# PROCESSING STRONGLY REDUCED FORMS
# IN CASUAL SPEECH

ISBN: 9789076203362
Cover design: Marije Brouwer and Marco Jongmans
Printed and bound by Ponsen & Looijen bv, Wageningen

# Processing strongly reduced forms in casual speech

Een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

**Proefschrift**
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op 20 december 2010
om 10:30 uur precies

door
**Susanne Brouwer**
geboren op 19 oktober 1981
te Zaanstad

| | |
|---|---|
| Promotor: | Prof. dr. A. Cutler |
| Copromotores: | Dr. H. Mitterer (MPI) |
| | Dr. F. Huettig (MPI) |
| Manuscriptcommissie: | Prof. dr. R. Schreuder |
| | Prof. dr. H. Reetz (Goethe University Frankfurt) |
| | Dr. N. Warner (University of Arizona) |

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

*Ad fontes*

'To the sources'

# ACKNOWLEDGEMENTS

May 21, 2006. I still remember this date vividly because on this date I was interviewed for a PhD position in the Language Comprehension Group at the Max Planck Institute for Psycholinguistics (MPI) in Nijmegen. A group of great scientists were sitting in front of me asking me questions about my past education and my future plans. During that interview I realized right away: this is the place where I would like to do my PhD, this is the place where I would like to learn, and this is the place where lots of interesting, innovative research is conducted.

Four months later my first day as a PhD student began. Anne Cutler, the director of the Language Comprehension Group, welcomed me warmly. Anne, I would like to thank you for giving me the opportunity to work in your group and for your belief in me that I could become a good scientist. You have created a unique, inspiring environment in the Language Comprehension Group of which many scientists can only dream of. I always enjoyed your never-fading enthusiasm when you presented your own work or when you were intrigued by somebody else's work.

I owe a lot to my two advisors, Holger Mitterer and Falk Huettig. Holger, where do I start? Day in and out I was amazed by your "brightness", your speed of processing things, and your inventive ideas. Especially in the beginning of my PhD you often dazzled me, but along the way things fell into place for me. Your expertise in programming and in statistics made my life so much easier. I also enjoyed seeing your happiness when I showed you the newest data. I found your interpretation of the data always very clever. You taught me to discover the interesting patterns in a complicated data set and you showed me how to highlight these in a paper. Holger, you are such a passionate scientist. Thanks for all your support.

Falk, you joined in a later, perhaps the most difficult, phase of my PhD project. Although I had run most of the experiments for my thesis, the papers were not so much in shape yet. You have supported me very much throughout this process. I would like to thank you for playing the "devil's advocate" (and these are your own words!) by asking me questions about my choice for certain experimental designs and my reasons for certain interpretations. Your critical view kept me sharp. I enjoyed looking at my findings from your (higher-level processing) perspective. And finally, I thank you for your great sense of humor, which made all of the above a very joyful experience for me.

I am also grateful to Mirjam Ernestus, James McQueen, and Rian Zondervan. Mirjam, you were my advisor in the very beginning of my PhD project. I would like to thank you for introducing me to the "world of reductions". Your comments on some of the papers have been very helpful and I hope you will enjoy reading this final version of the project. James, thanks for being my dean throughout my PhD years and for giving me support when necessary. I appreciate your critical questions during our group meetings and your ability to discuss all sorts of matter. Rian, thanks for helping me buying good tickets and finding safe accommodations to fly and reside all around the world. Thanks to the MPI, I presented at conferences in Germany, New Zealand, England, Spain, the United States of America, and France.

I would further like to thank all other members of the Language Comprehension Group. It is special to be part of a "family" that shares one common interest. The good atmosphere in the group contributes a great deal to this.

Very special thanks go out to two people who made my time at the MPI a very pleasant experience: Caroline Junge and Matthias Sjerps. Caroline, you are much more than a colleague to me. I always enjoyed your company, our conversations about our personal lives (and about our research!), and your caring, thoughtful nature. I am going to miss having you around. You are a true friend. Matthias, I especially remember all the fun hours we spend in the recording booth and our chit-chats on our bikes back home. And thank you for listening to my ups and downs of my research. Caroline and Matthias, I am very happy that you will be standing by my side during the defense as my paranymphs.

Other very important people to my life at the MPI were my roommates: Eva Reinisch, Caroline Junge, Attila Andics, Patrick van der Zanden, and Jiyoun Choi. Eva, we have shared our (bright, north-side, always cold, but beautiful view) room for many years. Thank you for being such an ambitious person, which helped me to remain being focused on my research as well. Patrick and Jiyoun, you were my roommates in the final phase of my PhD. I enjoyed having you two "youngsters" around. Your presence made me feel very happy in our office. Officially, Marijt Witteman was not one of my roommates, but it sometimes felt that way. Marijt, thanks for stopping by at my door and letting me share my social life with you. Thanks also for giving comments on my Dutch summary.

A very warm thank you goes out to three ladies, the "BOM-girls": Miriam Ellert, Sonja Gipper, and Nicole Altvater-Mackensen. Thanks to them I stayed active during all those years sitting behind my computer. It was not only working out that we did together, but also having light or trivial conversations during lunch, during coffee/tea breaks, or during a walk around the block. Miriam and Sonja, I would also like to thank you for the moments when we danced the night away...I had so much fun!

Setting up experiments also means a great need for technical support. I would therefore like to thank the MPI Technical Group. Thanks go particularly out to Ad Verbunt, Alex Dukers, Johan Weustink, Rick Viersen, Ronald Fischer, and John Nagengast. I would also like to thank one of the student assistants, Jessica Koppers, for spending hours and hours with me on the transcription of my stimuli. We were such a good team ("All files processed. That was fun!").

After two-and-a-half years working on my thesis at the MPI, I went abroad to spend six months working with people from Northwestern University in Evanston, IL, USA. Thanks to Anne Cutler, Janet Pierrehumbert and the financial support of the

Hugh Knowles Fund, I could work with Ann Bradlow from the Speech Communication Research group. It was a pleasure to see how life is in another research group. Ann, thank you for sharing all your knowledge with me and for giving me the opportunity to develop myself further in the field of speech perception. I am now enjoying my first postdoc position in your lab at Northwestern University to the fullest and I hope we will collaborate even more in the future. I also would like to express my gratitude to two members of the "EDAM"-team: Kristin van Engen and Lauren Calandruccio. Thank you for making the abroad-experience so joyful and for making the EDAM-project a (very productive) success. Finally, I would like to thank Bic Wirtz for literally giving me a home in Evanston, for all those delicious suppers, and for giving me insight into her exceptional view on life. And, of course, I would like to thank the Callahans for allowing me to change the male-female ratio in their house. I hope you enjoyed it as much as I did.

Besides people at the MPI, I would also like to thank three friends from Utrecht: Martijn van Drie, Liesbeth ten Klooster, and Lidwien Langenberg. Thanks to them my weekends were always packed with social activities. Martijn, thanks for all those phone calls when we talked about the things which really matter in life. Ahum. Liesbeth, thank you for letting me stay over so often and for all the dinners and breakfasts we had together. Lidwien, you are the funniest friend I have ever had, thanks for making me laugh.

My family has been another very important source of support for me: Steffen Brouwer, Margaret Zweedijk, and Marije Brouwer. Dear dad, I would like to thank you for taking me on all those fascinating trips when I was young. I think that these trips created my character to never be afraid for the unknown and to start new things with a smile. Dear mum, thank you for listening to me all these years and for giving me such good advice when I needed it. You are an inspiration to me, not only as a mother, but also as somebody who is fully devoted to her work. Dear Marije, thank you for being such a good example for me. I am impressed by the way you have developed yourself professionally and as a person. You are the kindest sister I could have imagined for. I would also like to thank you and your boyfriend, Marco Jongmans, for creating the cover of my thesis.

Finally, but certainly of very important value, my thankfulness goes out to my boyfriend, Brian Doornenbal. Brian, I had wished for somebody who could unconditionally love me and I found that person in you. I appreciate it so much that you support me to spread my wings in new directions and at the same time being there for me. Dear Brian, thank you for being you.

Susanne Brouwer

# TABLE OF CONTENTS


**Chapter 1** **19**

**Introduction**

Topic of the thesis …………………………………………………………….. 19

Research questions ……………………………………………………………. 25

The visual world paradigm ……………………………………………………. 27

Outline of the thesis …………………………………………………………... 29

**Chapter 2** **33**

**Speech reductions change the dynamics of spoken word recognition**

Abstract ……………………………………………………………….………. 33

Introduction ……………………………………………………..………….. 34

Experiment 1 …………………………………………………………………. 41

Method ………………………………………………………………..…….. 43

Participants ………………………………………………….………... 43

Materials……………………………………………………………………. 43

Procedure .……………………..……………………………………… 47

Design and analysis …………………….…….………….……….……. 48

Results ……………………………………………………….………….... 51

Accuracy and response time measures ……………………….……….. 51

Eye movements ……………………………………………………………. 52

Discussion …………………………………………………….…………. 55

Experiment 2 ………………………………………………………………. 58

Method ……………………………………………………………………….. 59

Participants ………………………………………………….………... 59

Materials and procedure ………………………………………………… 59

Design and analysis ........................................................ 60
Results ........................................................ 60
Accuracy and response time measures ........................................................ 60
Eye movements ........................................................ 61
Discussion ........................................................ 62
Experiment 3 ........................................................ 65
Method ........................................................ 65
Participants ........................................................ 65
Materials and procedure ........................................................ 65
Design and analysis ........................................................ 66
Results ........................................................ 66
Accuracy and response time measures ........................................................ 66
Eye movements ........................................................ 67
General discussion ........................................................ 68

**Chapter 3** **75**
**Phonological competition and the processing of reduced forms**

Abstract ........................................................ 75
Introduction ........................................................ 76
Experiment 1 ........................................................ 81
Method ........................................................ 81
Participants ........................................................ 81
Materials ........................................................ 82
Procedure ........................................................ 85
Design and analysis ........................................................ 86
Results ........................................................ 87
All trials ........................................................ 89
Correct and incorrect trials ........................................................ 90

Discussion ........................................ 92
Experiment 2 ........................................ 93
Method ........................................ 93
Participants ........................................ 93
Materials ........................................ 94
Procedure, design, and analysis ........................................ 95
Results ........................................ 95
Discussion ........................................ 98
Experiment 3 ........................................ 101
Method ........................................ 101
Participants ........................................ 101
Materials and procedure ........................................ 101
Design and analysis ........................................ 104
Results and Discussion ........................................ 105
General discussion ........................................ 108

**Chapter 4** **117**

**Discourse context and the recognition of reduced and canonical spoken words**

Abstract ........................................ 117
Introduction ........................................ 118
Pretest ........................................ 121
Method ........................................ 122
Results ........................................ 123
Experiment 1 ........................................ 124
Method ........................................ 124
Participants ........................................ 124
Materials ........................................ 125
Procedure ........................................ 127

Design and analysis ........................................................... 128
Results ........................................................... 129
Discussion ........................................................... 132
Experiment 2 ........................................................... 133
Method ........................................................... 133
Participants ........................................................... 133
Materials and procedure ........................................................... 133
Design and analysis ........................................................... 134
Results and Discussion ........................................................... 134
General discussion ........................................................... 135

**Chapter 5** **139**
**Shadowing reduced speech and alignment**

Abstract ........................................................... 139
Introduction ........................................................... 140
Method ........................................................... 143
Participants ........................................................... 143
Materials ........................................................... 143
Procedure ........................................................... 144
Design and analysis ........................................................... 144
Results ........................................................... 145
Error rate ........................................................... 145
Duration alignment ........................................................... 145
Alignment to segment realizations ........................................................... 146
Discussion ........................................................... 147

**Chapter 6** **151**

**Summary and conclusions**

Summary of results ........................................................ 151

Conclusions ................................................................ 157

**References** ................................................................ **163**

**Appendix** ................................................................ **173**

**Samenvatting en conclusies** ............................................ **177**

**Curriculum vitae** ........................................................ **189**

**MPI Series in Psycholinguistics** ........................................ **191**

# INTRODUCTION

**CHAPTER 1**

## Topic of the thesis

In our daily lives, much time is spent on using speech for communication purposes. Speakers produce on average about 16.000 words per day (and contrary to popular belief men produce just as many words as women: Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007). The speech used in such daily, informal interactions is produced in a casual way. An important characteristic of casual speech (also referred to as spontaneous, conversational, or natural speech) is that it contains large amounts of variation. Any one word is almost always pronounced differently on different occasions, and pronunciations can vary both from one speaker to another and from one situation to another. For example, a Dutch speaker may say the word *beneden* 'downwards' once in its canonical pronunciation /bəneːdə/ and once as [məneːə]. In the last case, the segment /b/ has changed into the segment [m] and the segment /d/ is completely absent. Such variation in production is called speech reduction, where segments, syllables and even whole words can be changed and/or deleted (e.g., Ernestus, 2000; Johnson, 2004; Shockey, 2003).

Sometimes a certain reduction can be ascribed to a particular speaker (one well-known person in the Netherlands for "swallowing" sounds in general is former Dutch Prime Minister Balkenende), but in one degree or another, everybody produces reduced forms. Although reduced forms can deviate strongly from their canonical forms (also referred to as citation, full, or unreduced forms), listeners typically understand each other without any difficulty. Most listeners (and speakers) are even unaware of the fact that reduced forms occur so often in conversational speech (e.g., Kemps, Ernestus, Schreuder, & Baayen, 2004). This apparent paradox leads to the question: *how are strongly reduced forms processed?* This thesis addresses this question.

Although listeners encounter casual speech more often, the speech style most frequently used in psycholinguistic research is laboratory speech (also referred to as careful, read, or oralized written speech; Mehta & Cutler, 1988). Laboratory speech is typically created in a soundproof room from which all noise is excluded. A selected speaker reads a prepared set of stimuli several times, whereby attention can be paid to the careful pronunciation of each segment of the target word or phrase. The decision to use a highly constrained set of careful speech materials is to a great extent motivated by methodological considerations because it means that the data can be studied in a highly controllable way. This also explains why studies with casual speech are comparably rare. Using casual speech stimuli necessarily means that a tight experimental control is lacking (see Warner, to appear, for an overview of methods for studying spontaneous speech). However, since read-aloud utterances produced in a soundproof booth may not be fully representative of casual speech production, research on everyday speech has grown in the last two decades. The availability of spontaneous speech corpora, databases of real-life conversations, helps to make this possible.

Mehta and Cutler (1988) were the first to compile a small spontaneous speech corpus to compare casual speech with laboratory speech. Stimuli were selected from an hour of recorded spontaneous conversation between two speakers. The two speakers returned to read the same sentences again in the laboratory. Participants were asked to perform a phoneme detection task on the stimuli from both speech styles (casual versus laboratory speech). The results showed that target recognition depended on the style of speech to which participants were listening. For example, in casual speech only, listeners responded faster to accented targets and strong syllables than to unaccented, unstressed targets. Based on their results, Mehta and Cutler argued for the importance of studying casual speech. Cutler (1998), however, pointed out that the research on spontaneous speech processes at that time still had the weakness that the materials used in most studies consisted of carefully controlled, laboratory speech (e.g., designing specific stimuli such as ‘gardenbench’ to trigger assimilation processes during a recording session). Evidently, such studies can contribute greatly to the knowledge of how casual speech is recognized, but one could question to what extent inferences based on laboratory speech can be generalized to more natural listening situations. The experiments reported in Chapter 2 aim to fill this scientific gap. Chapter 2 addresses the question of how listeners recognize casual speech extracted from a spontaneous speech corpus and how this is different from recognition of carefully-articulated speech.

Another apparent weakness of the previous research on spontaneous speech processes is that most work investigated the recognition of forms which minimally deviated from their canonical forms. Examples of such spontaneous speech processes are assimilation (e.g., ‘gardenbench’ /gɑrdnbɛntʃ/ → [gɑrdmbɛntʃ]), /t/-reduction (e.g., ‘postman’ /postmən/→ [posmən]), and schwa-deletion (e.g., ‘history’ /hɪstəɹɪ/ → [hɪstɹɪ]). In the assimilation example, the pronunciation of the syllable-final consonant /n/ is influenced by the initial consonant /b/ of the following syllable. In such a case, only one segmental change occurs (from /n/ to [m]) and therefore much of the acoustic evidence remains similar to the canonical form of the word.

In this thesis, the topic under investigation is strongly reduced forms (e.g., /bəneːdə/ → /məneːə/). Such forms are the result of spontaneous speech processes which have severely changed the canonical counterparts. Johnson (2004) has termed such large deviations from the canonical form ‘massive reductions’. Johnson found in a large database of American English conversational speech that complete syllables were deleted in no less than 6% of the words and that one or more segments were deleted in 25% of the words. Similar conclusions were drawn by Ernestus (2000) for Dutch, by Kohler (1990) for German, and by Shockey (2003) for British English.

These statistics show that 'massive reductions' are an unavoidable feature of human language (see also Warner, submitted, for explanations why it is relevant to investigate the phenomenon of reduction).

It is surprising that, although strong reductions occur so frequently in casual speech, listeners still understand each other with ease. Only a few studies investigated how listeners recognize strongly reduced forms in casual speech (e.g., Arai, 1999; Ernestus, Baayen, & Schreuder, 2002; Kemps et al., 2004). Ernestus et al. (2002) investigated how Dutch listeners recognize words such as [mok] for /moxələk/ *mogelijk* 'possible'. Participants listened to reduced word forms with differing amounts of context and had to write down the words they had heard. The results showed that listeners hardly recognized the reduced forms in isolation at all. Recognition performance increased when the reduced forms were presented in a phonetic context, but reached ceiling level only when the context was several words long. These results indicate that reduced forms need a semantic/syntactic context to be recognized. Subsequently, Kemps et al. (2004) provided evidence that listeners unconsciously reconstruct canonical forms from strongly reduced forms that are embedded in a sentence context. For example, they reported hearing the segment [l] in the strongly reduced form [mok] from /moxələk/. These studies thus show that the recognition of strongly reduced forms benefits greatly from a context of several words (Ernestus et al., 2002) and that the recognition involves the activation of the canonical form (Kemps et al., 2004). A drawback of these two studies is the use of offline tasks which reflect post-perceptual processing. The research reported in Chapter 2, 3 and 4 extends the previous offline findings by examining how listeners recognize strongly reduced forms *online*.

Both online and offline techniques have been used to investigate how listeners access and subsequently recognize spoken words. This type of research has mainly used laboratory speech to show how quickly and efficiently listeners analyze the continuous speech signal (e.g., Goldinger, Luce, & Pisoni, 1989; Zwitserlood, 1989). For example, Zwitserlood presented gated fragments of Dutch words (e.g., *kapitein* 'captain') in a cross-modal priming experiment. The auditory, gated fragments (e.g., /k/, /kɑ/, /kɑp/ etc.) were followed by visually presented target words for lexical decision. The results showed that word-initial, partial matches activated different lexical candidates from gate to gate. For example, hearing /kɑpi/ facilitates the recognition of words with overlapping onsets such as *kapitaal* 'capital', whereas phonologically unrelated words such as *schilderij* 'painting' are not activated. This finding has been often replicated and extended. Lexical access thus involves the continuous activation of multiple lexical candidates. Listeners evaluate which word in the mental lexicon, the dictionary in people's minds in which all lexical knowledge is stored, is most consistent with the speech signal. Words inconsistent with the speech signal compete less for recognition. This process of competition among the lexical candidates eventually provides the ultimate winner of the word recognition process (see McQueen & Cutler, 2001, for further discussion). A related, important finding is that lexical candidates with initial overlap with the word to be recognized (e.g., 'captain' for /kæp/) compete more strongly than words with medial or final overlap (e.g., 'apple' for /kæp/; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). However, almost all of the studies to date have used laboratory speech to investigate lexical competition. The experiments in Chapter 2 and 3 extend these studies by examining the temporal dynamics of phonological competition as strongly reduced forms of spoken words acoustically unfold.

Chapter 3 also addresses the issue of which mechanisms underlie the recognition of reduced word forms. In the literature, various accounts have been postulated for how pronunciation variants are recognized. The accounts differ in the way variants are represented in the mental lexicon. One main account argues that variants are not represented at all and that only full forms (e.g., [bəneːdə]) are stored in the mental lexicon. Pronunciation variants (e.g., [məneːə]) are recognized because reconstruction takes place. This process occurs at a prelexical level which mediates between the speech signal and the mental lexicon. Listeners use fine phonetic detail (e.g., Gow, 2002), the phonological context (e.g., Gaskell, 2003), or top-down processes (McClelland & Elman, 1986; Warren, 1970) to reconstruct canonical forms from reduced forms. For example, in the case of [məneːə], fine phonetic cues in the segment [m] may 'inform' listeners that this segment is actually a [b]. This process helps listeners to reconstruct the full form /bəneːdə/ from its reduced form [məneːə], which accordingly activates the canonical representation of this word in the mental lexicon.

Another main account argues that all pronunciation variants of a word are stored in the mental lexicon. This means that not only the canonical form /bəneːdə/ is stored; but that the reduced form [məneːə] is stored as well. At least two different views of this account exist. The episodic view argues for fine-grained storage of every variant (e.g., Bybee, 2001; Goldinger, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert 2001), whereas the other view argues that abstract variants are stored (e.g., Connine, 2004; McLennan, Luce, & Charles-Luce, 2003). Chapter 3 presents data relevant to the evaluation of these various accounts.

Chapter 4 extends previous studies which have demonstrated that the phonological and the sentential context contribute strongly to the successful recognition of reduced forms (Arai, 1999; Ernestus et al., 2002; Kemps et al., 2004). Ernestus and colleagues showed that when strongly reduced forms were presented without any context, recognition failed in about 50% of the cases. Performance improved when strongly reduced forms and their adjacent segments were presented (70% correct). Listeners only recognized strongly reduced forms well when these forms were presented in sentential contexts. However, the strongly reduced forms were still misidentified in almost 10% of the cases. The experiments reported in

Chapter 4 go a step further by presenting reduced and canonical forms in a wider discourse context (i.e., more than one sentence). The experiments test how the discourse context helps the recognition of reduced forms and canonical forms. The critical questions are whether reduced forms profit more from discourse context and whether this might be attributable to discourse processing or simply speaker adaptation.

The first three series of experiments (Chapters 2-4) examine how reduced forms are processed in perception. The experiment reported in Chapter 5 examines whether hearing reduced forms affects production. When, for example, a listener hears the reduced form [məneːə], is the listener most likely to produce an exact copy of this form or will the listener produce something closer to the full form [bəneːdə]? Previous research has shown that the link between perception and production is close in both non-social settings (e.g., Fowler, Brown, Sabadini, & Weihing, 2003; Goldinger, 1998; Porter & Castellanos, 1980) and in more natural situations (e.g., Branigan, Pickering, & Cleland, 2000; Pardo, 2006; Pickering & Garrod, 2004). The experiments in Chapter 5 extend these previous findings by examining whether listeners also imitate reduced forms in natural contexts.

## Research questions

After this brief introduction of the topics of each experimental chapter of this thesis, let us now summarize the four main research questions which will be addressed:

1. Does spoken word recognition during casual speech differ from laboratory speech?
2. Which phonological competitors take part in the competition process when listeners hear strongly reduced forms?
3. Does discourse context affect the recognition of reduced forms differently than the recognition of canonical forms?

4. Do listeners accommodate to reduced forms in their own subsequent production?

To answer these four questions we have to work with casual speech in which reductions often occur. We extracted our materials from a spontaneous speech corpus called the Spoken Dutch Corpus (Oostdijk, 2000). This corpus contains approximately 900 hours of speech of standard Dutch (circa 9 million words) spoken by Flemish and Dutch adult speakers. 225 hours of speech recordings are spontaneous, face-to-face conversations. All recordings have been aligned with orthographic transcriptions. Since the participants used in all our studies are Dutch, we restricted ourselves to Dutch (not Flemish) speakers.

From a pool of 100 polysyllabic, mid-to-high frequency content words, we selected 32 words for which we could find a reduced (e.g., [məneːə]) and a canonical realization (e.g., [bəneːdə]). These 64 word forms were produced by 59 different speakers. The canonical realizations were almost always fully realized. There is considerable variation in the reduced realizations: one or more segments were absent or changed in each reduction. For example, a reduced form could deviate from its canonical counterpart in its initial part (first or second segment), such as [məneːə] for [bəneːdə], or in a later part (third to fifth segment), such as [ʋɛs] for [ʋɛtstrɛit] *wedstrijd* ‘match’. The critical criterion for a reduced form to be included in the study was that it shared more initial segments with another existing Dutch word than its own canonical form. If the existing word shared more phonological onset overlap with the reduced form than with the canonical form, it was termed a “reduced form” competitor. However, if the existing word shared more phonological onset overlap with the canonical form than with reduced form, it was termed a “canonical form” competitor. For example, for the reduced form [məneːə] and its canonical counterpart [bəneːdə] the word [məneːr] *meneer* ‘mister’ was the “reduced form” competitor and the word [bəna:de:lə] *benadelen* 'to disadvantage' was the "canonical form" competitor. These competitors are visually displayed on a screen during an eye-tracking task in Chapter 2 to 4. In Chapter 2 and 4, the target and a phonologically unrelated distractor are also displayed on the screen. In Chapter 3, however, the target is not displayed, but replaced by another phonologically unrelated distractor. Note that the relationships among target, competitors, and distractor are relevant for all experiments in Chapter 2 to 4.

## The visual world paradigm

The first series of experiments examines how listeners recognize reduced forms in real time (Chapter 2-4). Previous work on the recognition of reduced forms has mainly used offline tasks such as phoneme monitoring task (Kemps et al., 2004) to investigate how listeners recognize strongly reduced forms. An advantage of offline tasks is that they are relatively easy to construct and administer; however, a disadvantage is that such tasks require listeners' meta-linguistic judgments. This may lead to an over- or underestimation of participants' language abilities. An online psycholinguistic technique that is well-suited to examine how listeners recognize reduced forms online is the visual world paradigm. This paradigm has three main advantages: 1) participants are not instructed to solve a metalinguistic task related to the research question; 2) the temporal resolution allows for real-time precision during spoken language processing; and 3) a topic can be studied under more natural conditions in which listeners hear words, sentences, or stories which are pragmatically relevant. This paradigm is used to address the first three main research questions concerning the recognition of strongly reduced forms in casual speech (Chapter 2-4).

In a seminal article, Cooper (1974) introduced the task now known as the visual world paradigm. He demonstrated that participants' eye movements are closely time-locked to the unfolding speech input. For example, participants looked at a picture of a lion in a visual display upon hearing the word 'lion' in a spoken story. It was more than twenty years later that the use of measuring eye movements was further applied to the study of spoken language comprehension. Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) used a task in which participants had to follow spoken instructions to move objects in a visual display, while their eye movements were recorded. This version has been used extensively over the last decade.

Allopenna, Magnuson, and Tanenhaus (1998) were the first to study the competition process during spoken word recognition in the visual world paradigm. A visual display in their study consisted of four pictures and four geometrical shapes. The task of the participants was to pick up and move one of the objects (e.g., 'Pick up the beaker. Now put it below the diamond'). The activation of words that shared initial phonemes with the target word (e.g., 'beetle' for 'beaker') and the activation of words that rhymed with the target word (e.g., 'speaker' for 'beaker') were measured over time. The findings showed that, upon hearing the target word 'beaker' both 'beetle' and 'speaker' become activated, but competition is stronger for 'beetle' than for 'speaker'. This result confirms previous work that lexical candidates with initial overlap with the word to be recognized compete more strongly than words with medial or final overlap (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). In addition, the results once again support the finding that lexical access involves the continuous activation of multiple lexical candidates (e.g., Goldinger et al., 1989; Zwitserlood, 1989). Lexical candidates consistent with the speech input (e.g., 'beetle' for 'beaker') compete more for recognition than candidates that are inconsistent with the input ('apple' for 'beaker').

Similar results to those found in Allopenna et al.'s (1998) study were obtained in a printed-word version of the paradigm (McQueen & Viebahn, 2007, see Huettig & McQueen, 2007, for further discussion and validation of the method). In such a variant, participants listen to an utterance while seeing printed words instead of pictures on the visual display. For example, listeners look more often at onset-matching competitors (e.g., *buffer* for *buffel* 'buffalo') than at offset-matching competitors (e.g., *lotje* 'lottery ticket' for *rotje* 'fire-cracker'). In this thesis, the printed-word version of the eye-tracking paradigm is used such that (target) words which are not (easily) depictable can be included.

In this thesis, yet another variant of the eye-tracking method is used: the target-absent version of the visual world paradigm (Huettig & Altmann, 2005; see Huettig & McQueen, 2007, for discussion). Huettig and Altmann have shown that excluding the target word from the visual display (e.g., hearing *beneden* but not seeing 'beneden' on the screen) greatly increases the likelihood of observing competition effects. In this thesis, the target word was therefore replaced in some studies by another unrelated distractor in the visual display.

In sum, this thesis combines two variants of the visual world paradigm (i.e., printed words rather than pictures and target-absent trials). Both variants have been used successfully in the past and both allow more flexibility for identifying usable competitors and thus increasing the magnitude of competitor effects. The online results from the visual world paradigm will therefore build further on the previous offline work on the recognition of reduced word forms.

## Outline of the thesis

The major objective of this thesis is to examine how listeners process strongly reduced forms in casual speech. Chapter 2-4 employed the visual world paradigm to examine how reduced forms are recognized and how these forms impact on the phonological competition process. The eye-tracking methodology is a promising new experimental tool for studying these questions. Chapter 5 reports a shadowing task to investigate how listening to reduced forms influences speech production. Previous work has used the shadowing task which provided evidence for imitation of segments, syllables, and words. The shadowing task is therefore well-suited to examine how listeners perceive and subsequently produce reduced forms.

Three eye-tracking experiments in Chapter 2 address the first research question: *Does spoken word recognition during casual speech differ from laboratory speech?* In the first experiment, canonical and reduced forms are presented in sentential and syllabic contexts, while four printed words are displayed on the screen: the target word (e.g., *beneden* 'downwards'), one competitor beginning similarly to the canonical form (e.g., *benadelen* 'to disadvantage'), one competitor beginning similarly to the reduced form (e.g., *meneer* 'mister'), and an unrelated distractor. In the second experiment, reduced forms are excluded and only canonical forms are presented. The canonical forms were embedded in sentences with different speech styles to test whether phonological competition is influenced by speech style. The third and final experiment intermixes canonical with reduced forms in casual speech to examine whether the competition pattern in the first experiment can be replicated. The critical question is whether the processing of laboratory and casual speech differs and what the impact of reductions is on the recognition process.

Chapter 3 reports a series of three eye-tracking experiments to examine the second research question: *Which phonological competitors take part in the competition process when listeners hear strongly reduced forms?* The design of the experiments in this chapter is similar to Chapter 2, except that the visual target is excluded and replaced by a second distractor to boost competitor effects. The visual display hence has the following structure: one competitor beginning similarly to the canonical form (e.g., *benadelen* 'to disadvantage'), one competitor beginning similarly to the reduced form (e.g., *meneer* 'mister'), and two unrelated distractors (e.g., *vakantie* 'holiday' and *juweel* 'jewel'). The first experiment presents canonical and reduced forms in isolation. This experiment also tests (as in Chapter 2) whether the attractiveness of different competitors can be influenced at all by the acoustic form of the target word (i.e., canonical versus reduced). Therefore, an experimental situation is created in which this seemed most likely (target-absent design). The second experiment presents the same word forms in sentential contexts to investigate the phonological competition process during the actual recognition of strongly reduced forms. The question is whether listeners also actively consider lexical candidates compatible with the acoustic structure of the reduced forms during online recognition. In the third experiment the reduced forms were manipulated. For example, the surface segment [m] in [məneːə] is replaced by a real, intended onset [m] from *met* 'with'. The aim is to examine whether listeners are sensitive to fine phonetic information in strongly reduced forms.

While Chapter 2 and 3 presented canonical and reduced forms in sentential contexts, Chapter 4 goes further by presenting target sentences alone or with the wider discourse context. The eye-tracking experiments reported in Chapter 4 hence address the third research question: *Does discourse context affect the recognition of reduced forms differently than the recognition of canonical forms?* The same visual displays are presented as in Chapter 2. The question is whether reduced forms benefit more from discourse context and whether this is due to discourse processing or adaptation to the target speaker.

The experiment reported in Chapter 5 deals with the fourth research question of this thesis: *Do listeners accommodate to reduced forms in their own subsequent production?* Participants perform a shadowing task in which they have to repeat back canonical and reduced forms embedded in target sentences. The aim is to give insight into the question how listeners perceive and subsequently produce reduced forms. Finally, Chapter 6 summarizes the results, attempts to give answers to all four research questions, and provides a general discussion of the main findings of this thesis.

# SPEECH REDUCTIONS CHANGE THE DYNAMICS OF SPOKEN WORD RECOGNITION

**CHAPTER 2**

This chapter is a slightly revised version of Brouwer, S., Mitterer, H., & Huettig, F. (under revision). Speech reductions change the dynamics of spoken word recognition. *Language and Cognitive Processes*.

## ABSTRACT

Three eye-tracking experiments investigated how phonological reductions (e.g., 'puter' for 'computer') influence phonological competition. Participants listened to sentences extracted from a spontaneous speech corpus and saw four printed words: a target (e.g., 'computer'), a competitor similar to the canonical form (e.g., 'companion'), one similar to the reduced form (e.g., 'pupil'), and an unrelated distractor. In Experiment 1, we presented canonical and reduced forms in syllabic and sentential contexts. Listeners directed their attention to a similar degree to both competitors independent of the target's spoken form. In Experiment 2, we excluded reduced forms and presented canonical forms only. In such a listening situation, participants showed a clear preference for the "canonical form" competitor. In Experiment 3, we presented canonical forms intermixed with reduced forms in sentence contexts, and replicated the competition pattern of Experiment 1. These data suggest that listeners penalize acoustic mismatches less strongly when listening to reduced speech than when listening to fully-articulated speech. We conclude that flexibility to adjust to speech-intrinsic factors is a key feature of the spoken word recognition system.

# INTRODUCTION

Most research on spoken word recognition has focused on careful speech read aloud by selected speakers (see Cutler, 1998). The advantage of using careful speech materials is that they are highly controllable and intelligible. Such materials have provided valuable insights into key constructs of spoken word recognition such as lexical competition (e.g., Goldinger, Luce, & Pisoni, 1989; Marslen-Wilson, 1987). However, in listeners' everyday communicative exchanges, they most often encounter casual speech, in which words are often pronounced with fewer segments than when they are produced in the laboratory. For example, the word 'hilarious' [hilɛriʌs] is realized as [hlɛrɛs] in a corpus of casually spoken English (Johnson, 2004). Nevertheless, people typically do understand each other with ease. Only a few attempts have been made to study speech "in the wild" (e.g., Ernestus, Baayen, & Schreuder, 2002; Mehta & Cutler, 1988). In this article, we investigate whether spoken word recognition during casual speech differs from spoken word recognition during carefully pronounced speech recorded in the laboratory.

Research using laboratory speech has been very successful. It has demonstrated that listeners rapidly analyze the speech signal and that the processing of speech is closely time-locked to the input (e.g., Goldinger et al., 1989; Zwitserlood, 1989). In a cross-modal priming experiment, for instance, Zwitserlood presented gated fragments of Dutch words, such as *kapitein* 'captain', which were followed by visually presented target words for lexical decision. The gated fragments were successively longer onsets (/k/, /kɑ/, /kɑp/, etc.) of words. The Zwitserlood study showed that partial information of onset fragments activated different matching candidate words from gate to gate. For example, when hearing *kapi*... listeners responded faster to words with overlapping onsets, such as *kapitaal* 'capital' than when they heard the beginning of a phonologically unrelated word. Lexical access thus involves the continuous activation of multiple lexical candidates. As more acoustic evidence becomes available, candidates inconsistent with the speech signal compete less for recognition than candidates that are consistent with the input. Thus the ultimate winner of the word recognition process emerges from a competition process among these candidates (see McQueen & Cutler, 2001, for further discussion).

An important finding of laboratory research is that lexical candidates with initial overlap with the word to be recognized compete more strongly than words with medial or final overlap (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). Upon hearing the spoken sequence /kæp../, all words that start with these sounds, such as 'captain', are activated in parallel but words that overlap later in time such as 'apple' /æp../ are less activated. Such effects have been particularly clearly demonstrated in eye-tracking studies that used the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, listeners' eye movements to pictures of objects on a computer screen are measured in response to concurrent speech. Proportion of fixations is typically taken to be related to underlying activation levels of word candidates. Eye movements are continuously recorded, so that it is possible to evaluate relative competitor activation over time. The paradigm thus provides closely time-locked measures of the ongoing spoken word recognition process.

Using this method, Allopenna, Magnuson, and Tanenhaus (1998) showed that listeners fixate more often on pictures with names similar to the target name than to phonologically unrelated names. In that study, participants' eye movements were tracked as they looked at four pictures on a computer screen (e.g., a 'beaker', a 'beetle', a 'speaker', and a 'carriage'). They listened to spoken instructions such as 'Pick up the beaker'. Participants looked at the pictures of both types of competitors, but more often to competitors matching at word onset (e.g., the 'beetle') than competitors matching at word offset (e.g., the 'speaker'; but see Connine, Blasko, & Titone, 1993). McQueen and Viebahn (2007) replicated these results using printed-word displays. In their study, participants' eye movements were recorded as they looked at four printed words on a computer screen. As in the study by Allopenna et al., participants looked more often at phonological competitors than at phonologically unrelated distractors and the effect was stronger for onset-matching competitors (e.g., *buffer* for *buffel* 'buffalo') than for offset-matching competitors (e.g., *lotje* 'lottery ticket' for *rotje* 'fire-cracker'). In the present study we use this printed-word version of the paradigm.

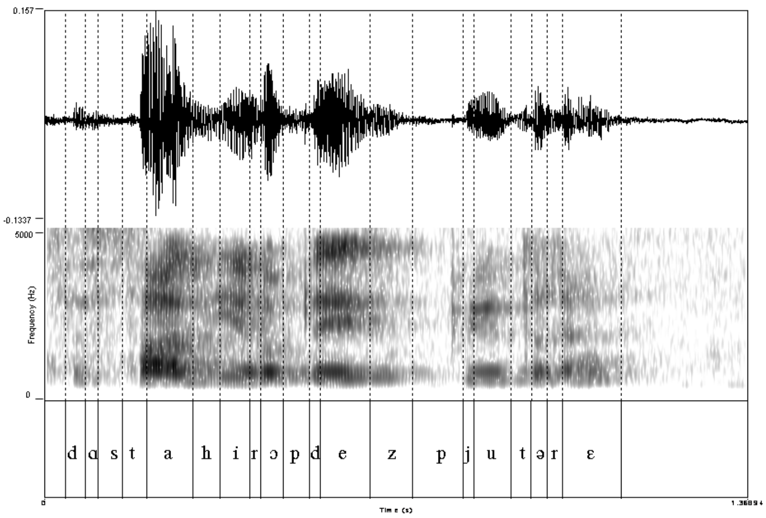
Huettig and McQueen (2007) have recently further validated this method through eye-tracking experiments with both picture and printed-word displays. Previous work showed that eye movements in the paradigm can be based on semantic (e.g., Huettig & Altmann, 2005), visual (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004; 2007), and phonological matches (Allopenna et al., 1998). Huettig and McQueen examined more closely the influence of these three types of matches. When they presented participants with the picture version of the paradigm they observed a strong influence of all three types of representations on participants' eye movements. Importantly for the present purposes, their study also showed that only phonological representations influence eye gaze when printed-word displays were used. Huettig and McQueen concluded that the printed-word version is more sensitive to phonological manipulations than the version using pictures. Weber, Melinger, and Lara Tapia (2007) provided further support for this view. They found that written displays produced stronger phonological competition effects than pictorial displays. Another reason to use the printed-word variant is that it is less sensitive to confounds from other types of item variability (e.g., semantic relatedness). The printed-word variant of the paradigm thus has been very successful for the investigation of phonological competition during carefully pronounced speech recorded in the laboratory.

To accommodate the finding of strong onset and weak offset competition, it is usually assumed that mismatches lead to strong deactivation of a target word. This assumption is explicitly made in the original Shortlist model (Norris, 1994), where the activation of a word candidate increases by one unit for every matching segment, but decreases by three units if there is a mismatch. In the TRACE model (McClelland & Elman, 1986), there is no such explicit penalty for a mismatch, but the winner-takes-all competition on a lexical level leads to a strong deactivation of a word if another one matches better.

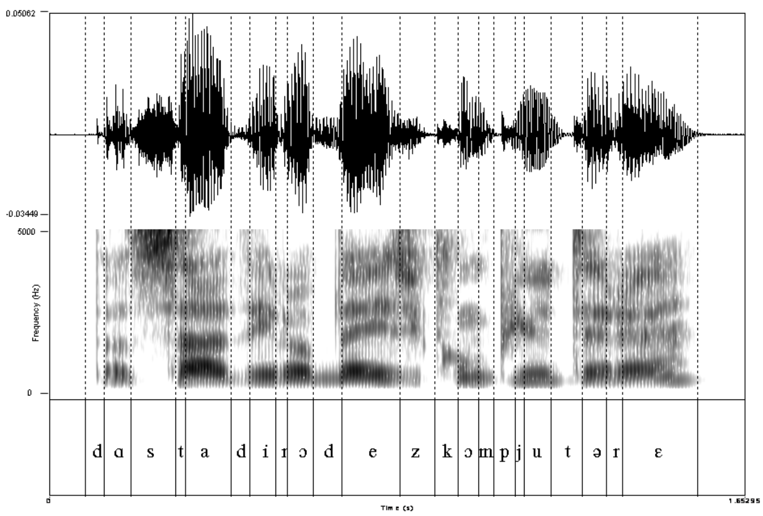
It is however as yet unknown to what extent the pattern of strong onset and weak offset competition also applies to casual speech. Given the huge amount of variation in casual speech, it might not be beneficial for the listener to weigh mismatches as strongly as some models of spoken word recognition suggest. Additionally, it is conceivable that the competitor words (the 'competitor set') may be rather different during casual speech in which speech reduction processes very frequently occur. Johnson (2004), for example, found that over 60% of the words in a spoken English corpus deviated from their citation form by at least one segment, and 28% of the words even deviated on two or more segments (see Ernestus, 2000, for convergent evidence for Dutch). To illustrate this phenomenon, Figure 1 shows a waveform and a spectrogram of the same Dutch sentence, once spoken casually, and once read out loud. We extracted the sentence *dat staat hier op deze computer, hè?* 'that is on this computer, isn't it?' from the Spoken Dutch Corpus (Oostdijk, 2000) and we re-recorded the same sentence in a laboratory setting. Figure 1 shows the waveform and spectrogram of both versions. Figure 1A shows the sentence from the spontaneous speech corpus, which is best transcribed as [dɑ staː ɦiːr ɔp deːz pjuːtər ɛ]. The same sentence read out loud was transcribed as [dɑ staːd iːr ɔ deːz kɔmpjuːtər ɛ]. Clearly, fewer segments are pronounced in the casually uttered sentence than the one recorded in the laboratory, resulting in a durational difference between the two speech fragments. These differences can best be illustrated if we focus on the word *computer* in these sentences (see Figure 2). The segments of the word *computer* in the read utterance are all fully pronounced (see Figure 2A). Figure 2B shows this word from the casually produced sentence. As can be seen, the first syllable [kɔm] of *computer* [kɔmpjuːtər] is missing. This is a clear example of a reduced realization of the target word *computer*.

In this paper, we address the question how such reductions in casual speech impact spoken word recognition. Given our analysis of casual speech, it is likely that word recognition in casual speech differs from word recognition in carefully articulated and fully pronounced speech. Consider for instance which words compete for recognition when the intended word is *computer*. According to the literature reviewed above, /k/-initial words such as *companion* should compete for recognition because they share initial overlap. However, it is unclear whether this is still the case when the word *computer* is intended but produced as /pjutər/. In such cases one may predict different competitor sets for canonical and reduced forms.

The aim of Experiment 1 hence is to examine whether phonological competition during casual speech is influenced by the exact phonetic form of the spoken word. In other words we examine the effect of hearing forms such as the reduced realization [pjuːtər] or the canonical realization [kɔmpjuːtər] of *computer* on competition processes during spoken word recognition.

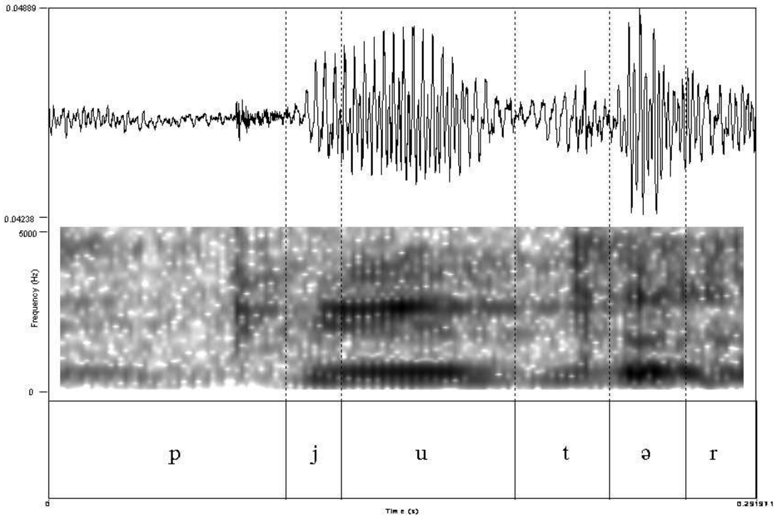



A)

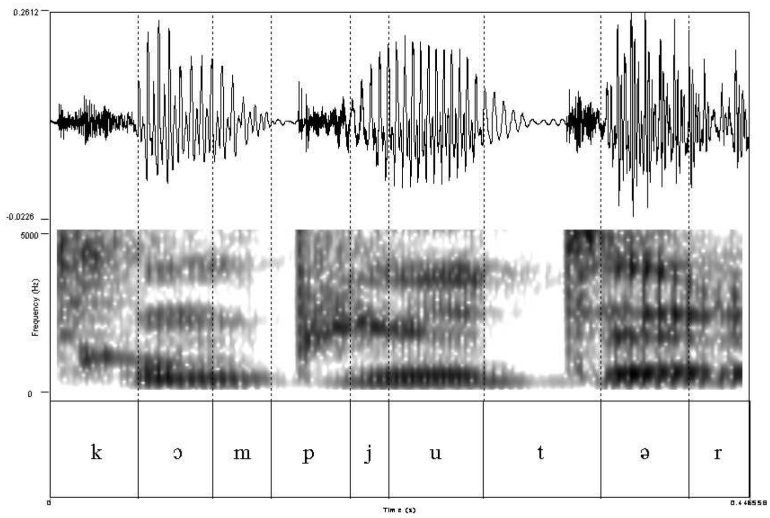



B)

Figure 1: *Realizations of the Dutch sentence* dat staat hier op deze <u>computer</u>, hè? *'that is on this <u>computer</u>, isn't it?' as produced in a spontaneous speech corpus (Fig. 1A) and as produced in the laboratory (Fig. 1B).*




A)




B)

Figure 2: *Realizations of the Dutch word* computer *as produced in a spontaneous speech corpus (Fig. 2A) and as produced in the laboratory (Fig. 2B). See text for details.*

# EXPERIMENT 1

We used casual speech in which the same target words appeared in either a canonical or in a reduced form. In order to investigate spoken word recognition in casual speech we have to work with extracts from speech corpora containing ecologically valid examples of casual speech. A disadvantage of using casual speech is that it is difficult to have a similar degree of control over stimulus selection as when creating new stimuli in the laboratory (for a discussion, see Warner, to appear). For example, it is important to establish which acoustic features in the casual speech fragments are precisely produced, a very time-consuming process involving the transcription of a great number of words. For the present study two independent raters transcribed more than 1400 tokens of 90 words.[1] On the basis of this corpus, we chose words which were produced (at least) once canonically and once in a reduced way.

A requirement for our selected stimuli was that a word exists in the Dutch language that has more phonological onset overlap with the canonical form than with the reduced form, henceforth called a "canonical form" competitor, and another word that has more phonological onset overlap with the reduced form than with the canonical form, henceforth called a "reduced form" competitor. For example, for the canonical form of the English word 'computer' [kɔmpjuːtər] the word 'companion' [kɔmpɛnjən] is a "canonical form" competitor, whereas for the reduced form

[1] Note that the two raters used IPA transcription as a way to represent which segments were present in the reduced speech. As a result, a single IPA transcription for a token (whether it is based on auditory judgment and/or on visual cues in the spectrogram) presumes that one or the other type of information is inaccurate, and encodes a single representation of a token. IPA transcription thus forces reduced speech into categories that may not really be appropriate. This has a direct consequence for the relationships among targets and their "reduced form" and "canonical form" competitors. That is, it is unsure how well the target and the competitors precisely overlap. However, the basic effects do work in the experiments. Therefore, IPA transcription of reduced speech seems to be fine-grained enough to provide a helpful representation of it.

[pjuːtər] of *computer* the word 'pupil' [pjuːpɪl] functions as a "reduced form" competitor.

Note that for 75% of the items (24 out of the 32, see Appendix) both competitors overlap phonologically at onset. As a result, some "reduced form" competitors are also to some extent competitors of the target word's canonical form. That is, they function as offset overlap competitors. The "canonical form" competitor, however, always had more phonological onset overlap with the canonical form than with the reduced form and the "reduced form" competitor always had more phonological onset overlap with the reduced form than with the canonical form. For instance, the word *directeur* 'director' was pronounced canonically as /diːrɛktøːr/, and in a reduced way as [dɪktøː] in the spontaneous speech corpus. The "canonical form" competitor *dirigere*n [diːriχeːrə] 'to conduct' shares the first three segments with the canonical form but shares only the first two segments with the reduced form. The "reduced form" competitor *dictator* [dɪktaːtɔr], however, shares three initial segments with the reduced form but shares only two initial segments with the canonical form. It is therefore crucial to compare the relative strength of the two competitors under different conditions.

The prediction from previous studies using laboratory speech (e.g., Allopenna et al., 1998; McQueen & Viebahn, 2007) therefore is that during listening to canonical forms our "reduced form" competitors will attract *less* overt attention than our "canonical form" competitors because they share no onset overlap (25% of the items) with the target or smaller onset overlap (75% of the items) than the "canonical form" competitors with the target words. It is, however, unclear what happens during listening to reduced forms. What matters more in such a case? If the acoustic input is crucial, the "reduced form" competitors should attract more overt attention than our "canonical form" competitors because in this condition the "reduced form" competitors overlap to a greater extent with the acoustic signal than the "canonical form" competitors. If, however, the canonical form of a word is still crucial, even if the input is reduced, then the "canonical form" competitors should attract more overt attention than the "reduced form" competitors. This may seem unlikely at first sight, but previous research indicated that listeners may fill in missing phonemes in the input (Warren, 1970; Samuel, 1996; Kemps et al., 2004), so that the input is restored to its canonical form. For looks to targets, we predict listeners to look more often and earlier in time to targets in the canonical than in the reduced form conditions, replicating earlier findings of reduction costs (cf. Ernestus et al., 2002; Kemps et al., 2004).

# METHOD

## Participants

Twenty-five participants from the Max Planck Institute's subject pool, mostly undergraduates at the Radboud University Nijmegen, took part in this experiment. All were native speakers of Dutch without any hearing problems and with normal or corrected-to-normal vision. They were paid for their participation.

## Materials

We selected 32 polysyllabic, mid-to-high frequency content words for which we could find reduced and canonical pronunciations in the spontaneous speech subcorpora of the Spoken Dutch Corpus (Oostdijk, 2000). For each reduced realization one or more segments were absent or changed (e.g., [məneːə] for [bəneːdə] *beneden* 'downwards'). There is considerable variation in the reductions (see Appendix). For example, a reduced form could either deviate from the canonical form in its initial part (first or second segment), such as [məneːə] for [bəneːdə], or in a later part (third, fourth or fifth segment), such as [ʋɛs] for [ʋɛtstrɛit] *wedstrijd* 'match'. The critical criterion for a reduced form was that it shared more initial segments with another existing word than with its own canonical form. Note that the amount of variability in our materials is the norm in work on casual speech (see Warner, to appear, for a discussion).

All target words were spoken by Dutch (not Flemish) speakers and were not masked by overlapping (speech) sounds. The words of interest were transcribed separately by two independent raters. Spectrograms were made with the software package PRAAT (Boersma, 2001) to observe the signal in auditory and visual form. The independent transcriptions were compared to verify agreement. In case of disagreement, the transcribers were required to reach consensus. The transcribers again examined the spectrum carefully. Moreover, they listened to the full sentence, parts of the sentence, the target word, and each segment in isolation. Note also that the discrepancies which were encountered were rather minimal. For example, differences were found in where the onset of a segment started.

Note that listeners can hardly recognize reduced word forms on the basis of the acoustic signal for that word alone (e.g., Arai, 1999; Ernestus et al., 2002). Listeners also find it difficult to recognize highly reduced forms in a limited context in which only the adjacent vowels and intervening consonants around the target word are present. Therefore the target forms were presented either in full contexts with several words around the target (e.g., *ook naar beneden, die sluit dan aan* 'also going downwards this connects then to'), or (to reduce the predictability of the target word) in syllable contexts with only the syllables directly neighbouring the target (e.g., *naar beneden die*). Often these single syllables consisted of existing words (e.g., *naar* 'to'). Note that the context for a canonical item always differed from that of a reduced item because they occurred in different natural utterances. We conducted a cloze test (web-based) to investigate whether the different contexts induce preferences for certain word types (i.e., target, "canonical form" competitor, "reduced form" competitor, and distractor), which might have caused confounds in our material. This test measured the predictability of the target word given the preceding context in canonical and reduced sentences. For both types of sentences the words preceding the target were presented on the screen. In the first part, participants (n = 35) had to finish the sentence freely with three to seven words suitable to the context. In the second part, the sentence was again shown on the screen, but now the potential target, the two competitors and the distractor were provided. The participants had to rank these words in order by how likely they were to complete the sentence.

In the first open-ended part of the cloze test, participants named the target word on 5.8% of the trials (5.4% in the reduced form sentences, 6.2% in the canonical form sentences). These results suggest that some target words were indeed somewhat predictable given their linguistic context. The target words were, however, not more predictable in the sentence in which they happened to be reduced. The participants never named a competitor with the exception of one occurrence of a "reduced form" competitor (< 1%).

In the second forced-choice part, participants rated the target word as the most likely option (in 81.6% of the trials). The mean rank of the target word was hence close to 1, and this did not differ between sentences with reduced forms (1.30) and sentences with canonical forms (1.25). To test whether there was a difference in terms of semantic predictability of the "canonical form" competitor and the "reduced form" competitor, we compared the mean rank of both competitors for both types of sentences (i.e., sentences with reduced forms and sentences with canonical forms). The mean rank in all four cases was approximately 3 ("canonical form" competitors: 3.07 in the canonical form sentences, and 2.94 in the reduced form sentences; "reduced form" competitors: 2.94 in the canonical form sentences, and 2.84 in the reduced form sentences). It is hence unsurprising that there were no significant differences as evaluated with a two-by-two repeated measures ANOVA with competitor and sentence as predictors ($F_{Sentence}(1, 30) = 1.68$, $p > 0.1$, all other $F$s < 1).

During the experiment the computer screen displayed four different word types: the target word (e.g., *beneden* 'downwards'), a phonologically unrelated distractor (e.g., *vakantie* 'holiday') and two types of competitors (see Figure 3). A "canonical form" competitor shared more initial segments with the canonical form than with the reduced form (e.g, *benadelen* 'to disadvantage' [bənaːdeːlə] for [bəneːdə]), whereas a "reduced form" competitor shared more initial segments with the reduced form than with the canonical form (e.g., *meneer* 'mister' [məneːr] for [məneːə]). As a consequence, the display always contained two to three phonologically related words, of which one was the target. To mask this pattern, we used filler trials. On filler trials, displays also contained four printed words of which two to three were phonologically related. For half of the filler trials, the target appearing in the auditory sentence, however, was not one of the set of phonologically related words on the screen, but rather was the unrelated word. These fillers were included to prevent participants from developing any expectation that items sharing certain phonological attributes would be mentioned. Fillers were also included to prevent listeners from predicting the upcoming target word due to repetition of visual displays. The visual displays of the fillers were, as the experimental items, repeated. For example, the same visual four-word grid (e.g., *familie* 'family', *seizoen* 'season', *strijden* 'to fight', *strijdlustig* 'quarrelsome') was displayed when listeners heard the target word *familie* and when they heard the target word *seizoen*.


Figure 3: *Example of a printed-word display presented to participants. The spoken target word in this example was* beneden *'downwards'. The four printed words are the target (*beneden *'downwards'), a distractor (*vakantie *'holiday'), a "canonical form" competitor (*benadelen *'to disadvantage'), and a "reduced form" competitor (*meneer *'mister').*

We created two different item lists. Both lists had half of the canonical forms and half of the reduced forms with full contexts and the other half with syllabic contexts. Each subject received one list. The trials in each list were randomized, so that each subject received a different order of presentation. Besides sixty-four fillers (16 fillers in each condition) we also selected 12 practice trials from the spontaneous subcorpora. The positions of the four printed words on the screen were randomized for each participant.

### Procedure

Participants were tested individually, seated at a comfortable viewing distance from the computer screen. The eye-tracking system was mounted and calibrated (an SMI EyelinkII system, sampling at 250 Hz). The auditory stimuli were presented over headphones using the NESU software.

Participants received written instructions on the screen. They had to click on the printed word in the visual display representing the word they heard, using the computer's mouse. The location of the printed words was randomized over the four quadrants on the screen to avoid cues to the position of the target. On each trial, the four printed words (24pt Courier) were first presented on the centres of the quadrants on the screen. After 2500 ms, the auditory stimulus was presented. Note that the preview time in the current study was much longer than the one used in McQueen and Viebahn's (2007) study. There are two reasons why we chose to use this longer preview time. First, our target sentences are more complex than their target sentences (e.g., *ook naar beneden, die sluit dan aan* 'also going downwards that connects then too' versus *Klik op het woord lotje* 'Click on the word lottery ticket'). Second, the position of the target was unpredictable in our sentences, whereas in McQueen and Viebahn's study it was predictable, i.e. the target word always followed after the sentence frame 'Click on the word'. When participants clicked with the mouse on a word, they initiated the next trial. After every five trials, a central fixation cross appeared centred on the screen. Participants were instructed to look at it, so that the experimenter could correct drifts in the calibration of the eye tracker. Each participant first completed the 12 practice trials. Subsequently, we presented the 64 experimental and 64 filler trials (the two lists described above). The experimental session took 20 minutes.

## Design and analysis

For the click responses, we calculated the percentage of correct identifications. The response times on the correct detections were measured from target word offset instead of onset because of the durational differences between the canonical and the reduced form of the same target word. Canonical forms were always longer in duration than reduced forms. The response times would be confounded if we had measured from target onset. A statistical analysis of the error pattern and the response times was carried out with linear mixed effects models. A logistic linking function was used for the error patterns (cf. Dixon, 2008).

For the eye-tracking data, we analyzed only those trials for which the participants clicked on the correct target. We analyzed the data from the right eye of the participants and discarded blinks and saccades. It is estimated that an eye movement is typically programmed about 200 ms before it is launched (Matin, Shao, & Boff, 1993). Thus eye fixations before 200 ms after target onset are unlikely to be driven by acoustic information from the target word. Following Allopenna et al. (1998) and McQueen and Viebahn (2007) we choose to analyze proportion of fixations during the 200-800 ms time window after the acoustic onset of the target word. For all four Word Types (i.e., target, "canonical form" competitor, "reduced form" competitor, and distractor) we allowed a deviation of 100 pixels in height and 150 pixels in width around the centre of each printed word in the visual display. The screen resolution was 1024 * 768 pixels.

For the analysis we first transformed the proportion data with the empirical logit function (Barr, 2008; see formula (6), p. 14) because proportions are problematic in any statistical technique that assumes a linear relation between predictor and outcome variables. From these data, we created three linearly independent measures: 1) looks to the target, to investigate the ease of recognition; 2) mean of looks to both competitors vs. looks to the distractor, to assess the existence and strength of overall competition effects; and 3) looks to the "canonical form" competitor vs. looks to the "reduced form" competitor, to test for the specificity of the competition effects.

Note that the latter two are difference measures, so that a difference from zero indicates a preference for one type of stimulus.[2]

We tested whether these measures were influenced by Word Form (i.e., canonical versus reduced forms) and Context (i.e., full versus syllable context) using linear mixed effects models (Baayen, Davidson, & Bates, 2008), with participants and items as random effects. This technique is designed to overcome the language-as-fixed-effect problem (Clark, 1973). As Baayen et al. show, the LMER technique is more powerful without producing more false positives. Word Form and Context were coded as numeric contrasts (-0.5 and 0.5, cf. Barr, 2008). We estimated *p*-values by using Markov Chain Monte Carlo simulations (Baayen et al., 2008). Canonical forms and full context were coded as -0.5, whereas reduced forms and syllable context were coded as 0.5. Thus, we contrasted four conditions: 1) Canonical forms in full contexts, 2) Canonical forms in syllable contexts, 3) Reduced forms in full contexts, and 4) Reduced forms in syllable contexts. A negative beta indicates that the dependent variable has a higher value for the canonical forms and for the full context condition whereas a positive beta indicates that the dependent variable has a higher value for the reduced forms and for the syllable condition. Note that the interpretation depends on the dependent measure. In the case of the response time measure, a positive beta would mean longer response times for reduced forms and for the syllable condition—and hence that these conditions are more difficult, while for target fixation proportion, a positive beta indicates for the Word Form factor that the target is more

[2] Clearly, other contrasts may be of interest, too. For instance, if the competitors are different overall from the distractors, one might wonder if this difference could be driven by one of the competitors. One might then compare each competitor individually with the distractor. There are two reasons not to do this. First, this would generate linearly-dependent contrasts and the necessary correction of the statistical tests would reduce the statistical power. Secondly, if only one of the competitors gives rise to competition effects, this should lead to a significant difference between the two competitors. Hence, with the two contrasts—competitors versus distractor and "canonical form" competitor versus "reduced form" competitor—we ascertain whether there are measurable competition effects at all, and whether they are mainly carried by one of the competitors.

often fixated in the reduced form condition than in the canonical form condition. For the Context factor, a positive beta indicates more target fixation in the syllable condition than in the full context condition. Effects must be interpreted in opposite directions: Greater fixation represents better recognition, but greater response time represents more difficult recognition.

## RESULTS

### Accuracy and response time measures

Table 1 displays the percentages of mouse-click responses to the different word types and the average response times per condition. The error analysis showed that participants provided significantly more correct responses for the canonical forms than for the reduced forms ($\beta_{Word\ Form} = -5.91$, $p < 0.01$) as indicated by the negative beta. We found no other main or interaction effects (all *p*'s > 0.1).

The analysis of the response time data (measured from target word offset) showed that listeners took significantly more time to recognize reduced versus canonical targets ($\beta_{Word\ Form} = 254.5$, $p = 0.0001$), which is indicated by the positive beta. There were no other main or interaction effects found for this measure (all *p*'s > 0.1).

Table 1: *Task performance in Experiment 1.*

| | Canonical forms | | Reduced forms | |
|---|---|---|---|---|
| % Click-responses | Full context | Syllable context | Full context | Syllable context |
| Target | 99.75 | 99.75 | 92.50 | 81.00 |
| Canonical comp | 0 | 0.25 | 3.00 | 2.50 |
| Reduced comp | 0.25 | 0 | 4.50 | 15.75 |
| Distractor | 0 | 0 | 0 | 0.50 |
| RT in ms | 977 (467) | 974 (368) | 1213 (515) | 1193 (464) |

*Standard deviations appear in parentheses.*

### Eye movements

Figure 4 presents the proportion of fixations over time for all four conditions from acoustic target onset (0 ms) to 1200 ms thereafter. In the 200-800 ms time window we tested the effects of condition on three linearly independent measures: looks to targets (i.e., ease of recognition), looks to competitors versus distractor (i.e., overall competition), and looks to "canonical form" competitor versus "reduced form" competitor (i.e., specific competition). We first analyzed whether looks to targets differed by condition. We found a main effect of Word Form ($\beta_{Word\ Form}$ = -1.23, $p_{MCMC}$ < 0.001). The negative beta reveals that targets attracted more looks in the canonical form condition than in the reduced form condition. Further, we found a main effect of Context ($\beta_{Context}$ = -0.56, $p_{MCMC}$ < 0.001). The negative beta reveals that targets attracted more looks in full contexts than in syllable contexts. The analysis also revealed an interaction effect of Word Form by Context ($\beta_{Word\ Form \times Context}$ = 0.98, $p_{MCMC}$ < 0.05). This interaction shows that the context effect is larger for canonical forms than for reduced forms.

We also analyzed whether the two competitors attracted more looks than the distractor. This analysis (competitors - distractor) showed an effect of overall competition ($\beta_{Intercept}$ = 0.28, $p_{MCMC}$ < 0.01), independent of Word Form and Context (all $p_{MCMC} > 0.1$). We found no interaction between Word Form and Context ($\beta_{Word\ Form\ x\ Context} = 0.37$, $p_{MCMC} > 0.1$).

Finally, a comparison between looks to the "canonical form" competitor versus the "reduced form" competitor ("canonical form" competitor – "reduced form" competitor) showed that the mean difference between looks to the "canonical form" competitor and the "reduced form" competitor was not larger than zero, i.e. the competitors did not differ from each other ($\beta_{Intercept} = 0.17$, $p_{MCMC} > 0.1$), and this pattern was not influenced by the phonetic form of the input ($\beta_{Word\ Form} = 0.01$, $p_{MCMC} > 0.1$) or by context ($\beta_{Word\ Form} = 0.18$, $p_{MCMC} > 0.1$). These two factors did not interact with each other ($\beta_{Word\ Form\ x\ Context} = -0.14$, $p_{MCMC} > 0.1$).

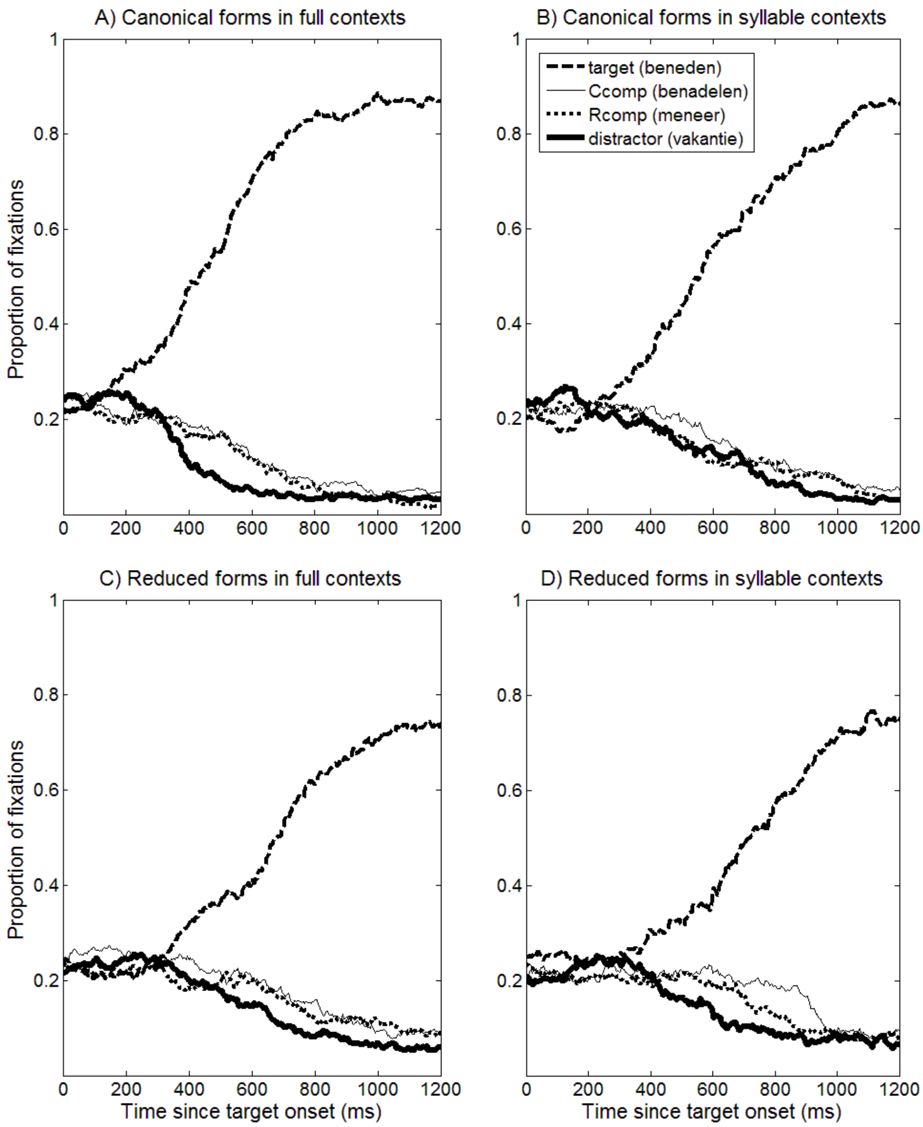



Figure 4: *Proportion of fixations to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor, in (A) Canonical forms presented in full contexts* ook naar beneden, die sluit dan aan *'also going downwards this connects then to', (B) Canonical forms presented in syllable contexts* naar beneden die *'going downwards this', (C) Reduced forms presented in full contexts* buigt het zo af en dan valt het naar beneden, dat is echt *'it bends like this and then it falls down, that is really', and (D) Reduced forms presented in syllable contexts* naar beneden dat *'going downwards that'.*

# DISCUSSION

The accuracy data show that it is harder to recognize reduced forms than canonical forms and that listeners benefit from more linguistic context. Similarly, the response time data reveal that listeners need more time to recognize reduced forms than canonical forms. The eye-movement data also support the conclusions drawn from the offline data. Listeners looked more often to targets in the canonical than in the reduced conditions. All of these findings replicate earlier findings that listeners find it difficult to recognize reduced forms on the basis of the acoustic signal alone (cf. Arai, 1999; Ernestus et al., 2002; Kemps et al., 2004).

More interestingly, our eye-movement data suggest that differences in the exact phonetic form of the acoustic input have no detectable influence on phonological competition. While we anticipated that the phonetic form of the input might not influence the pattern of competition, we had at least expected to replicate the pattern found in other eye-tracking studies (Allopenna et al., 1998; McQueen & Viebahn, 2007), with a preference for onset overlap competitors. In the current case, we therefore had expected that the "canonical form" competitor would attract more overt attention than the "reduced form" competitor, at least when the target word was pronounced canonically. The data show, however, that the "canonical form" competitor attracted as much overt attention as the "reduced form" competitor when the target word was pronounced canonically (i.e., when hearing *beneden* participants directed as much attention to the "canonical form" competitor *benadelen* as to the "reduced form" competitor *meneer*). This finding contrasts with the results from laboratory-speech research that candidates with initial phonological overlap with the target word compete more strongly than candidates with medial or final overlap (e.g., Allopenna et al., 1998; McQueen & Viebahn, 2007). Why do our results using spontaneous speech differ from the results predicted based on laboratory speech?

One possibility is that the style of speech changes listeners' tolerance for mismatch. If listeners are confronted with casual speech (such as the corpus speech in our experiment), they may be more tolerant of acoustic mismatches in the speech signal. As discussed in the Introduction, previous research has interpreted listeners' preference for competitors with an onset overlap over competitors with an offset overlap as evidence for intolerance to acoustic mismatch. It is conceivable however that listeners are more tolerant of such mismatches when the speech style indicates that reductions are possible. In such a listening situation *overall* match between the input and the candidate words may be the prime influence on phonological competition rather than the amount of *onset* overlap.

Can such an assumption explain why the "canonical form" and "reduced form" competitors in Experiment 1 attracted similar levels of attention? In first instance, it seems surprising that the "reduced form" competitor was as active as the "canonical form" competitor when the target form was pronounced canonically. An analysis of whether the "reduced form" and the "canonical form" competitors differ with respect to their total segmental overlap with the target forms was therefore performed. The overlap of number of phonemes between the "reduced form" competitors and their target forms was first calculated. This analysis took the segmental order into account, but did not require an exact match of the position. For example, the "reduced form" competitor *persoon* 'person' [pərsoːn] - matching the reduced form [pəsiːpə] - shares 3 out of 6 phonemes with its target form *principe* 'principle' [prɪnsiːpə]. The shared phonemes between the "reduced form" competitor and the target form are [p], [r], and [s], which appear in the same order in both words. If the order of the phonemes were not taken into account, the segment [n] as well as the schwa would also have been included in this calculation. The number of matching phonemes was then divided by the total number of phonemes of the "reduced form" competitor. Similar comparisons were made between the "canonical form" competitors and their target forms. A t-test showed no differences in segmental overlap between the overlap values for the "reduced form" and the "canonical form" competitors ($t(62) = -0.18$, $p > 0.1$ ). Thus this result is consistent with the notion that overall match between input and candidate words rather than onset overlap is of prime importance when listening to casual speech. This could then explain why there was no difference in looks to the two types of competitors.

A second possibility is that the results of experiment 1 reflect a lack of power. There are two factors that may have reduced experimental power. First, the cloze test showed that the target words are to some extent predictable. As we needed valid examples of strong reduction, we were forced to use sentences from a speech corpus. It was hence not possible to prevent some predictability of the target word. There is however evidence that contextual predictability can constrain lexical activation (e.g., Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Tabossi & Zardon, 1993). A second potential problematic issue is that our manipulation of the "canonical form" versus "reduced form" competitor is less strong than the manipulation of onset vs. offset overlap in previous experiments (Allopenna et al, 1998; McQueen & Viebahn, 2007). Three quarters of our pairs of "canonical form" and "reduced form" competitors both shared the initial segment with the target, and often the difference in amount of onset overlap was small. This may also make a difference between the two types of competitors less likely. Experiment 2 was designed to test these two possible explanations for the results of Experiment 1.

## EXPERIMENT 2

Listeners were again presented with the canonical forms in the full context condition of Experiment 1 (henceforth, canonical forms in casual speech condition). Note that this condition is identical to the canonical form-full context condition of Experiment 1. For a second condition, we re-recorded these same spontaneous sentences under laboratory conditions such that all (target) words were carefully pronounced (henceforth, canonical forms in laboratory speech condition). These conditions hence differ neither in amount of reduction on the target words—the target word is always fully pronounced—nor in the predictability of the target word in the sentences—the sentences were after all identical. The only difference between the two conditions is the speech style. Note that the factor Context, which was included in Experiment 1, is not included in this experiment.

Importantly, the experiment was blocked by speech style. The laboratory speech condition was presented before the casual speech condition. These conditions enable us to distinguish the two accounts for the results of Experiment 1. According to the first account listeners are more tolerant of acoustic mismatch when they hear casual speech (reducing the preference for the "canonical form" competitor). If this account is correct the "canonical form" competitor should attract more overt attention than the "reduced form" competitor in the laboratory speech condition but not in the casual speech condition. According to the second account of the data in Experiment 1, the lack of a preference for the "canonical form" competitors in Experiment 1 was due to lack of power (because of target predictability and/or lack of sufficient difference in onset overlap between "canonical form" and "reduced form" competitors). If this account is correct both conditions should replicate the finding of Experiment 1: Competition effects should be as strong for the "canonical form" as for the "reduced form" competitors.

# METHOD

## Participants

Twenty-six native Dutch speakers from the Max Planck Institute's subject pool participated in this experiment. They reported normal hearing and vision and were paid for their participation. None of them participated in Experiment 1.

## Materials and procedure

We used the same 32 sentences of the Canonical forms in full context condition of Experiment 1 for the casual speech condition of Experiment 2. For the laboratory speech condition we re-recorded these sentences in the laboratory. To do this, the casual speech sentences were orthographically transcribed. We took typical casual speech characteristics like hesitations (e.g., uh) and repetitions out of the sentences to make them clearer and to make it easier for the speaker to pronounce the target words fully. A female native speaker of Dutch was asked to read the sentences carefully out loud while being recorded in a sound-attenuated booth. Her speech was recorded directly to a computer (sampling rate at 44.1 kHz). The speaker was naive to the purposes of the experiment and did not hear the casual speech sentences beforehand, so she was unable to mimic the speech rate, prosody, or intonation of the original sentences.

We used the same procedure as in Experiment 1. Each participant listened to half of the laboratory and half of the casual speech sentences, counterbalancing this assignment across participants. Trials were blocked by speech style (i.e., laboratory versus casual speech). The casual speech block immediately followed after the laboratory speech block. Before each block participants completed 3 practice trials. Next the 16 experimental and 16 filler trials were presented. Order of presentation within each block was randomized. The total duration of the experimental session was 10 minutes.

### Design and analysis

We examined whether the results were influenced by Speech Style (i.e., laboratory versus casual speech), using linear mixed effects models with participants and items as random effects (Baayen et al., 2008). Speech style was coded as a numerical contrast (-0.5 and 0.5, cf. Barr, 2008) in which laboratory speech was coded as -0.5 and casual speech as 0.5. We used the same measures as in Experiment 1 (i.e., errors, response times, target activation, overall competition and specific competition) and we analyzed proportion of fixations during the 200-800 ms time window after the acoustic onset of the target word (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007).

## RESULTS

### Accuracy and response time measures

Table 2 (left column) shows the error rates and the average response times per Speech Style. Listeners made no errors. The reaction time analysis (measured from target word offset) revealed that listeners clicked faster on canonical targets in the laboratory speech condition than in the casual speech condition ($\beta_{Speech\ Style} = 59.52$, $p < 0.05$). This can be explained by the significantly longer word durations for the canonical forms in the laboratory speech condition ($M = 601$, $SD = 88$) than for the canonical forms in the casual speech condition ($M = 489$, $SD = 105$; $\beta_{Speech\ Style} = -110.1$, $p < 0.0001$).

Table 2: *Task performance in Experiment 2 and 3.*

| | Experiment 2 | | Experiment 3 | |
|---|---|---|---|---|
| | Canonical forms | | Casual speech | |
| | Lab speech | Casual speech | Canonical forms | Reduced forms |
| % Correct | 100 | 100 | 99 | 93 |
| RT in ms | 906 (536) | 975 (475) | 1008 (430) | 1192 (479) |

*Standard deviations appear in parentheses.*

## Eye movements

Figure 5 presents the proportion of fixations from acoustic target onset (0 ms) to 1200 ms thereafter for (A) Canonical forms in laboratory speech and (B) Canonical forms in casual speech. We analyzed whether looks to targets differed between the two conditions. The analysis showed no difference in target looks between the conditions ($\beta_{Speech\ Style} = -0.32$, $p_{MCMC} > 0.1$).

An analysis of whether listeners looked more often to the competitors than to the distractor showed an effect of overall competition ($\beta_{Intercept} = 0.31$, $p_{MCMC} < 0.01$). The significant intercept indicates that the mean difference between looks to competitors and distractor is larger than zero, and hence that the competitors attracted more looks than the distractors. No difference was found between the laboratory speech and the casual speech condition ($p_{MCMC} > 0.05$).

Finally, a comparison between looks to the competitors ("canonical form" competitor – "reduced form" competitor) revealed that the "canonical form" competitor attracted more looks than the "reduced form" competitor ($\beta_{Intercept} = 0.53$, $p_{MCMC} < 0.001$). The significant intercept shows that the mean difference between looks to the "canonical form" competitor and the "reduced form" competitor is larger than zero, indicating that the "canonical form" competitor is more strongly activated than the "reduced form" competitor. No main effect was found for Speech Style ($p_{MCMC} > 0.1$).

In sum, the data of Experiment 2 are very clear. We observed a significant preference for the "canonical form" competitor over the "reduced form" competitor in both the laboratory and the casual speech condition. There were no differences on any other measure between the two conditions of Experiment 2.

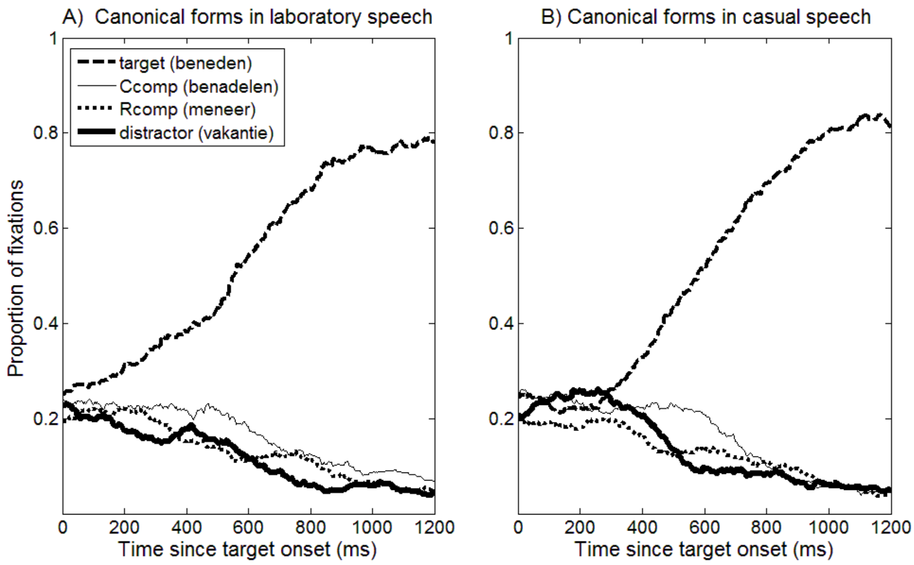



Figure 5: *Proportion of fixations to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor, in (A) Canonical forms in laboratory speech, and (B) Canonical forms in casual speech for the sentence* ook naar <u>beneden</u>, die sluit dan aan *'also going <u>downwards</u>, this connects then to'.*

# DISCUSSION

The data of Experiment 2 reveal a preference for the "canonical form" competitor in both the casual speech condition and the laboratory speech condition. These results are both expected and unexpected. On one hand, this pattern replicates earlier results showing a preference for onset over offset overlap competitors (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007). As such, it is expected. Therefore it is possible with the present materials—despite predictable targets and relatively small differences in onset overlap between the two types of competitors—to obtain differentiating competition effects. We observed a clear preference for the "canonical form" competitor in both conditions of Experiment 2.

On the other hand, the results are unexpected because we had predicted either no preference in both conditions or a preference for the "canonical form" competitor only in the laboratory speech condition. The latter prediction was driven by the expectation that the casual speech condition of Experiment 2 would replicate the results of the full context-canonical form condition of Experiment 1 (i.e., no preference for the "canonical form" competitor). This expectation was based on the fact that the stimuli in these two conditions were *identical*. The only difference between the two conditions were that, in Experiment 1, reduced forms were presented randomly intermixed with the canonical forms, while in Experiment 2, participants only heard canonical pronunciations.

To ascertain that the difference caused by the experimental situation is real we performed a statistical comparison between the two experiments. We compared the results of the casual speech condition of Experiment 2 with the canonical forms in full context condition of Experiment 1. To reiterate, the stimuli in these two conditions are identical, only the experimental context varies. This cross-experiment analysis examined whether the results were different for the canonical forms in Experiment 1 and Experiment 2, using linear mixed effects models with participants and items as random effects (Baayen et al., 2008). Experiment was coded as a numerical contrast (-0.5 and 0.5, cf. Barr, 2008) in which Experiment 1 was coded as -0.5 and Experiment 2 as 0.5. We used the same measures and analyzed the same time window as in the within-experiment analysis.

The response time analysis revealed no difference between the two identical conditions ($\beta_{Experiment}$ = -2.17, $p_{MCMC} > 0.1$). The analysis of target fixations showed that listeners looked more often to the target words in the canonical forms in full context condition of Experiment 1 than in the casual speech condition of Experiment 2 ($\beta_{Experiment}$ = -0.92, $p_{MCMC} < 0.01$) as indicated by the negative regression weight. There also was an effect of overall competition ($\beta_{Intercept}$ = 0.83, $p_{MCMC} < 0.05$), independent of Experiment ($p_{MCMC} > 0.1$). Importantly, a comparison in strength between the two types of competitors showed a significant difference between the two conditions ($\beta_{Experiment}$ = 0.59, $p_{MCMC} < 0.05$). The positive beta indicates that listeners looked more often to the "canonical form" competitor than the "reduced form" competitor in the casual speech condition of Experiment 2 than in the canonical forms in full context condition of Experiment 1. This is a crucial result. It shows that there is a preference for the "canonical form" competitor in Experiment 2, but not in Experiment 1.

Note that differences for identical conditions in different experimental contexts are not unprecedented (for a classical example, see Van der Heijden, Hagenaar, & Bloem, 1984). To account for such effects, modelling approaches (see Phaf, Van der Heijden, & Hudson, 1990) typically assume that participants adjust their processing strategy to the most difficult condition. Such an interpretation fits well with our data. In Experiment 1, listeners had to deal with reduced forms and, therefore, put less confidence in mismatches between input and canonical form. This led to similar levels of activation for "canonical form" and "reduced form" competitors. In Experiment 2, listeners encountered little reduction and hence took mismatches more seriously, leading to a preference for "canonical form" over "reduced form" competitors.

In order to further confirm that this interpretation of the data is correct, we conducted a final experiment. Experiment 3 was designed to test directly that listeners are more tolerant of acoustic mismatches in a listening situation in which they encounter reduced speech. We again presented the corpus sentences with the canonical forms of the target words (Casual speech condition of Experiment 2), but now intermixed with reduced forms. If our interpretation of the data of Experiment 1 and 2 is correct, we should again observe no preference for the "canonical form" competitor. In other words, the same target words in canonical form, which led to increased eye gaze to “canonical form” competitors when intermixed with clearly spoken sentences in Experiment 2, should produce no such preference when intermixed with casual speech sentences containing reduced forms.

# EXPERIMENT 3

# METHOD

## Participants

Twenty-four native Dutch speakers from the Max Planck Institute’s subject pool participated in this experiment. They reported normal hearing and vision and were paid for their participation. None of them took part in the previous experiments.

## Materials and procedure

We used the same 32 sentences of the Casual speech condition of Experiment 2 (henceforth, Canonical forms in Casual speech) and intermixed these sentences with the Reduced forms in full context condition of Experiment 1 (henceforth, Reduced forms in Casual speech). The same procedure was used as in the previous experiments.

Participants were exposed to either the canonical or the reduced form of each target word. The four-word display thus only appeared once, as in Experiment 2, in the course of the experiment. Note that this presentation is different from Experiment 1 in which the four-word display was presented twice to participants. An anonymous reviewer was concerned that the increased target predictability in Experiment 1 might have reduced participants’ consideration of either competitor, thereby washing out any differences in their consideration of either competitor as a function of the phonetic realization of the target word. Experiment 3 tested this possibility. Another difference between this experiment and Experiment 1 is that the factor Context is not included.

The presentation order of the stimuli in Experiment 3 was randomized. Participants started with 3 practice trials after which the 32 experimental and the 32 filler trials were presented. The total duration of the experimental session lasted 15 minutes.

### Design and analysis

We examined whether the results were influenced by Word Form (i.e., canonical versus reduced), using linear mixed effects models with participants and items as random effects (Baayen et al., 2008). Word Form was coded as a numerical contrast (-0.5 and 0.5, cf. Barr, 2008) in which canonical forms were coded as -0.5 and reduced forms as 0.5. We used the same measures as in the previous experiments (i.e., errors, response times, target activation, overall competition and specific competition) and we analyzed proportion of fixations during the 200-800 ms time window after the acoustic onset of the target word (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007).

## RESULTS

### Accuracy and response time measures

Table 2 (right column) shows the error rate and the average response times per Word Forms. Listeners made more errors in the reduced form condition than in the canonical form condition ($\beta_{Word\ Form}$ = -4.13, $p < 0.05$) as indicated by the negative beta. The reaction time analysis (measured from target word offset) showed that listeners took significantly more time to recognize reduced versus canonical targets ($\beta_{Word\ Form}$ = 211.4, $p < 0.001$).

## Eye movements

Figure 6 presents the proportion of fixations over time for (A) Canonical forms in casual speech and (B) Reduced forms in casual speech from acoustic target onset (0 ms) to 1200 ms thereafter. We first analyzed whether looks to targets differed by conditions. We found a main effect of Word Form ($\beta_{Word\ Form}$ = -0.52, $p_{MCMC} < 0.05$), indicating that listeners looked more often to targets in the canonical form condition than in the reduced form condition.

Second, we analyzed whether there is an effect of overall competition (competitors – distractor). We found an effect of overall competition ($\beta_{Intercept}$ = 0.40, $p_{MCMC} < 0.01$), independent of Word Form ($p_{MCMC} > 0.1$). The significant intercept indicates that the mean difference between looks to competitors and distractor is larger than zero, and hence that the competitors attracted more looks than the distractors.

Finally, and most importantly, we compared listeners' fixations to the competitors ("canonical form" competitor – "reduced form" competitor). The analysis revealed no difference between looks to the "canonical form" competitor and the "reduced form" competitor ($\beta_{Intercept}$ = 0.06, $p_{MCMC} > 0.1$), and this pattern was not influenced by Word Form ($p_{MCMC} > 0.1$).

Experiment 3 thus confirms our interpretation of the results of the first two experiments. The competitors that are activated upon hearing a given word not only depend on the sentential context and the phonetic form of that word. The data from the present experiments are strong evidence that competition processes are also influenced by the amount of reduction the listener encounters in a given listening situation. In addition, the results of Experiment 3 rule out that the absence of a preference for the "canonical form" competitor given a canonical form in the auditory input—as observed in Experiments 1 and 3—was due to repetition of target words. Repetition of target words in the visual display occurred in Experiment 1 but not in

Experiment 3, yet the absence of a preference for the "canonical form" competitor given a canonical form in the input was found in both experiments.

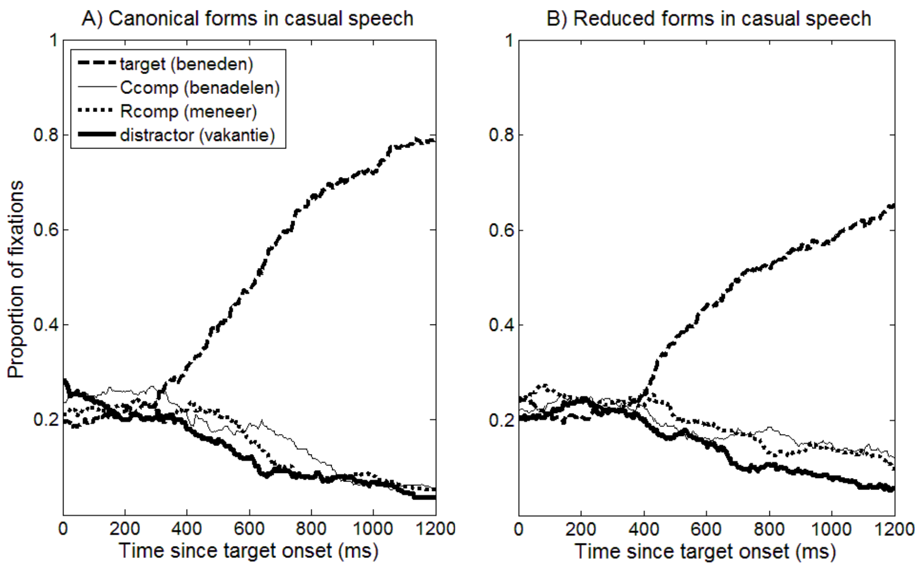



Figure 6: *Proportion of fixations to the target, the "canonical form" competitor, the "reduced form" competitor, and the distractor, in (A) Canonical forms in casual speech* ook naar <u>beneden</u>, die sluit dan aan *'also going <u>downwards</u>, this connects then to', and (B) Reduced forms in casual speech* buigt het zo af en dan valt het naar <u>beneden</u>, dat is echt *'it bends like this and then it falls <u>down</u>, that is really'.*

## GENERAL DISCUSSION

In three eye-tracking experiments we examined whether spoken word recognition in casual speech is different from spoken word recognition in laboratory speech. Participants heard spoken sentences while they saw four printed words in a visual display. Sentences originated from a spontaneous speech corpus or from carefully pronounced laboratory speech. Eye movements were measured while participants listened to sentences containing a critical target word—also presented visually on the screen—which was realized in its canonical or in its reduced form.

Experiment 1 examined whether phonological competition is influenced by the exact phonetic form of the target word (canonical versus reduced). The data showed that on either hearing the reduced realization [pjuːtər] or the canonical realization [kɔmpjuːtər] of *computer*, listeners directed their attention to a similar degree to the same competitors. We interpreted this finding as indicating that when listening to reduced speech, listeners are more tolerant of acoustic mismatches.

Experiment 2 was designed to further investigate this hypothesis. We compared the recognition of canonical forms in laboratory speech with casual speech. Importantly, in contrast to Experiment 1, we did not include any reduced forms in the experiment. We observed that in such a listening situation there was no influence of speech style on competition processes. Listeners directed significantly more overt attention to the "canonical form" competitor than the "reduced form" competitor not only in the laboratory speech condition but also in the casual speech condition of Experiment 2. In the identical condition of Experiment 1 there was no such bias (see Fig. 4a). The only difference between the experiments was that, in Experiment 1, the canonical forms were intermixed with reduced forms, whereas, in Experiment 2, listeners only heard carefully articulated fully pronounced canonical forms. In Experiment 2, participants first listened to a block of laboratory speech before they listened to a block of casual speech. This suggests that participants adjusted to listening to carefully pronounced canonical forms.

The results of Experiment 3 provided further support for the account that speech-intrinsic variation such as reduced speech affects the recognition of clearly articulated words. In Experiment 3, in which the canonical forms of Experiment 2 were intermixed with reduced forms, we replicated the competition pattern of Experiment 1. Once again there was no difference between listeners' fixations to "canonical form" and "reduced form" competitors. This shows that in a listening situation with casual speech which includes a great deal of reduced forms, listeners are more tolerant to acoustic mismatches between input and canonical form. As a consequence, medial and offset overlap competitors become stronger candidates in casual speech (than in a listening situation in which listeners are exposed to carefully articulated fully pronounced speech only) because the initial mismatch is less bad.

It is important to note that we did not compare cohort with rhyme competitors as in the Allopenna et al. (1998) and McQueen and Viebahn (2007) studies. In our materials, some competitors shared onset overlap *both* with the canonical form (e.g., "canonical form" competitor *wetboek* 'statute book' for *wedstrijd* 'match') and the reduced form (e.g., "reduced form" competitor *wesp* 'wasp' for *wedstrijd* 'match'). Importantly however the "reduced form" competitor always deviated from the canonical form by more segments than the "canonical form" competitor (see Appendix). With such an item set, kept constant across all experiments, we found results similar to those of Allopenna et al. and McQueen and Viebahn in the laboratory speech condition and the casual speech condition of Experiment 2. This shows that our weaker manipulation of "canonical form" versus "reduced form" competitors was still able to produce qualitatively similar results relative to the manipulation of onset versus offset competitors in these earlier experiments.

Why do speech reductions change the dynamics of spoken word recognition? Interestingly, previous research on assimilation suggests that listeners are also more tolerant of phonological changes leading to mismatches if the context allows the phonological change. Gaskell and Marslen-Wilson (1996), for example, examined how listeners deal with assimilations (e.g., 'lean bacon' → 'leam bacon'). In a cross-modal priming experiment, they found an effect of priming for unassimilated (e.g., 'lean') and assimilated auditory primes (e.g., 'leam') presented in isolation. A second experiment presented the assimilated tokens in two contexts: a viable context (e.g., 'leam bacon'), allowing for assimilation, or an unviable context (e.g., 'leam gammon'). In the viable context, a priming effect was found for both assimilated and unassimilated primes.

However, in the unviable context, the assimilated primes showed reduced priming effects as compared to unassimilated primes.

Mitterer and Blomert (2003) also investigated how listeners cope with the variation caused by place assimilation in continuous spoken word recognition. Participants had to indicate whether the Dutch word *tuin* 'garden' was pronounced canonically or as [tœym] due to nasal place articulation. These target words were presented in a context which allowed assimilation (*tuinbank* 'garden bench') or in a context that did not (*tuinstoel* 'garden chair'). In the viable-context condition, listeners (incorrectly) perceived the target *tuimbank* as *tuinbank*, (see Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1998; Gow, 2003; Mitterer, Csépe, Honbolygo, & Blomert, 2006, for similar findings). These results suggest that listeners tolerate variation in the input if the context allows the variation.

Our results indicate another form of mismatch tolerance based on speech-intrinsic factors, but on a much larger time-scale. The experiments on assimilation showed that listeners take the immediately following context—in the range of fractions of seconds—into account to license a mismatch between input and canonical form. Our experiments reveal that a general tolerance for mismatch can also be based on the time range of minutes. If participants listen to a mix of canonical and reduced forms embedded in casual speech sentences (as in Experiment 1 and Experiment 3), listeners tolerate onset mismatches to a greater extent than when listeners are first confronted with speech that is carefully produced in a laboratory setting before they listen to casual speech (as in Experiment 2).

The present findings also fit well with recent data about the influence of extrinsic factors on spoken word recognition. Huettig and McQueen (2009) investigated listener flexibility by comparing the dynamics of the spoken word recognition process in clear speech and speech disrupted by radio noise. In their Experiment 1, Dutch participants listened to clearly articulated spoken Dutch sentences which each included a critical word, while their eye movements to four visual objects were measured. There were two critical conditions. In the first, the objects included a cohort competitor (e.g., *parachute*) with the same onset as the critical spoken word (e.g., *paraplu*, 'umbrella') and three unrelated distractors. In the second condition, a rhyme competitor (e.g., *hamer*, 'hammer') of the critical word (e.g., *kamer*, 'room') was present in the display, again with three distractors. Their Experiment 2 was identical to their Experiment 1 except that phonemes in the spoken sentences were replaced with radio-signal noises (as in AM radio listening conditions). Importantly (as in our present study) the critical words (and the immediately surrounding words) were not changed. Huettig and McQueen observed a significant experiment by competitor type interaction. In Experiment 1 (no noise) participants fixated both kinds of competitors more than unrelated distractors, but there were more and earlier looks to cohort competitors than to rhyme competitors (as in the Allopenna study). In Experiment 2 (with radio noise) participants still fixated cohort competitors more than rhyme competitors but the early cohort effect was reduced and the rhyme effect was stronger and occurred earlier.

Their results suggest that speech-extrinsic factors such as AM radio noise also change the dynamics of spoken word recognition. Thus, the well-attested finding of stronger reliance on word onset overlap in speech recognition appears to be due in part to the use of carefully articulated fully pronounced and noise-free speech in most experiments. When onset information becomes less reliable, either because of speech-intrinsic factors such as reduced speech or speech-extrinsic factors such as noise, listeners appear to depend on it less. A core feature of the speech recognition system thus appears to be its flexibility.

We conclude that listening to phonological reduced speech changes the dynamics of spoken word recognition. In such a listening situation, listeners penalize acoustic mismatches less strongly than when listening to fully pronounced laboratory speech. Our data demonstrate that speech-intrinsic variation such as reduced speech influences phonological competition. Flexibility to adjust to speech-intrinsic (and speech-extrinsic) factors is a key feature of the spoken word recognition system.

# PHONOLOGICAL COMPETITION AND THE PROCESSING OF REDUCED FORMS

**CHAPTER 3**

## ABSTRACT

Three experiments examined phonological competition during the recognition of strongly reduced forms such as [pjuːtər] for *computer*, using a target-absent variant of the visual world paradigm, which maximizes the likelihood of observing such competition effects and their time-course. Listeners' eye movements were tracked during presentation of canonical and reduced forms as they looked at displays of four printed words. One of the words was phonologically similar to the canonical pronunciation of the target word, one word was similar to the reduced pronunciation, and two words served as phonologically unrelated distractors. When spoken targets were presented in isolation (Experiment 1) or in sentential contexts (Experiment 2), competition was influenced as a function of the target word form (canonical vs. reduced). When reduced targets were presented in sentential contexts (Experiment 2), listeners first preferentially fixated "reduced form" competitors before shifting their eye gaze to "canonical form" competitors. Experiment 3, in which the original /p/ from [pjutər] was replaced with a "real" onset /p/, showed an effect of cross-splicing in a late time window. These data suggest that speech reductions initially activate competitors which are similar to the phonological surface form of the reduction, but that listeners nevertheless can exploit fine phonetic detail to reconstruct canonical forms from strongly reduced forms.

# INTRODUCTION

The speech we encounter most often in daily life is casual speech. Although there is a growing interest in studying spoken word recognition with this type of speech, it remains a relatively unexplored area. A critical feature of casual speech is that it contains phonological reductions; that is, speakers pronounce words with fewer phonemes than their canonical transcription in a dictionary would prescribe. For example, the four-syllable word 'apparently' /əpʰɛrəntli/ is realized in only two syllables [pʰɛɾ̃ɪ̃] in a corpus of spontaneously spoken English (Johnson, 2004). Reduction processes can thus significantly modify the way words are produced, and consequently, affect how listeners recognize these words. Reductions are also very common. Johnson, for example, found in a corpus of English conversational speech that more than 60% of the words deviated from their citation form by at least one segment, and another 28% of the words deviated even on two or more segments. The listener's challenge is to recognize words in spite of this variability. In the present study, we examine phonological competition during the recognition of strongly reduced forms such as [pjuːtər] from the canonical form [kɔmpjuːtər] *computer*.

Only a few studies have investigated listeners' comprehension of strongly reduced forms in spontaneous conversation. Ernestus, Baayen, and Schreuder (2002) examined how listeners recognize highly reduced forms in Dutch such as [ifɑl] for [ɪn idər χəfɑl] *in ieder geval* 'in any case'. They presented such forms in differing amounts of context. The listeners' task was to write down the form they heard. Results showed that when listeners did not have any supporting context, they hardly recognized the forms. When the forms were presented in a phonetic context, recognition performance increased, but listeners reached ceiling level only when the context was several words long. Their results suggest that highly reduced forms cannot be recognized on the basis of their acoustic forms alone; only when there is a semantic/syntactic context available can one recognize reduced forms correctly (see also Arai, 1999). In a subsequent study, Kemps, Ernestus, Schreuder, and Baayen (2004) provided evidence that listeners unconsciously reconstruct highly reduced forms, i.e., they compute their canonical counterparts. Listeners had to monitor for the phoneme /l/ in highly reduced forms such as [ɛik] from /ɛiχələk/ *eigenlijk* 'actually'. When such forms were embedded in sentence contexts, listeners often incorrectly reported hearing the phoneme /l/. However, listeners did not report /l/ when the reduced forms were presented in isolation.

Together, these two studies suggest that people only recognize strongly reduced forms within a context of several words, and when they do so, they also activate the canonical word forms. The results of Kemps et al. (2004) seem to indicate that reduced forms are linked to the canonical representation in the mental lexicon and not to a more veridical reflection of the actual input, reflecting the acoustic signal itself. It is, however, possible that these results reflect their use of offline tasks. Such tasks require listeners' meta-linguistic judgments, which are conscious and controlled and thus take time to develop. Listeners only make a decision after the acoustic offset of the target words. Studies using offline tasks are therefore unable to measure whether listeners also actively consider lexical candidates compatible with the acoustic structure of reduced pronunciations early on in the recognition process. For example, listening to the reduced form [pjuːtər] of the canonical form *computer* may activate lexical candidates that sound similar in onset such as 'pupil' and 'pure'.

A useful technique to investigate the *online* processing of strongly reduced forms is visual world eye tracking (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this methodology, listeners' eye movements are measured as they listen to speech and see pictures of objects on a computer screen. The timing and proportion of fixations to pictures of objects reveal which lexical candidates the listener is entertaining as speech unfolds over time. Allopenna, Magnuson, and Tanenhaus (1998), for instance, showed that listeners fixate pictures with names similar to the target name more often than phonologically unrelated names. In Allopenna et al.'s study, participants saw four pictures on a computer screen (e.g., a 'beaker', a 'beetle', a 'speaker', and a 'carriage') and listened to spoken instructions such as 'Pick up the beaker'. Participants looked more at the pictures of both types of competitors than to unrelated pictures, but this was more pronounced for pictures of onset-match competitors (e.g., the 'beetle') than for pictures of offset-match competitors (e.g., the 'speaker'). Recently, a printed-word version of the visual world paradigm has been developed (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Using this variant of the eye-tracking method, McQueen and Viebahn replicated the phonological effects found by Allopenna et al. (1998): participants looked more often to onset-matching (e.g., *buffer* for *buffel* 'buffalo') than to offset-matching competitors (e.g., *lotje* 'lottery ticket' for *rotje* 'fire-cracker').

In a recent study, Brouwer, Mitterer, and Huettig (under revision, see Chapter 2 in this thesis) used the printed-word version of the visual world paradigm to examine whether spoken word recognition in casual conversational speech with many speech reductions differs from carefully articulated laboratory speech (as used in many psycholinguistic experiments). Following the example of Ernestus and colleagues (2002; 2004), the stimulus material was compiled from a spontaneous speech corpus. Whereas Ernestus and colleagues used an offline task to study the effect of spontaneous speech on word recognition, Brouwer and colleagues used an online task to tap directly into the time course of processing. In Brouwer et al., they compared the recognition of reduced and canonical forms of mid-to-high frequency content words in a four-word display of which one of the words was the target word. They constructed "canonical form" competitors (e.g., [kɔmpɛnjən] 'companion' for [kɔmpjuːtər]), which phonologically overlapped more at onset with the canonical form than with the reduced form of the spoken word; and "reduced form" competitors (e.g., [pjuːpɪl] 'pupil' for [pjuːtər]), which phonologically overlapped more at onset with the reduced form than with the canonical form of the spoken word. In Brouwer et al.'s Experiment 1, listeners directed their attention to a similar degree to both competitors on either hearing the reduced realization [pjuːtər] or the canonical realization [kɔmpjuːtər] of *computer*. In their Experiment 2, reduced forms were not included in the experiment. Instead, the recognition of canonical forms in laboratory speech was compared with the recognition of canonical forms in casual speech. Here, listeners directed significantly more overt attention to the "canonical form" competitor than to the "reduced form" competitor in both the laboratory speech condition and in the casual speech condition. In Experiment 3, they intermixed the canonical forms in casual speech of Experiment 2 with reduced forms, and found (as in Experiment 1) that there was no difference between listeners' fixations to "canonical form" and "reduced form" competitors. They concluded that during casual speech, which includes a great deal of reduced word forms, listeners are more tolerant of acoustic mismatches between input and canonical form. These data therefore showed that speech-intrinsic variation (e.g., the overall reliability and quality of the phonetic input) can influence phonological competition.

Nevertheless, the absence of a preference for "canonical form" competitors (e.g., *benadelen*) over "reduced form" competitors (e.g., *meneer*) in the study by Brouwer et al. (under revision) is quite surprising because visual world paradigm studies using carefully-articulated laboratory speech have found very strong effects of onset overlap (Allopenna et al., 1998; McQueen & Viebahn, 2007). There are two possible explanations why such a result was not observed in the Brouwer et al. study. One possibility is that the task situation could affect phonologically-mediated eye gaze. In the visual world paradigm, visual and auditory information jointly determine attention and eye gaze (see Huettig, Rommers, & Meyer, submitted, for extensive review). Thus, listeners' eye gaze not only reflects the processing of the spoken input, but is also affected by the processing of the stimuli in the visual display. Huettig and Altmann (in press) have shown that the properties of all the (partly) matching objects in the display affect the magnitude and timing of eye gaze. In the Brouwer et al. study, there were three items in the display for which there was some phonological overlap with the spoken word (the target; the "canonical form" competitor; and the "reduced form" competitor). In other words, the combination of three at least (partly) matching items in the display with the great number of speech reductions in the spoken stimuli may have created a task situation in which listeners are more tolerant of phonological mismatches.

The influence of the task situation could be reduced by using a target-absent version of the visual world paradigm (Huettig & Altmann, 2005). In this version of the paradigm, a fully-matching target word referent is excluded from the visual display (e.g., hearing *computer* and not seeing a 'computer' on the screen), which greatly increases the magnitude of competition effects for related words (cf. Figure 2 in Huettig & Altmann, 2005). The question then becomes whether there are task situations in which words (e.g., *pupil*) matching the phonological surface form of the speech reduction (e.g., [pjuːtər]) compete more than words matching the phonological surface form of the canonical form (e.g., *companion*) of the target word.

There is, however, another explanation why the input form may not have influenced competitor activations in Brouwer et al. (under revision). It could be that the reduced segments in the reduced forms (e.g., [pjuːtər] for *computer*) carried fine phonetic detail that indicate that the [p] is not word-initial, but that the word starts with a weak syllable starting with /k/. In fact, Brouwer et al. observed for the reduced form [pjuːtər] for *computer* that the closure duration for /p/ was rather long (> 100 ms) for connected speech. This may signal that a weak syllable was literally "swallowed" in this closure; consequently, /p/-initial words would not serve as strong competitors for either reduced or canonical word forms because neither carries sufficient evidence for a p-initial word. Note that, given this explanation, changing the task situation should not influence competitor activations.

In the present study, we therefore investigated whether phonological competition for reduced words in casual speech can ever be influenced by the exact phonetic form of the spoken word. We used the target-absent version of the visual world paradigm described above to maximize the likelihood of observing competitor effects and their time course. The visual display in the current study therefore had the following structure: given the target word *computer* (in either canonical or reduced form), the visual display contained a "canonical form" competitor (*companion*), a "reduced form" competitor (*pupil*), and two phonologically unrelated distractors (*jewel*; *holiday*). Note that the actual items were in Dutch and the same target words were used as in Chapter 2 of this thesis. The visual display, however, was different from the one used in Chapter 2 of this thesis. We examined when and to what extent "reduced form" and "canonical form" competitors play a role in the online recognition of naturally reduced words. If we observe that the overt attention to different competitors is influenced by the input form, this may indicate that reduced forms do not carry sufficient phonetic-detail cues to prevent the activation of words that are similar to the reduced form in the auditory input.

# EXPERIMENT 1

# METHOD

## Participants

Twenty-four participants from the Max Planck Institute's subject pool, undergraduates at the Radboud University Nijmegen, were paid to participate in this experiment. All were native speakers of Dutch and reported normal hearing and normal or corrected-to-normal vision. None of these subjects had participated in any of the experiments reported in Chapter 2.

**Materials**

We selected 32 polysyllabic, mid-to-high frequency content words from the Spoken Dutch Corpus (Oostdijk, 2000). Note that these are the same stimuli as used in Chapter 2 of this thesis. We took both a canonical (e.g., [bənedə] for *beneden* 'downwards') and a reduced realization (e.g., [məneə]) of every target word (see Appendix). Recordings with background noise, overlapping speech, or with unfamiliar dialects such as Flemish were excluded. The target words were transcribed by two independent raters to observe the signal in auditory and visual spectrographic form. The independent transcriptions were compared to verify agreement between the transcriptions. In case of disagreement, the transcribers reached consensus.

The segments of the canonical forms were almost always fully realized, whereas their reduced counterparts were missing one or more segments. The critical criterion for a reduced form to be included in the study was that it shared more initial segments with another existing Dutch word than with its own canonical form. To illustrate this, the reduced form [ʋɛtstrɛi] for the canonical form [ʋɛtstrɛit] *wedstrijd* 'match', in which the final /t/ is deleted, does not live up to the criterion. In this case, no other Dutch word exists that phonologically matches the reduced form [ʋɛtstrɛi] except its own canonical form *wedstrijd*. As a consequence, the reduced form [ʋɛtstrɛi] could not be included in our material. An example of a reduced form that would live up to our criterion is [ʋɛs] for *wedstrijd*. In this case, for example, the Dutch word *wesp* 'wasp' matches phonologically better with the reduced form [ʋɛs] than with the canonical form [ʋɛtstrɛit].

There is considerable variation in the reductions. Reduced forms either differed in the initial part (i.e., first or second segment) such as [məneə] from [bənedə] for *beneden* 'downwards', or in a later part (i.e., third, fourth or fifth segment) such as [ʋɛs] for [ʋɛtstrɛit] *wedstrijd* 'match' from the canonical form. The Appendix lists all target items including their canonical and reduced transcriptions. Note that the degree of variability in our materials is the norm in work on spontaneous speech. Such materials often lead to greater variability in results, but it is necessary in order to study real, spontaneous speech (see Warner, to appear, for a discussion of this trade-off).

For each trial, the computer screen displayed three different word types: a "canonical form" competitor (e.g., *beneden* 'downwards'), a "reduced form" competitor (e.g., *meneer* 'mister') and two phonologically unrelated distractors (e.g., *juweel* 'jewel' and *vakantie* 'holiday'; see Figure 1). Note that in 75% of the cases (24 out of 32, see Appendix) the "canonical form" competitor and the "reduced form" competitor overlap phonologically at onset. Some "reduced form" competitors are therefore also to a certain degree competitors of the target word's citation form. They thus function as offset overlap competitors. However, the "canonical form" competitor always overlapped more at onset with the canonical form than with the reduced form, and the "reduced form" competitor always overlapped more at onset with the reduced form than with the canonical form. For example, the word *wedstrijd* 'match' was realized canonically as [ʋɛtstrɛit] and in a reduced way as [ʋɛs]. The first three segments of the "canonical form" competitor *wetboek* [ʋɛtbuːk] are shared with the canonical form, but only the first two segments are shared with the reduced form. The first three segments of the "reduced form" competitor *wesp* [ʋɛsp], however, are shared with the reduced form, but only the first two segments are shared with the canonical form. It is therefore important to compare the relative strength of the two types of competitors under different conditions. Note again that the fairly subtle differences in which competitor has how much overlap depend heavily on IPA transcription being an adequate way to represent reductions.

The target word mentioned in the casual speech fragments was absent from the visual display (Huettig & Altmann, 2005). The displays of the experimental trials never contained printed target words corresponding fully to the spoken target words. In these cases the participants' task was to click the centre of the screen. To prevent participants from getting less involved in the task we added filler items in which one of the visual words on the screen matched with the auditory target stimulus. We created twice as many filler (128) as experimental items (64). Thus, on average, participants had to click in two out of three cases on one of the four visually presented words on the screen, and in one out of three cases to click in the middle of the screen. Since the display for the experimental trials contained two phonologically related words (i.e., the two competitors), we masked this pattern in the filler items in which a visual target was always present. Participants had to click half of the time on one of the phonologically similar words and half of the time on a word that was phonologically unrelated. In this way, the fillers prevented participants from developing any expectation that items sharing certain phonological attributes would be mentioned.

The experimental and filler items were put into one list and the order was randomized, so that each participant got a different order of presentation. The position of the three types of printed words was randomized over the four quadrants on the screen. That is, the "reduced form" competitor, the "canonical form" competitor and the distractors appeared with equal probability on each of the four screen position over the course of an experimental run. Besides experimental and filler items, we also selected six practice items from the Spoken Dutch Corpus (Oostdijk, 2000) to familiarize participants with the task. Half of the practice items contained a target on the screen, the other half did not.

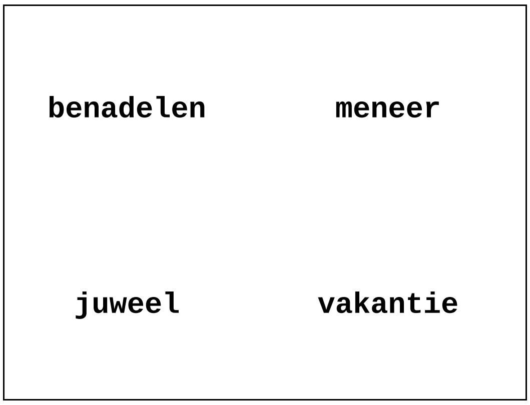



Figure 1: *Example of a printed-word display without a visual target presented to participants (the spoken target word in this example is* beneden *'downwards').*

## Procedure

Participants were tested individually. They were seated in a sound-attenuated booth at a comfortable viewing distance from the computer screen. The eye-tracking system was mounted and calibrated. Eye movements were monitored using an SMI EyeLinkII system, sampling at 250 Hz. Auditory stimuli were presented to the participants over headphones using the NESU software.

Participants received written instructions on the screen. Participants had to click with the computer's mouse on the printed word in the visual display representing the word they heard in the auditory stimulus. If none of the printed words matched with the auditory stimulus - as for all experimental trials - participants had to click the centre of the screen.

Each trial had the following structure. First, a grid with four printed words appeared in a 24-point Courier font on the screen. The centres of the printed words corresponded, independently of the length of the words, to the centres of the quadrants on the screen. After 2500 ms the auditory stimulus was presented. The next trial was initiated after participants clicked with the mouse on the screen. Participants were put under no time pressure to perform this action. Every five trials a central fixation cross appeared centred on the screen, permitting for drift correction in the calibration.

After the six practice trials, the 64 experimental and 128 filler items were presented in random order. The experimental session took approximately 15 minutes.

### Design and analysis

The dependent variables were click-responses and eye movements. For the click-responses we calculated the percentage of correct rejections, i.e. the percentage of clicks in the centre of the screen, and the percentage of incorrect clicks on the three word types. A statistical analysis of the error pattern was carried out with linear mixed effects models using a logistic linking function (cf. Dixon, 2008). Since the auditory target word was never present on the screen for the experimental trials, response times were uninformative, and hence not included in our analyses.

For the eye-tracking data, we discarded blinks and saccades and analyzed the data from the right eye of the participants. Although it is estimated that an eye movement is typically programmed about 200 ms before it is launched (e.g., Matin, Shao, & Boff, 1993), we choose to start analyzing our data from 400 ms onwards. As is apparent from Figure 2, both competitors start to diverge from the averaged distractors around 400 ms and the pattern 'late' in time looks quite differently from the early pattern. We therefore choose to statistically analyze proportion of fixations during two time windows: an early time window (400 to 800 ms) and a late time window (800 to 1200 ms).

For the analysis we transformed the proportion data with the empirical logit function (cf. Barr, 2008). From these data we constructed two linearly independent measures: 1) overall competition effects: mean of looks to both competitors vs. mean of looks to both distractors; and 2) specificity of the competition effects: mean of looks to the "canonical form" competitor vs. mean of looks to the "reduced form" competitor. All measures are difference measures, so that a difference from zero indicates a preference for one type of stimulus.

We tested whether these measures were influenced by Word Form (i.e., canonical forms versus reduced forms) using linear mixed effects models (Baayen, Davidson, & Bates, 2008) with participants and items as random effects and in which Word Form was coded as a numeric contrast (-0.5 and 0.5, cf. Barr, 2008). Canonical forms were coded as -0.5 and reduced forms as 0.5. A negative beta would indicate that the dependent variable has a higher value for the canonical form condition whereas a positive beta would indicate that the dependent variable has a higher value for the reduced form condition. Note that the interpretation of the beta depends on the dependent measure. In the case of the accuracy measure, a positive beta would mean more errors in the reduced form condition than in the canonical form condition, suggesting that recognizing reduced forms is more difficult than recognizing canonical forms. In the case of the overall competition measure, a positive beta would indicate more overall competition in the reduced form condition than in the canonical form condition. A similar interpretation holds for the specific competition measure: a positive beta would imply more specific competition in the reduced form condition than in the canonical form condition. We estimated *p*-values by using Markov Chain Monte Carlo simulations (Baayen et al., 2008).

## RESULTS

Table 1 shows the percentages of correct rejections and of incorrect click responses to the three word types. An analysis on the error rates showed a main effect of Word Form ($\beta_{Word\ Form}$= -4.26, $p < 0.0001$), indicating that listeners made more errors in the reduced form condition than in the canonical form condition. Listeners clicked 31% of the time on the "reduced form" competitor when listening to reduced forms.

Table 1: *Task performance in Experiment 1.*

| | Forms presented in isolation | |
|---|---|---|
| % Click-responses | Canonical forms | Reduced forms |
| Correct rejections | 97.1 | 65.9 |
| "Canonical form" competitor | 1.6 | 2.7 |
| "Reduced form" competitor | 1.3 | 31 |
| Distractors | 0 | 0.4 |

*Note that 'Correct rejections' correspond to clicks in the centre of the screen.*

For the eye-movement data, we plot the fixation proportion of all trials (including the correct and incorrect responses) as well as for correct responses separately. In typical eye-tracking experiments with careful speech, error rates tend to be low (< 5%) and errors are typically discarded. In the current case, however, with more than 30% of errors, simply discarding the errors is problematic because it would exclude the (apparently) most difficult trials with the most severe reductions. Nevertheless, misidentifications obviously lead to prolonged looks at a competitor, simply because the competitor is clicked on. Figure 2 shows therefore one plot for the results for the Canonical forms (all trials; 2A) as well as three different plots for reduced forms: Figure 2B shows the data for all trials; and Figure 2C and 2D show the data for only correct or only incorrect trials, respectively. These additional plots give us insight into how the competition pattern changes depending on participants' performance on a trial. All plots give mean fixation proportion to the two types of competitors and the averaged distractors from acoustic target onset (0 ms) to 1400 ms thereafter. Proportion of fixations was analyzed during an early (400-800 ms) and a late time window (800-1200 ms).

## All trials

We first analyzed whether there was an effect of overall competition (competitors - distractors). We found an overall effect in both time windows (early: $\beta_{Intercept} = 0.87$, $p_{MCMC} < 0.001$; late: $\beta_{Intercept} = 1.17$, $p_{MCMC} < 0.001$). Overall competition was dependent on Word Form. In the early time window, overall competition was strongest in the canonical form condition as indicated by the negative regression weight ($\beta_{Word\ Form} = -0.54$, $p_{MCMC} < 0.001$). In the late time window, however, overall competition was strongest in the reduced form condition as indicated by the positive regression weight ($\beta_{Word\ Form} = 0.97$, $p_{MCMC} < 0.001$).

Second, we analyzed whether listeners looked more often to the "canonical form" competitors than to the "reduced form" competitors ("canonical form" competitor – "reduced form" competitor). We found an overall effect of specific competition in both time windows (early: $\beta_{Intercept} = 0.34$, $p_{MCMC} < 0.01$; late: $\beta_{Intercept} = -0.60$, $p_{MCMC} = 0.05$). This effect was influenced by Word Form in both time windows (early: $\beta_{Word\ Form} = -1.30$, $p_{MCMC} < 0.001$; late: $\beta_{Word\ Form} = -1.68$, $p_{MCMC} < 0.001$). That is, it varied over conditions whether the "canonical form" or "reduced form" competitor received more looks.

To further investigate this pattern, we analyzed the effect of specific competition in each condition separately. This strategy is analogous to the breaking-down of an interaction in factorial ANOVA designs. The analysis showed that the effect of specific competition was significant in the canonical form condition only in the early time window ($\beta_{Intercept} = 0.99$, $p_{MCMC} < 0.001$), indicating more looks to the "canonical form" than the "reduced form" competitor. The specific competition effect was no longer significant in the late time window ($\beta_{Intercept} = 0.25$, $p_{MCMC} > 0.1$). In the reduced form condition we found the opposite pattern: the preference for the "reduced form" competitor was not significant in the early time window ($\beta_{Intercept} = -0.30$, $p_{MCMC} > 0.1$), but it was in the late time window ($\beta_{Intercept} = -1.43$, $p_{MCMC} < 0.05$).

## Correct and incorrect trials

For the correct trials, we first analyzed whether listeners looked more often at the competitors than the distractors (overall competition). We found an overall effect of competition in both time windows (early: $\beta_{Intercept} = 0.83$, $p_{MCMC} < 0.001$; late: $\beta_{Intercept} = 0.78$, $p_{MCMC} < 0.001$), dependent on Word Form in the early time window ($\beta_{Word\ Form} = -0.57$, $p_{MCMC} < 0.01$). In this time window, overall competition was strongest in the canonical form condition as indicated by the negative regression weight.

Second, we analyzed whether there was an effect of specific competition. The analysis showed an overall effect in the early time window ($\beta_{Intercept} = 0.59$, $p_{MCMC} < 0.001$). This effect was influenced by Word Form ($\beta_{Word\ Form} = -0.85$, $p_{MCMC} < 0.01$). We analyzed this effect in each condition separately and found that specific competition was only significant in the canonical form condition ($\beta_{Intercept} = 0.99$, $p_{MCMC} < 0.001$). The specific competition effect was not significant in the reduced form condition ($\beta_{Intercept} = 0.16$, $p_{MCMC} > 0.1$).

These results suggests that even in the trials in which participants made the correct decision they may not (or not always) have recognized the target word. Participants may instead have based their decision on overlap in perceived phonemes without making contact with the lexicon. By contrast, for the incorrect trials, we observe a clear preference for the “reduced form” competitors, suggesting that incorrect trials reveal numerous and sustained fixations to the word type that was chosen (i.e., “reduced form” competitor), as people guide the mouse cursor toward the object they intended to click on.

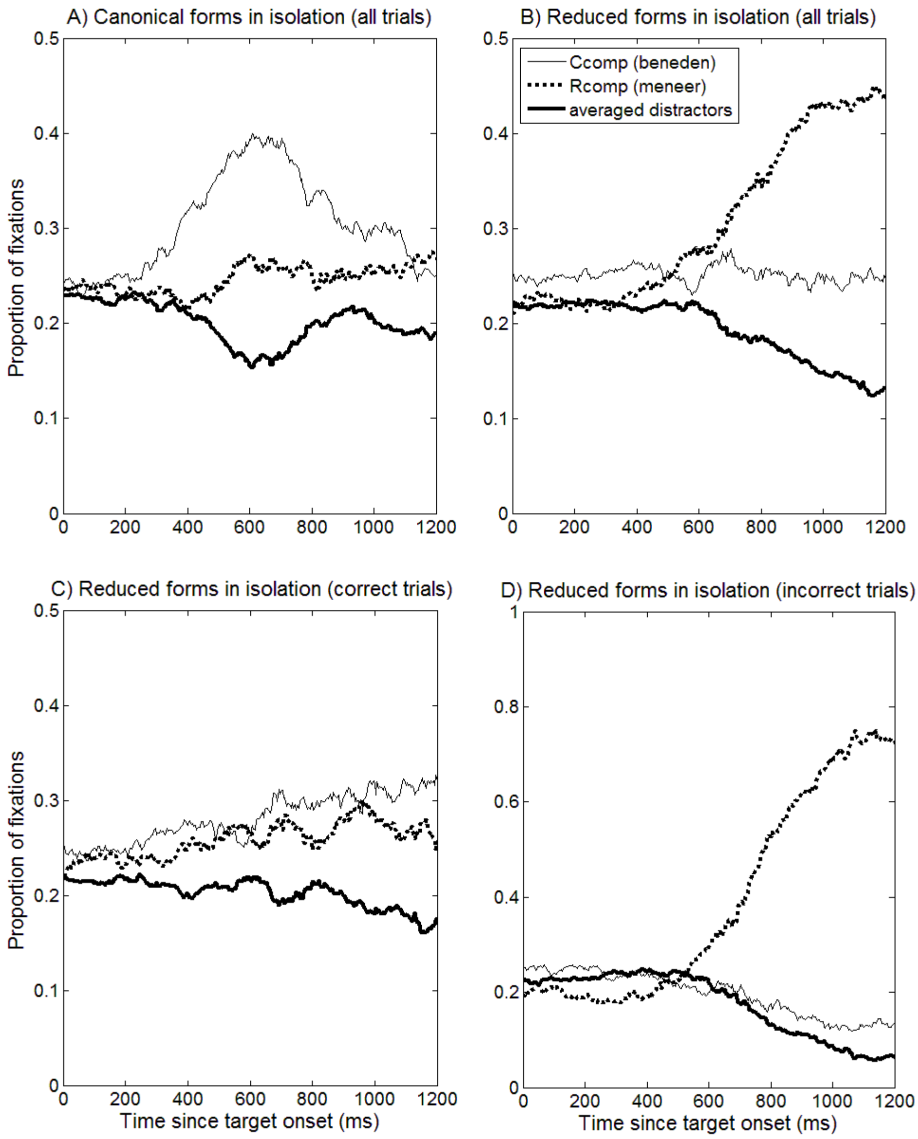



Figure 2: *Proportion of fixations to the "canonical form" competitor (Ccomp), the "reduced form" competitor (Rcomp), and the averaged distractors, in (A) Canonical forms in isolation (all trials); (B) Reduced forms in isolation (all trials); (C) Reduced forms in isolation (correct trials); and (D) Reduced forms in isolation (incorrect trials).*

## DISCUSSION

In Experiment 1, we observed that overt attention to "canonical form" and "reduced form" competitors in a printed-word display can be influenced by the exact form of the acoustic input. This therefore suggests that, at least in task situations such as the present which maximize the likelihood for competition, canonical forms can activate different competitors more strongly than reduced forms. This shows that the previous findings by Brouwer et al. (under revision) were most likely due to the task situation rather than to phonetic detail in the stimuli. There is, however, one alternative interpretation. It is conceivable that by presenting the reduced fragments in isolation, we limited the listeners' ability to exploit the phonetic detail. Consider the example of the reduced form (e.g., [pjuːtər] for *computer*) we used in the Introduction. We had thought that the long closure duration might be a cue that tells listeners that the /p/ is not (underlyingly) word-initial. However, with a single-word presentation, the closure duration is not even audible, making it unlikely that this phonetic detail could influence the competition process. In addition to this, listeners might need surrounding speech rate, or coarticulation with preceding sounds, in order to interpret the reductions.

Moreover, the click-responses showed that listeners made more errors in the reduced form condition than in the canonical form condition. Listeners often clicked on the "reduced form" competitor in the reduced form condition. This offline preference for the "reduced form" competitor is in line with the online eye-movement data. Listeners looked most often at the "reduced form" competitor when listening to reduced forms. These results accord with the findings of Ernestus et al. (2002) and of Arai (1999) who showed that reduced forms are difficult to recognize on the basis of the acoustic form alone.

These two observations pave the way for our subsequent experiments. Given the fact that performance for reduced forms in isolation is seriously compromised

(34% error rate), the question arises how much context is needed for successful recognition. In Experiment 2, we examined phonological competition processes when reduced forms (and canonical forms) are presented in sentential contexts. Recall that Kemps et al. (2004) showed that performance increases with more given context. In our second experiment we therefore examine the phonological competition process during the actual *recognition* of strongly reduced forms in their original context. For canonical forms, we predict the same competition pattern as in Experiment 1: more looks to the "canonical form" competitor than to the "reduced form" competitor (cf. Allopenna et al., 1998; McQueen & Viebahn, 2007). Two outcomes are possible for reduced forms. If listeners in Experiment 1 were simply unable to make use of the phonetic detail in the short fragments, we should now again replicate the pattern observed in Brouwer et al. (under revision): input form does not influence competitor activations. If, however, the absence of the target word better revealed the dynamics of the competition process, we should still observe differences in competitor activation depending on the input form.

## EXPERIMENT 2

## METHOD

### Participants

Twenty-four undergraduates were paid to participate in this experiment. All participants were native speakers of Dutch. They reported no hearing problems and normal or corrected-to-normal vision. None of the listeners participated in the previous experiment or in the experiments of Chapter 2 in this thesis.

### Materials

We used the same 32 canonical and 32 reduced realizations as in Experiment 1, but they were now embedded in a context of several words. Note that the context for a canonical form (e.g., *ook naar beneden, die sluit dan aan* 'the one going downwards, as well, this connects then to') always differed from that of a reduced form (e.g. *buigt het zo af en dan valt het naar beneden, dat is echt* 'it bends like this and then it falls down, that is really') because they occurred in different corpus utterances. We therefore conducted first a cloze test (web-based) to investigate whether the different contexts induced preferential bias for certain word types (i.e., target, "canonical form" competitor, "reduced form" competitor, or for one of the distractors), which might have caused confounds in our material. Note that this is the same cloze test (with the same results) as reported in Chapter 2 of this thesis. This test measured the predictability of the target word given the preceding context in canonical and reduced utterances. For both types of utterances, the words preceding the target were presented on the screen. In the first part, participants (n = 35) were asked to complete each sentence with three to seven words that would fit the context. In the second part the sentence was again shown on the screen, followed by four possible continuations of the sentence: the eventual target, the two types of competitors and one of the distractors. The participants had to rank these words in the order of plausible endings.

In the first open-ended part of the cloze test, participants named the target word on 5.8% of the trials (6.2% in the canonical form sentences, 5.4% in the reduced form sentences). These results suggest that some target words were indeed somewhat predictable given their preceding linguistic context. Crucially, however, the target words were not more predictable in reduced form sentences than in canonical form sentences. The participants never named a phonological competitor, except for one single occurrence of a "reduced form" competitor (< 1%).

In the second part, participants rated the target word as the most likely option for 81.6% of the trials. The mean rank of the target word was hence close to 1. Again, there was no difference between sentences with canonical forms (1.25) or sentences with reduced forms (1.30). We compared the mean rank of both competitors for both types of sentences (i.e., sentences with canonical forms and sentences with reduced forms) to test whether there was a difference in terms of semantic compatibility of the "canonical form" competitor and the "reduced form" competitor. The mean rank in all four cases was around 3 ("canonical form" competitors: 3.07 in the canonical form sentences and 2.94 in the reduced form sentences; "reduced form" competitors: 2.94 in the canonical-form sentences and 2.84 in the reduced form sentences). Statistical analyses confirm this: there were no significant differences as evaluated with to a two-by-two repeated measures ANOVA with competitor type and sentence type as predictors ($F_{SentenceType}(1,30) = 1.67$, $p > 0.1$, the other $F$s > 1).

### Procedure, design, and analysis

The procedure and design were identical to the previous experiment. For the analyses, we used the same two measures as in the previous experiment (i.e., overall competition and specific competition), but we analyzed only the correct trials, i.e. in which subjects clicked in the centre of the screen. The experimental session took approximately 25 minutes.

## RESULTS

Table 2 shows the percentages of correct rejections and of incorrect click responses for the three word types. An analysis on the error rates showed a main effect of Word Form ($\beta_{Word\ Form}$= -2.83, $p < 0.0001$), indicating that listeners performed better in the canonical form (99.5%) than in the reduced form condition (93.8%).

Table 2: *Task performance in Experiment 2.*

| % Click-responses | Forms presented in sentential contexts | |
|---|---|---|
| | Canonical forms | Reduced forms |
| Correct rejections | 99.5 | 93.8 |
| "Canonical form" competitor | 0.1 | 1.8 |
| "Reduced form" competitor | 0.3 | 4.4 |
| Distractors | 0 | 0 |

*Note that 'Correct rejections' correspond to clicks in the centre of the screen.*

For the eye-movement data, we plot and analyze only the fixation proportion of correct trials only for the canonical and reduced forms, since participants hardly made any errors. Figure 3 shows the mean fixation proportion to the two types of competitors and the averaged distractors from acoustic target onset (0 ms) to 1200 ms thereafter for (A) Canonical forms and (B) Reduced forms in sentential contexts. As in the previous experiment, we analyzed fixations during an early (400-800 ms) and a late time window (800-1200 ms).

We first analyzed whether there was an effect of overall competition (competitors - distractors). We found an overall effect in both time windows (early: $\beta_{Intercept} = 0.92$, $p_{MCMC} < 0.001$; late: $\beta_{Intercept} = 0.66$, $p_{MCMC} < 0.001$), independent of Word Form ($p_{MCMC} > 0.1$).

Second, we analyzed whether listeners looked more often to the "canonical form" competitors than to the "reduced form" competitors ("canonical form" competitor - "reduced form" competitor). We found no overall effect of specific competition (all $p_{MCMC}$'s $> 0.1$). Specific competition was, however, influenced by Word Form only in the early time window ($\beta_{Word\ Form} = -1.14$, $p_{MCMC} < 0.001$). The negative beta indicates that specific competition was strongest in the canonical form condition. Importantly, the results from the reduced form condition patterned similarly to the results in the canonical form condition in the late time window ($\beta_{Word\ Form}$ = 0.26, $p_{MCMC}$ > 0.1).

As in the previous experiment, we further analyzed the effect of specific competition in each condition separately. Analyses on the early time window showed that the effect of specific competition was only significant in the canonical form condition ($\beta_{Intercept}$ = 0.80, $p_{MCMC}$ < 0.001). In the reduced form condition we found no effects of specific competition (all $p_{MCMC}$ > 0.1), although there was a preference for the "reduced form" competitor in the early time window and a preference for the "canonical form" competitor in the late time window.

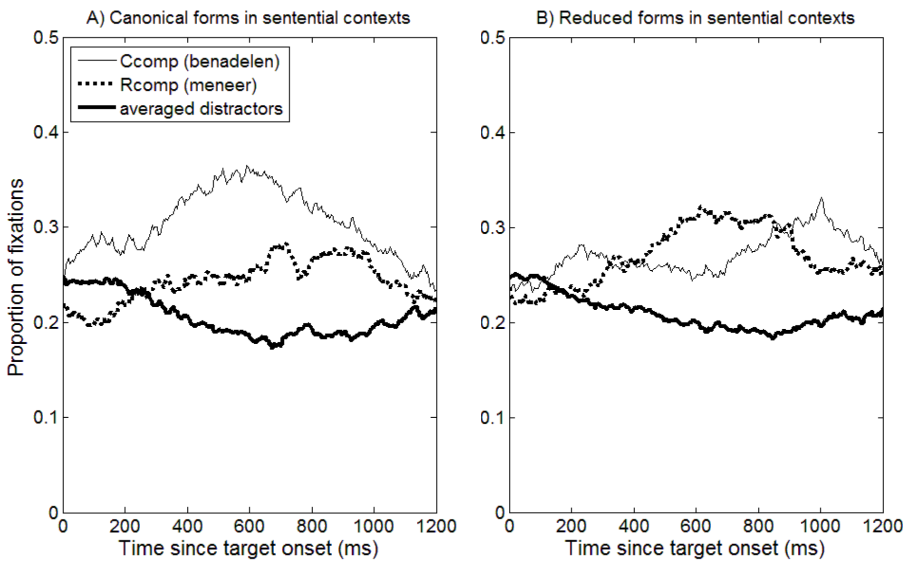



Figure 3: *Proportion of fixations to the "canonical form" competitor (Ccomp), the "reduced form" competitor (Rcomp), and the averaged distractors, in (A) Canonical forms, and (B) Reduced forms for the correct trials presented in sentential contexts.*

# DISCUSSION

Experiment 2 investigated phonological competition when listeners hear strongly reduced forms in sentential contexts. The error pattern showed that reduction still inhibits word recognition (6% error rate), but to a much lesser extent than in Experiment 1 (34% error rate). This result is in line with the findings by Arai (1999) and Ernestus and colleagues (2002): listeners benefit from phonetic and semantic/syntactic context during the recognition of strongly reduced forms.

The eye-movement data revealed a similar competition pattern for canonical forms as in Experiment 1: the "canonical form" competitors attracted more attention than the "reduced form" competitors in the early time window. These results are in line with the competitor effects that have been found by Allopenna et al. (1998) as well as McQueen and Viebahn (2007) because our "canonical form" competitors are similar to the "cohort" competitors in those studies. Over time, as the acoustic form of the "canonical form" competitors became inconsistent with the acoustic input, the preference of looks to the "canonical form" competitors disappeared.

When the input was a reduced form, however, we observed that, in the early time window, the "reduced form" competitors attracted more looks than the "canonical form" competitors. Note, however, that the interaction was significant, but that the specific competition effect did not reach significance in the reduced form condition separately. There was a preference for the "reduced form" competitors even though participants heard a complete dialogue fragment; that is, there was now additional phonetic context that should have enabled participants to better exploit any fine phonetic detail cues. There is a better match of the acoustic signal, (i.e., [məneə]) with the phonological representation activated from the "reduced form" competitors (i.e., *meneer*), than with the "canonical form" competitors (i.e., *benadelen*). This suggests that "reduced form" competitors (i.e., unrelated words overlapping in phonemes with the reduced form of the spoken target word) can compete early during the recognition of reduced forms.

In contrast, later in time “canonical form” competitors received more attention than phonologically unrelated distractors. This result is similar to the offset-matching (or “rhyme”) competitor effects that have been reported by Allopenna et al. (1998) and McQueen & Viebahn (2007). Nevertheless, there is a difference between these studies and our study: the offset-matching competitors in the Allopenna et al. and McQueen and Viebahn’s studies always attracted *less* looks than the onset overlap competitors, which was not the case in our Experiment 2. This suggests that an increase in looks to our “canonical form” competitors (i.e., *benadelen*) reflects more than just overlapping phonemes. It could be that participants reconstruct the canonical (or “citation”) form (i.e., [bənedə]) from its reduced form (i.e., [məneə]. Such a reconstruction process may be time-consuming, which would explain why we only observe a shift in eye gaze to the “canonical form” competitor in the later time window. This also explains why the previous work using offline tasks always only documented a role for the citation form when listeners were confronted with strongly reduced forms. Kemps et al. (2004), for instance, found that listeners judge phonemes that are phonetically absent in reduced forms as present in a phoneme monitoring task. When asked to monitor for /l/, they responded with ‘yes’ to the phonetic form [tyk], which is a reduced form of /natyrlʏk/ *natuurlijk* ‘naturally’. This result is in line with our assumption of a late reconstruction process. The offline task used by Kemps et al. only revealed what happens late in time when the acoustic input form had no longer any influence.

Experiment 2 revealed that “reduced form” competitors competed for eye gaze even when participants had the opportunity to make use of additional phonetic context indicating a speech reduction. Does this mean that there is little phonetic detail to exploit? Experiment 3 tests this directly. We examined whether listeners are sensitive to fine phonetic detail information in interpreting whether a reduced form was heard or not. Previous research showed that listeners are good at exploiting the fine phonetic detail of utterances to recognize intended words even when spontaneous speech processes have changed them so that they deviate from their canonical form. Gow (2002), for example, showed that listeners make use of fine phonetic detail to solve the lexical ambiguity produced by place assimilation. He showed, for example, when the compound noun *right berries* is assimilated to *ripe berries*, the assimilated [p] differs from the unassimilated form [rɑipbɛriz] from 'ripe berries'. The assimilated [p] still bears some cues for an alveolar place of articulation. This finding has two consequences. First, when the listener is presented with a strong [p] in the phrase [rɑipbɛriz], *ripe* is accessed; however, a slightly weaker [p] also activates *right*. In addition, the slightly weaker [p] facilitates the recognition of the upcoming labial segment (see also Gow, 2001; 2003). If listeners make use of fine phonetic detail in strongly reduced forms, this could potentially help them in interpreting whether a reduced form was heard or not, and thus whether they should attempt a reconstruction process.

In Experiment 3, we take a similar approach as in Gow's (2002) experiments. Analogous to the use of assimilated and intended segments in Gow's experiments, we used cross-splicing to replace the acoustic realization of a "surface" segment in a reduced form with an "intended" segment. For example, in the reduced form [məneə] from the canonical form [bənedə] *beneden* 'downwards' the "underlying" segment [b] has changed into the "surface" segment [m]. We replaced this "surface" segment [m] with an "intended" segment /m/ from the same speaker which did not arise from reductions (e.g., [mɛt] *met* 'with'). We examined whether listeners are sensitive to the subtle difference between the "surface" segment [m] and an "intended" segment /m/. In other words, will the "surface" segment [m] be comparable to an "intended" segment /m/ or will the "surface" segment [m] still contain traces of the "underlying" segment [b]? We predicted that the cues of the "intended" segment /m/ would bias listeners' interpretation of the reduced form [mənеə] more towards the "reduced form" competitor (e.g., *meneer* 'mister').

# EXPERIMENT 3

# METHOD

## Participants

We tested 33 Dutch native speakers, who were paid for their participation. The participants reported normal hearing and normal or corrected-to-normal vision. None of the participants took part in the previous experiments of this chapter or in any of the experiments reported in Chapter 2 of this thesis.

## Materials and procedure

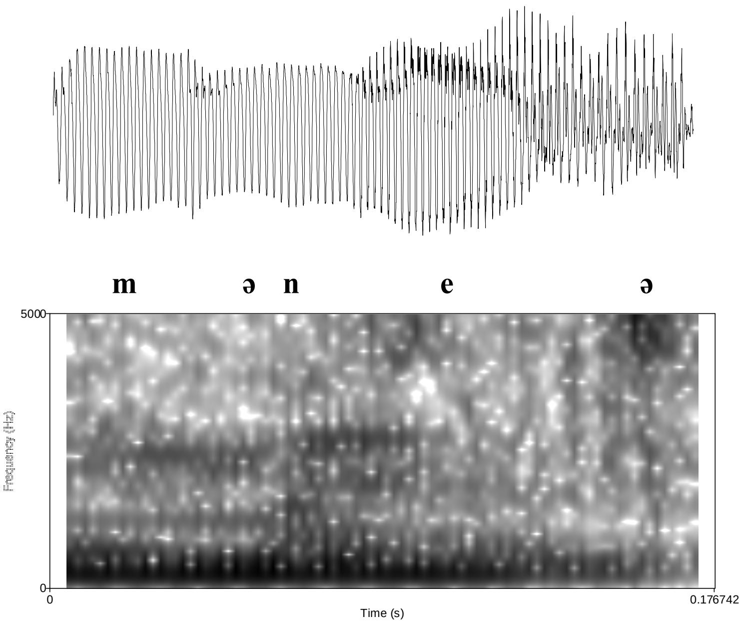
We searched for the same segment from the same speaker in the Spoken Dutch Corpus (Oostdijk, 2000) to replace the "surface" segment [m] in the reduced form [mənеə] with an "intended" segment /m/. Additionally, we attempted to find the same segmental context surrounding a "surface" segment as in the original speech fragment.

For 23 out of the 32 reduced forms, we found appropriate "intended" segments to do the cross-splicing manipulation. This experiment thus used a subset of the same materials used in the previous experiments. For example, for the reduced form [mənеə] (from the context: *buigt het zo af en dan valt het naar beneden, dat is echt* 'it bends like this and then it falls down, that is really') we found an onset /m/ in the word *met* /mɛt/ 'with' in the context *Je kan altijd een keer met korting reizen* 'You can always travel sometime with discount'. If the crucial reduction in a word form occurred in medial or offset position, such as in [ʋɛs] for [ʋɛtstrɛit] *wedstrijd* 'match', we looked for the critical segments in the same position. Thus, for the reduced form [ʋɛs] (from the context: *'t ergste is nog als de wedstrijd dus afgelopen is* 'the worst thing is if the match is finished') we found the "intended" segments /ɛs/ in the word blessuretijd 'injury time' in the context *en uh en en blessuretijd* 'and uh and and injury time'.

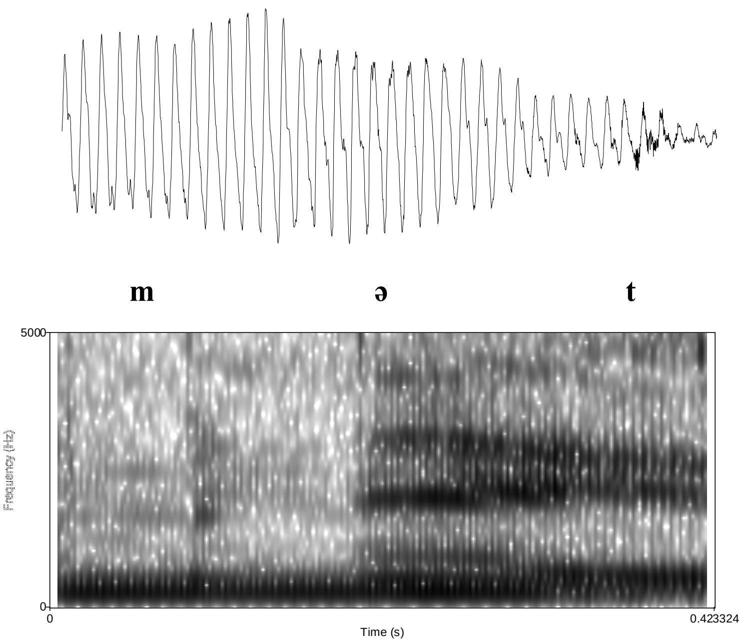
After finding the appropriate "intended" segments we, for example, deleted the "surface" segment /m/ in [mənеə] and replaced this segment with an "intended" onset /m/ from *met* 'with'.[3] Similarly, for the "surface" segments /ɛs/ in [ʋɛs] we replaced this with an "intended" mid /ɛs/ from *blessuretijd* 'injury time'. However, before replacing the "surface" segment with the "intended" segment, we edited the "intended" segment with the PSOLA component of the PRAAT software package (Boersma, 2001) to make the fit as good as possible. First, we made the "intended" segment as long as the "surface" segment. Secondly, we re-synthesized the "intended" segment with the original pitch contour of the "surface" segment. Additionally, we gave the "intended" segment the same amplitude as the "surface" segment. In case it was necessary, we also added noise to the "intended" segment to approximate the noise level of the "surface" segment (see Figure 4). The segments were spliced at zero-crossings and we kept the glottal phases intact to avoid splicing problems.

---

[3] Note that it is problematic to do controlled acoustic measurements on the "surface" and on the "intended" segments. Obviously, it is possible to do measurements, but there is a need for good control tokens. All segments come out of different contexts; therefore, any obtained measure depends on different speakers, different prosodies, and different quality of the sounds. Most of the sentences contain quite some noise, which also prevented us from doing good controlled measurements on the segments.


m ə n e ə
5000
0
Frequency (Hz)
0
Time (s)
0.176742

A)


m ə t
5000
0
Frequency (Hz)
0
Time (s)
0.423324

B)

C)

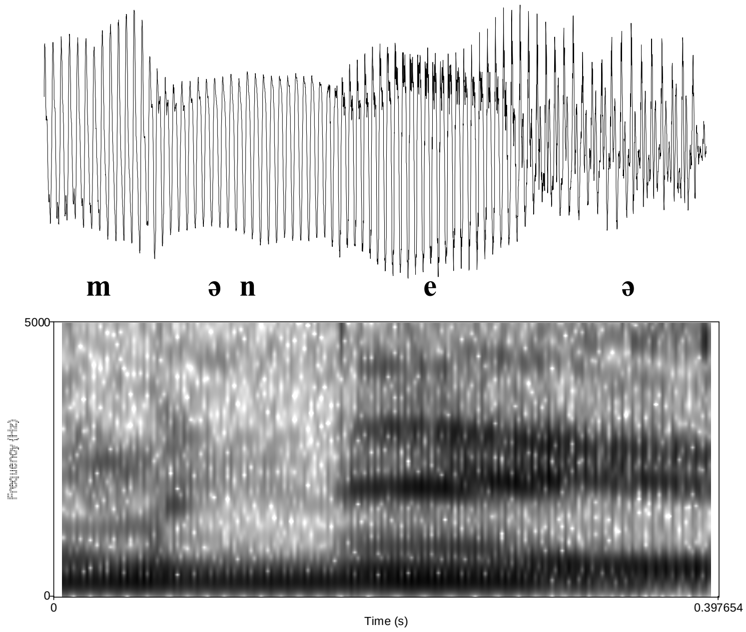



Figure 4: *Realizations of the "surface" segment [m] (Fig. 4A), the "intended" segment /m/ (Fig. 4B), and the cross-spliced [m] (Fig. 4C). See text for details.*

As in Experiment 2, the cross-spliced forms were presented in sentential contexts. We used the same eye-tracking display as in the previous experiments. Note that the experimental items consisted only of the 23 cross-spliced forms. We also selected 36 filler items. The procedure was identical to the previous experiments. The experimental session took approximately 10 minutes.

## Design and analysis

This experiment consisted of only one condition, i.e. the cross-spliced form condition. A comparison was made between this condition and the reduced form condition of

Experiment 2 to investigate whether competition works differently for both forms. Note that only a set of 23 cross-spliced items was presented. In the analysis we therefore compared these cross-spliced items with the same 23 reduced forms of Experiment 2. This data will be reported here as well. Word Form was coded as a numeric contrasts, in which reduced forms were coded as -0.5 and cross-spliced forms as 0.5. The analyses were performed as in Experiment 2.

## RESULTS and DISCUSSION

Table 3 shows the percentage of correct rejections and the percentage of incorrect click responses for the three word types. An analysis on the error rates showed a main effect of Word Form ($\beta_{Word\ Form} = -1.25$, $p < 0.0001$), indicating that listeners were more accurate in the reduced form condition (93.7%) than in the cross-spliced form condition (88%).

Table 3: *Task performance in Experiment 3.*

| | Forms presented in sentential contexts | |
|---|---|---|
| % Click-responses | Experiment 3: Cross-spliced forms | Experiment 2: Reduced forms (23 items) |
| Correct rejections | 88 | 93.7 |
| "Canonical form" competitor | 1.8 | 1.4 |
| "Reduced form" competitor | 10.2 | 4.9 |
| Distractors | 0 | 0 |

*Note that 'Correct rejections' correspond to clicks in the centre of the screen.*

For the eye-movement data, we plot and analyze the proportion of fixations of correct trials only for the cross-spliced and the reduced forms, since there were only few errors. Figure 5A shows the mean fixation proportion to the two types of competitors and the averaged distractors from acoustic target onset (0 ms) to 1200 ms thereafter for the cross-spliced form condition. In addition, we plotted the same 23 items of the reduced form condition of Experiment 2 in panel 5B. Note that this subset shows a similar competition pattern as the 32 reduced items in Experiment 2 (see Figure 3B). In the analysis, we compared fixations in the cross-spliced form condition with the reduced form condition containing only those 23 items. As in the previous experiments, we analyzed fixations during an early (400-800 ms) and a late time window (800-1200 ms).

We first analyzed whether there was an effect of overall competition. We found an overall effect in both time windows (early: $\beta_{Intercept} = 0.64$, $p_{MCMC} < 0.001$; late: $\beta_{Intercept} = 0.81$, $p_{MCMC} < 0.001$). Overall competition was not dependent on Word Form in both time windows ($p_{MCMC} > 0.1$).

Second, we analyzed whether listeners looked more often to the "canonical form" competitors than to the "reduced form" competitors. We found no overall effect of specific competition (all $p_{MCMC}$'s $> 0.05$). Specific competition was, however, influenced by Word Form in only the late time window ($\beta_{Word\ Form} = 0.67$, $p_{MCMC} < 0.05$). We further analyzed how the input form affected the effect of specific competition in the late time window by looking at each condition separately. The analysis showed a marginally significant effect of specific competition in the reduced form condition ($\beta_{Intercept} = 0.43$, $p_{MCMC} < 0.06$), but not in the cross-spliced form condition ($p_{MCMC}$'s $> 0.1$).

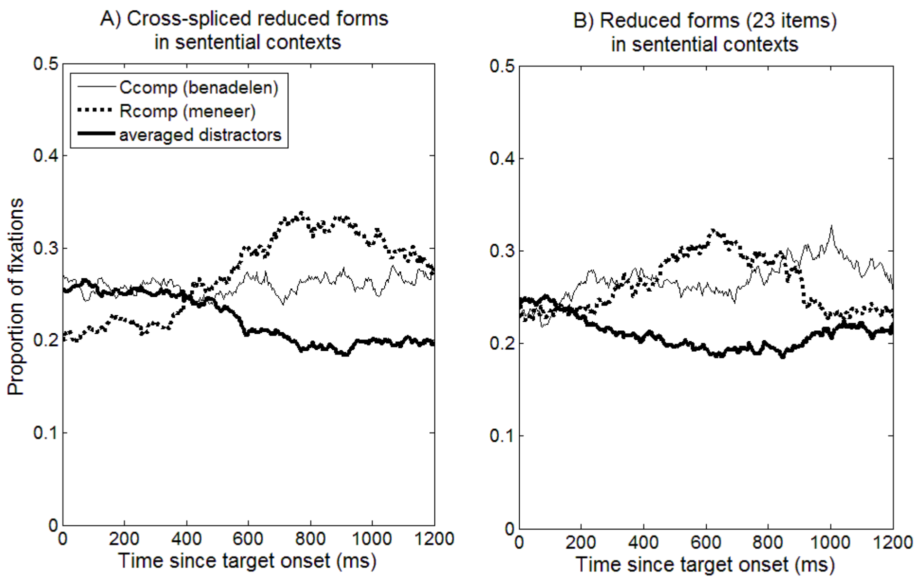



Figure 5: *Proportion of fixations to the "canonical form" competitor (Ccomp), the "reduced form" competitor (Rcomp), and the averaged distractors, in (A) Cross-spliced reduced forms, and (B) 23 Reduced forms for the correct trials presented in sentential contexts.*

In sum, Experiment 3 examined whether phonological competition is influenced if we purposely change segmental information in the reduced form itself. The error pattern showed that listeners made more errors in the cross-spliced form condition than in the reduced form condition. Participants often clicked on the "reduced form" competitors when they heard cross-spliced forms, indicating that listeners interpret the cross-spliced [m] more often as the "intended" segment [m] than as the "surface" segment [m].

The eye-movement data showed that the late rise of the "canonical form" competitor in the reduced form condition was not present in the cross-spliced form condition. Instead, listeners remained looking at the "reduced form" competitor in the cross-spliced form condition. This indicates that the cross-spliced segments are interpreted as real segments, and that reduced forms contain residual cues with fine phonetic information of the canonical form. Thus listeners are sensitive to fine phonetic information in interpreting whether a reconstruction process is likely to be involved or not. The initial onset /m/ suggests that *beneden* is unlikely to have been intended and thus no (or a reduced) late reconstruction process takes place.

Note that it is difficult to find out what the listeners were exactly picking up from the cross-spliced segments. The differences between the original and the spliced stimuli were very small because we matched their duration, pitch, and amplitude. We therefore only showed that these cues do not contribute significantly to the present findings. It is most likely that spectral differences influenced the results. Such differences also appear to be important for the interpretation of assimilated segments (e.g., Gow, 2002). For further research it would be interesting to investigate the role of duration more explicitly. This is difficult to achieve for the current purposes because the duration differences between the reduced and canonical utterances not only differed by being reduced and unreduced, but also by position in the sentence, speaker differences, and many other factors. From an experimental point of view, it was therefore best to keep the duration similar between the original and the cross-spliced segments.

## GENERAL DISCUSSION

Three eye-tracking experiments were conducted to investigate the nature of phonological competition during the recognition of strongly reduced forms. Competition processes were measured using a printed-word, target-absent variant of the visual world paradigm (Huettig & McQueen, 2007). Participants' looks were tracked to four printed words on a computer screen: a "canonical form" competitor (e.g., *benadelen* 'to disadvantage', phonologically similar to the canonical form), a "reduced form" competitor (e.g., *meneer* 'mister', phonologically similar to the reduced form), and two phonologically unrelated distractors, while listening to canonical (e.g., [bənedə]) and strongly reduced forms (e.g., [məneə]) of a spoken target word (e.g., *beneden* 'downwards').

A recent study (Brouwer et al., under revision) demonstrated that listeners penalize acoustic mismatches less strongly when listening to reduced speech than when listening to fully articulated speech. When faced with a listening situation in which phonological reductions frequently occurred, listeners directed their eye gaze to a similar degree to both types of competitors ("canonical form" or "reduced form" competitors) independent of the target's *exact* spoken form. In the present research we examined whether phonological competition during casual speech, which typically contains many phonological reductions, can *ever* be influenced by the exact phonetic form of the spoken word.

In contrast to the experiments in Brouwer et al. (under revision), the printed target word was removed from the visual display in the experimental trials of the experiments in the current study: although during filler trials the (printed) target word was present, in one third of the trials (the experimental trials) the target word was absent from the visual displays. When the target word was absent participants were required to click the centre of the screen. Such an experimental set up has been shown to greatly increase competition effects in the visual world paradigm (Huettig & Altmann, 2005; Huettig & McQueen, 2007; and Huettig et al., submitted, for discussion).

By maximizing the likelihood of observing phonological competition in this way, we observed that eye gaze to the different types of competitors was influenced by the input form. In Experiment 1, when a canonical input form was presented in isolation, participants made fewer errors and the "canonical form" competitor attracted more overt attention than the "reduced form" competitor. When a reduced input form was heard in isolation, participants were incorrect on almost 35% of trials. Thus, recognizing isolated reduced forms was, perhaps unsurprisingly, harder than recognizing isolated canonical forms. On trials in which participants made no errors, there was no significant difference in overt attention between "reduced form" and "canonical form" competitors. In sum, when listeners hear isolated canonical forms, they look more to "canonical form" competitors, but when they hear isolated reduced forms (and made the correct mouse click response), they do not differ in their looks to any of the competitor types.

In Experiment 2, we presented the spoken target words in a context of several surrounding words. Here, we observed again that "canonical form" competitors attracted more overt attention than "reduced form" competitors when canonical forms were presented, just as in Experiment 1. When reduced forms were presented, we observed significantly more looks to the "reduced form" competitors than to "canonical form" competitors during an early time window. In a later time window however, this difference between the two competitors disappears: as in Experiment 1, there was no significant difference in overt attention between "reduced form" and "canonical form" competitors.

Therefore, when participants encountered reduced forms in the present experimental set-up, phonological competitors with a quite different surface form from the canonical counterpart competed early on during the acoustic duration of the reduced form for visual attention. In other words, we see that listeners shortly activate canonical forms of other words that are acoustically similar to the reduced forms taken from casual speech. Such activation may underlie the delays reported for recognition of reduced forms in previous studies using offline techniques (Ernestus et al., 2002; Kemps et al., 2004). These "reduction costs" are also reflected in the error patterns: listeners made more errors in the reduced form condition than in the canonical form condition, even though performance was considerably improved compared to that in Experiment 1.

This competition pattern changed later in time, after when participants had heard reduced forms. We observed that during 800 to 1200 ms after the acoustic onset of the reduced form, the "canonical form" competitor attracted more visual attention than the "reduced form" competitors. One way to interpret this pattern of results is that canonical forms could be successfully reconstructed from their reduced forms. We conjecture that this process is time-consuming due to the early, momentary activation of unwanted competitors; competitors which in contrast are not strongly activated by a canonical pronunciation of the same word.

The results of Experiment 2 thus suggested that initially, listeners are not sensitive to the fine phonetic detail in phonological reductions to block competition from "reduced form" competitors. Only in a later time window they reconstruct the reduced form to the canonical form, as an increase in looks to the "canonical form" competitor suggests. To find out whether listeners are sensitive to the fine phonetic detail at any stage of word recognition, we replaced the reduced segment by a fully pronounced segment in Experiment 3. Here, we examined how fine phonetic detail in reduced forms affects word recognition. We replaced the "surface" segment /m/ from [məneə] with an "intended" segment /m/ from a canonical form, and presented these cross-spliced forms to listeners. The early eye-movement data showed the same rise of fixations to the "reduced form" competitors as in Experiment 2. However, the late eye movements were influenced by the cross-splicing: the late rise of the "canonical form" competitor observed in Experiment 2 was absent in Experiment 3. This seems to indicate that the cross-splicing impeded the reconstruction process and therefore the recognition of the intended word. Consistent with this account, participants also made more errors with the cross-spliced than with the original stimuli. Note that the later result cannot be explained as a splicing artifact, as participants more often (falsely) recognized the cross-spliced reduced form as a different intended word. The results of Experiment 3 then revealed that there are subtle phonetic differences between a given phoneme in a reduced form (e.g., the /m/ in the reduced form of *beneden*) and the same phoneme when produced as part of an intended canonical form (e.g., the /m/ in the canonical form of *met*). Listeners thus appear to be sensitive to these differences when listening to cross-spliced forms, but only at a later period, when they are reconstructing the form.

The results of Experiment 3 also rule out an alternative explanation for the pattern of results in Experiment 2 (i.e., early rise of the "reduced form" competitor, and late rise of the "canonical form" competitor). It could be argued that this pattern reflects a "form matching" strategy. It is conceivable that listeners just match the strongly reduced form they hear with the first best-matching word they see on the screen (i.e., "reduced form" competitor) and then they look at the second best-matching word on the screen (i.e., "canonical form" competitor). In other words, listeners may have strategically cut the lexicon down to the four items on the screen. In Experiment 3, we only slightly manipulated the acoustic input, but this manipulation had a large influence on the results. Looks to the "canonical form" competitor never increased over time, whereas this did happen in the reduced form condition of Experiment 2. It is therefore difficult to see how a "form-matching" strategy would be influenced by such a subtle phonetic manipulation that it could explain the difference in results between Experiment 2 and 3. Additionally, the results of Experiment 2 themselves also provide an argument that invalidates such a strategic account of the results of Experiment 2. If participants were using a "form-matching" strategy, we would have found a similar pattern in the canonical form condition, with looks to the "reduced form" competitor rising late in time when the "canonical form" competitor has been ruled out as a potential target. In the canonical form condition we observed instead that the "canonical form" competitor attracted more overt attention than the "reduced form" competitor in both time windows.

How do our results relate to the ongoing debate of how pronunciation variants are recognized? Different views on how listeners recognize reduced forms are postulated in the literature. Two main classes of accounts focus on different mechanisms. One class of accounts proposes that a reconstruction process occurs at a prelexical level, which mediates between the speech signal and the lexicon, on the basis of fine phonetic detail in the signal, phonological context (e.g., Gaskell, 2003), or by top-down restoration (McClelland & Elman, 1986; Warren, 1970). For example, upon hearing the reduced pronunciation [mənеə], listeners may reconstruct the corresponding canonical pronunciation *beneden*. This full form then activates the representation of the word *beneden* in the lexicon, and competes with other /b/-initial words for recognition.

A second class of accounts assumes that phonological variants are stored in the mental lexicon. Two different versions of this account exist. On the episodic account, the entry for a given word in the mental lexicon consists of detailed and concrete episodic memories of pronunciations of that word that have been encountered previously (e.g., Bybee, 2001; Goldinger, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert, 2001). More precisely, a "grainy spectrogram" of such variants would be stored in the mental lexicon. Such episodic traces of phonological variants are stored in the mental lexicon next to traces of canonical forms of those words. For example, episodic traces such as [mənеə] for *beneden* are stored in the mental lexicon next to traces of canonical forms of those words. Proponents of the second lexical-storage account argue that different pronunciation variants are stored as abstract phonological forms (e.g., Connine, 2004; Ranbom & Connine, 2007). On this view, both the phonological variants (e.g., [mənеə]) and the canonical form (e.g., [bənedə]) would be stored, but as abstracted variants of the canonical representation that do not include indexical properties of spoken words such as voice quality, speech rate, pitch, and so on, as would be the case for episodic traces of each variant.

It may well be the case that both mechanisms play roles in the recognition of reduced forms. Gaskell and Marslen-Wilson (1998) already found that phonological reconstruction of pronunciation variants is more efficient for words than nonwords. Recently, more evidence is accruing that even for the same pronunciation variant, phonological and lexical processes may operate together. Snoeren, Gaskell, and Di

Betta (2009) showed that phonological reconstruction for variants with place assimilation works more efficiently on known words than for nonwords. In a similar vein, Pitt (2009) provided evidence that variants with nasal flaps (*center* pronounced as *cenner*) are recognized by a combination of lexical and phonological processes. He taught participants new words with medial /t/ that could be flapped (e.g., *senty*). Participant did not accept flapped variants (*senny*) as instances of the same word unless they had previously been exposed to the variant form. While this highlights the importance of lexical storage of variant forms, an additional experiment showed that phonological processing plays a role as well. Nasal flapping is much more likely to occur if the /t/ is followed by a reduced vowel (e.g., *center*) then if it is followed by a full vowel (e.g., *content*). Pitt showed that this phonological conditioning matters. Variants with nasal flaps were only accepted if followed by a reduced vowel, so that exposure to a variant form was not sufficient for recognition.

Although the current data cannot distinguish conclusively among these alternative accounts, they appear to be more in line with a reconstruction mechanism. The late looks to the "canonical form" competitors in Experiment 2 fit best with the notion that listeners reconstruct canonical forms from reduced forms. In addition, Experiment 3 revealed that this late rise of the "canonical form" competitor only appears when fine phonetic detail is preserved in the signal. It is important to note here that our results cannot distinguish whether reduced forms are stored in the mental lexicon or not. Such storage may greatly depend on, for example, how strongly reduced a word form is or how frequent it is. Further research is required to clarify the contributions of the two mechanisms during the comprehension of strongly reduced forms.

To conclude, the current study addressed the implications of reduction processes for phonological competition. Using an experimental set-up that maximizes the likelihood of measuring phonological competition, we observed that strongly reduced forms in casual speech can activate competitors which are similar to the phonological surface form of the reduction. These same competitors are not strongly activated by a canonical pronunciation of the same word. We conjecture that this added competition is one of the causes of the delay during the recognition of strongly reduced forms. Although this delay demonstrates that processing speech reductions is cognitively costly, our results also show that listeners can exploit fine phonetic detail to reconstruct canonical forms from their strongly reduced counterparts. This provides further evidence for the efficiency of the spoken word recognition system.

# DISCOURSE CONTEXT AND THE RECOGNITION OF REDUCED AND CANONICAL SPOKEN WORDS

**CHAPTER 4**

This chapter is a slightly revised version of Brouwer, S., Mitterer, H., & Huettig, F. (submitted). Discourse context and the recognition of reduced and canonical spoken words.

## ABSTRACT

In two eye-tracking experiments we examined whether wider discourse information helps the recognition of reduced pronunciations (e.g., 'puter') more than the recognition of canonical pronunciations of spoken words (e.g., 'computer'). Dutch participants listened to sentences from a casual speech corpus containing canonical and reduced target words. Target word recognition was assessed by measuring eye fixation proportion to four printed words on a visual display: the target, a "reduced form" competitor, a "canonical form" competitor and an unrelated distractor. Target sentences were presented in isolation or with a wider discourse context. Experiment 1 revealed that target recognition was facilitated by wider discourse information. Importantly, the recognition of reduced forms improved significantly when preceded by strongly rather than by weakly supportive discourse contexts. This was not the case for canonical forms: listeners' target word recognition was not dependent on the degree of supportive context. Experiment 2 showed that the differential context effects in Experiment 1 were not due to an additional amount of speaker information. Thus, these data suggest that in natural settings a strongly supportive discourse context is more important for the recognition of reduced forms than the recognition of canonical forms.

# INTRODUCTION

Casual speech used in everyday conversations is highly variable and contains many phonological reductions (Ernestus, 2000; Johnson, 2004). For example, during a casual conversation a speaker of Dutch may pronounce the word *beneden* [bəneːdə] 'downwards' as [məneːə]. Reduced forms can thus be substantially different from their canonical counterparts. Surprisingly, however, reductions do not seem to hinder the communication between speaker and listener. An obvious reason for this may be that phonological and sentential context help listeners to recognize reduced forms. However, reduced forms in sentential contexts are still misidentified in almost 10% of the cases (Ernestus, Baayen, & Schreuder, 2002). The question therefore arises how listeners recognize the meaning of reduced forms successfully. In the present research, we test whether a supportive wider discourse context is a key factor for successful recognition. More specifically, we examine the hypothesis that a strongly supportive discourse context is more important for the recognition of reduced forms (e.g., [məneːə]) than it is for the recognition of canonical forms (e.g., [bəneːdə]).

Most of the past research using carefully pronounced laboratory speech has investigated the effect of context for the *prediction* of upcoming words rather than the effect of the wider discourse context on the *recognition* of spoken words (e.g., Altmann & Kamide, 1999; Altmann & Kamide, 2007; Altmann & Kamide, 2009; Kamide, Altmann, & Haywood, 2003). Altmann and Kamide (1999), for example, showed that when listeners hear a sentence such as 'The boy will eat the cake' in the context of a scene depicting a boy and a cake (and other things), they shift their eye gaze towards the cake even before "cake" starts to acoustically unfold. Altmann and Kamide interpreted this finding as evidence that selectional information conveyed by a verb can be used to predict an upcoming theme. Kamide et al. (2003) further explored whether the combination of verb information with the preceding grammatical subject can be used for prediction. They found increased fixations to a motorbike when participants heard 'The man will ride...', but increased fixations to a carousel when they heard 'The girl will ride...'. Therefore information provided by the grammatical subject and by the verb can jointly constrain anticipation (at least when a visual context is present, see also Kamide, Scheepers, & Altmann, 2003, for evidence that case-marking can be used for prediction).

Most studies investigating the effects of (semantically predictable) context have used isolated sentences only. Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005), however, investigated how wider discourse context (i.e., more than one sentence) can be used to predict an upcoming noun (e.g., as in 'The burglar had no trouble locating the secret family safe. Of course, it was situated behind a ...'). Event-related potentials to determiners and adjectives were measured for prediction-consistent (e.g., 'big$_{NEU}$ painting$_{NEU}$') and prediction-inconsistent nouns (e.g., 'big$_{COM}$ bookcase$_{COM}$'). The results showed an N400 effect for adjectives inconsistent with the discourse-predictable noun relative to adjectives consistent with the discourse-predictable noun. The N400 component is typically associated with difficulty during semantic integration of words in a sentence context (Kutas & Hillyard, 1984). When these stories were presented in a self-paced reading task, prediction-inconsistent adjectives also slowed readers down. These data suggest that people use wider discourse context to predict upcoming words.

The focus of our present research is not on whether people can use such context for prediction (evidently they can) but on the effect of wider discourse context on the actual *recognition* of spoken words. Note in particular that very few studies have looked at the use of context for the recognition of *reduced forms* during casual speech. An exception is a study by Ernestus et al. (2002) who selected samples from a spontaneous speech corpus to examine how listeners recognize highly reduced forms (e.g., [mok] from [moxələk] *mogelijk* 'possible') in Dutch. Participants listened to such forms in sentential contexts (e.g., [zo snɛl mok na ɔː] *zo snel mogelijk naar eh* 'as fast as possible to uhm'), in phonetic contexts (e.g., [ɛl mok na] *el mogelijk naa* 'ast possible to'), and without any context (e.g., [mok] *mogelijk* 'possible'), and were asked to write down the form they heard. The results showed that listeners hardly recognize reduced forms on the basis of the acoustic signal for that word alone. Identification performance increased when highly reduced forms were presented in a phonetic context. However, only when presented in sentential contexts performance for highly reduced forms improved substantially. Nevertheless, listeners still misidentified reduced forms in almost 10% of the cases (see also Kemps, Ernestus, Schreuder, & Baayen, 2004).

Our experiment goes further than Ernestus et al. (2002) by presenting wider discourse context to participants. Moreover, no study so far has compared the recognition of reduced forms with the recognition of canonical forms. Since reduced forms are more difficult to recognize, it is conceivable that discourse information aids reduced and canonical forms to a different degree. Note that Ernestus et al. (2002) used an offline task (self-paced listening) rather than an online task to investigate reduced speech. In the current study we investigate target recognition online by using eye tracking. Participants listen to target sentences, while four printed words are displayed on the screen: the target word (e.g., *beneden* 'downwards'), a phonologically unrelated distractor (e.g., *vakantie* 'holiday'), a "canonical form" competitor (e.g., *benadelen* 'to disadvantage'), and a "reduced form" competitor (*meneer* 'mister'). The critical experimental manipulation was whether the target sentences were preceded by discourse context or not.

Comparing the recognition of a given word in a sentence context with recognition in a wider discourse context leads to a possible confound. The preceding contexts not only contain additional discourse information but also additional speaker information. Many studies have shown that speaker information can be an important aid for the listener to recognize spoken words (e.g., Bradlow, Nygaard, & Pisoni, 1999; Mullenix, Pisoni, & Martin, 1989; Nygaard, Sommers & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993). Furthermore, a large body of research has shown that listeners adapt to speaker-specific characteristics on the time scale of minutes (e.g., Norris, McQueen, & Cutler, 2003; Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2005) and even seconds (Ladefoged & Broadbent, 1957). For instance, Mitterer (2006) showed that adaptation to a speaker is stronger when more information about this speaker is available. Therefore, it is essential to show that the advantage in the processing of reduced (and possibly canonical) forms in a wider discourse context over the processing of the same form in the sentence context is not solely due to more efficient adaptation to the speaker. After all, by presenting wider discourse context, we also expose the listener longer to a given speaker. In Experiment 2, we thus presented the same target sentences with different contexts. Instead of using the actual context in which the word occurred, we selected another arbitrarily chosen sample from the same speaker. These control contexts provided the same amount of speaker information but no matching discourse information. Experiment 2 therefore allows us to measure how much benefit speaker information provides for the recognition of reduced and canonical forms.

In sum, the present research examined the effects of wider discourse context on the recognition of reduced and canonical forms. Critically, we predict that the recognition of reduced forms relies more on strongly supportive contexts than the recognition of canonical forms. To assess how contextually supportive the different contexts were (i.e., both the actual discourse context and the control contexts), we first performed a pre-test with these materials.

## PRETEST

In the present research we use ecologically valid examples of reductions in casual speech. To be able to do this, we have to work with stretches of speech extracted from a spontaneous speech corpus. A downside of using spontaneous speech materials is the lack of control one has over such stimuli. We extracted target sentences and the discourse context directly preceding these target sentences. The discourse contexts provided minimally five seconds of speech of the target speaker. We conducted a pretest to examine whether the selected samples provide supportive discourse information to listeners, which they can use to recognize targets successfully. A second purpose of the pretest was to empirically confirm that the "speaker-only" contexts—to be used in Experiment 2—do in fact not contain any supportive discourse information.

In this pre-test, listeners were asked to rate how well the contexts preceding the target sentences (e.g., *Ja, dat is echt uh... Nou we hebben daar ook nog gestaan. Ik heb daar ook nog gefilmd. En dan komt dat water komt echt zo naar je toe en dan* 'Yes, that is really, uhm… Well, we have also been standing there. I have also made a movie there. And then the water really approaches you and then') matched with the target sentences (e.g., *buigt het zo af en dan valt het naar <u>beneden</u>, dat is echt* 'it bends like this and then it falls <u>down</u>, that is really').

## Method

Eighteen members of the Max Planck Institute subject pool participated in the pretest for which they were paid. None of them reported any hearing disorders and all had normal or corrected-to-normal vision. Listeners were tested individually in a sound-attenuated booth. The presentation of the stimuli was controlled by Presentation software. The auditory stimuli were presented to the participants over headphones.

We presented 112 preceding discourse contexts followed by their accompanying target sentence. Half of the items were experimental items (context A), whereas the other half of the items were control items (context B, to be used in Experiment 2). For the control items we selected random contexts from the same speaker which did not directly precede the target sentences. This presentation mode created an AXBX task, in which A and B were the preceding contexts and X the target sentences. The presentation of A and B was counterbalanced. Each participant received a different random ordering of the stimuli, but started with the same three practice trials to familiarize themselves with the task. A fixation cross appeared for 300 ms between the presentation of the preceding context and the target sentence. This fixation cross was an indication for the participant that the target sentence would start. After the presentation of the target sentence, participants saw a vertical line crossing out a horizontal bar on the screen. The horizontal bar represented a continuum from mismatch (-5) to match (+5). Participants were asked to indicate with the scroll wheel on the computer's mouse whether the preceding contexts matched with the target sentences or not. The scroll wheel enabled participants to move the vertical line on the continuum to the left (-5) or to the right (+5). Once participants made a decision, they had to confirm the position of the vertical line on the continuum with the left mouse button. After they clicked on the left mouse button, the next trial initiated. Participants were put under no time pressure to perform this action. There was a short pause half way through the experimental list. The pretest lasted about 35 minutes.

**Results**

Table 1 shows the rating scores for Word Form (canonical versus reduced) and for Information (discourse versus control). A mixed effect logistic regression model was used to test whether the target sentences were rated to match better with the discourse contexts than with the control contexts. This was the case ($\beta_{Information}$ = -2.76, $p_{MCMC}$ = 0.0001). We found no main effect of Word Form or an interaction between Information and Word Form (all $p_{MCMC}$'s > 0.1). The results indicate that our stimulus selection was appropriate: contexts with discourse information (to be used in Experiment 1) provide more useful information for listeners than our control contexts (to be used in Experiment 2).

Note that the range of ratings for our selected experimental items was wide for both target types (for reduced targets: ranging from -2.11 to 3.83; for canonical targets: ranging from -2.67 to 4.11). This shows that some contexts were strongly supportive whereas other contexts were only weakly supportive. Thus, as in real world situations, not all discourse contexts provide supportive information to a similar extent. We therefore took into account how supportive a given context is for a given item in our data analysis. For visualization purposes, we used a median-split to label the canonical and reduced items below the median as weakly supportive contexts and those above the median as strongly supportive contexts. For the statistical data analysis, we used the degree of support as a covariate to examine whether this influences target recognition as measured by fixation proportion.

Table 1: *Mean ratings in the pretest (scale from mismatch (-5) to match (+5)).*

| | Word form | |
|---|---|---|
| Information | Reduced | Canonical |
| Discourse (Experiment 1) | 1.90 (1.99) | 2.19 (2.11) |
| Weakly supportive | 0.63 (1.14) | 1.14 (1.62) |
| Strongly supportive | 3.16 (0.50) | 3.24 (0.40) |
| Control (Experiment 2) | -0.52 (2.44) | -0.91 (2.28) |

*Note: Standard deviation between parentheses.*

# EXPERIMENT 1
# METHOD

## Participants

Forty-eight participants from the Max Planck Institute's subject pool, undergraduates at the Radboud University Nijmegen, were paid to participate in this experiment. All participants were native speakers of Dutch and had normal or corrected-to-normal vision. No participant reported any (history of) hearing problems. None of the participants took part in the pretest or in any of the experiments reported in Chapter 2 or 3 in this thesis.

### Materials

Twenty-eight polysyllabic, mid-to-high frequency content words were selected from the Spoken Dutch Corpus (Oostdijk, 2000) as target words, of which 28 were realized canonically (e.g., [bəneːdə] for *beneden* 'downwards') and 28 were pronounced in a reduced way (e.g., [məneːə]). This experiment thus used a subset of the same target words as in Chapter 2 and 3 of this thesis. We selected Dutch recordings (and not Flemish) because this variant of Dutch is most familiar to the participants in our subject pool. Recordings with background noise or overlapping speech were excluded. Two independent raters transcribed the target words. Spectrograms were made with the software package PRAAT (Boersma, 2001) to observe the signal in auditory and visual form. The raters compared their independent transcriptions to verify agreement between the transcriptions. In case of disagreement, the transcribers reached consensus. Canonical targets were selected if (almost) all segments were fully realized, whereas their reduced counterparts were selected if one or more segments were missing.

Each target was embedded in a sentence. For each of the target sentences, we searched in the spontaneous speech corpus for the discourse context directly preceding the target sentence. A preceding context was included in the study if the speech of the target speaker in the preceding context consisted of a minimum duration of 5 seconds. Participants listened to the target sentence alone (sentence only condition) or to the additional context and the target sentence (wider discourse condition).

We used the printed-word variant (Huettig & McQueen, 2007; McQueen & Viebahn, 2007) of the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). On each trial, participants were presented with a visual display containing four printed words. Each display contained the printed target word (e.g., *beneden* 'downwards'), a "canonical form" competitor (e.g., *benadelen* 'to disadvantage'), a "reduced form" competitor (e.g., *meneer* 'mister'), and a phonologically unrelated distractor (e.g., *vakantie* 'holiday'). Figure 1 shows an example of a visual display. We included competitors of the target word to make the task ("Click on the target word that appears in the sentence") more challenging to participants. A "canonical form" competitor shared more onset overlap with the canonical target (e.g, *benadelen* 'to disadvantage' [bənaːdeːlə] for [bəneːdə]), whereas the initial segments of a "reduced form" competitor overlapped better with the reduced target (e.g., *meneer* 'mister' [məneːr] for [məneːə]). In such a display, there are always two to three phonologically related words, of which one was the target. We therefore masked this pattern by adding filler items, which we also selected from the spontaneous speech corpus. Each filler trial also consisted of two to three phonologically similar words and one to two unrelated words, but half of the time one of the dissimilar words was the target and half of the time one of the similar words was the target. In this way, the fillers prevented participants from developing any expectation that items sharing certain phonological attributes would be mentioned.

Two lists were created. One list contained 28 reduced targets and 28 fillers, whereas the other list contained 28 canonical targets and 28 fillers. The order of each list was randomized, so that each participant received a different order of presentation. The position of the four types of printed words was also randomized over the four quadrants on the screen. That is, each printed word appeared with equal probability on each of the four screen positions over the course of an experimental run. To familiarize participants with the task, the experimental run started with a warm-up session containing 6 practice items. These items were also selected from the corpus.

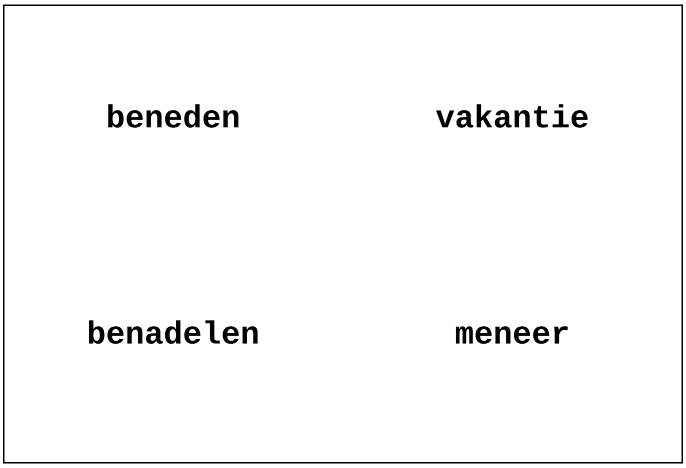



Figure 1: *Example of a printed-word display presented to participants. The spoken target word in this example was* beneden *'downwards'. The four printed words are the target (*beneden *'downwards'), a distractor (*vakantie *'holiday'), a "canonical form" competitor (*benadelen *'to disadvantage'), and a "reduced form" competitor (*meneer *'mister').*

## Procedure

Participants were tested individually in a sound-attenuated booth. They were seated at a comfortable viewing distance from the computer screen. Eye movements were monitored at a sampling rate of 1 kHz with an SR-Research EyeLink1000 eye tracker (used in the tower-mount version). The presentation of the auditory and visual stimuli was controlled with SR-Research programme Experiment builder. Note that this is a different type of eye-tracker than the one we used in the experiments in the previous chapters. The auditory stimuli were presented to the participants over headphones.

Participants received written instructions on the screen. They were instructed that they would first see a cross in the centre of the screen. During the presentation of this cross they either listened to an auditory fragment (i.e., the preceding context) or to a 300 ms silence. After the auditory fragment or the silence, the target sentences were presented. During this presentation, the four printed words appeared in a 24-point Courier font on the screen. The centres of the printed words corresponded, independently of the length of the words, to the centres of the quadrants on the screen. The participants had to use the computer's mouse to click on the printed word that appeared in the target sentence. After they clicked with the mouse on one of the words, the next trial initiated. Participants were put under no time pressure to perform this action. A central fixation cross appeared centred on the screen after every ten trials, permitting for drift correction in the calibration. Note that the experiments in the previous chapters presented the fixation cross every five trials.

Each participant first completed 6 practice trials. After that, one of the two lists was presented in random order. The experimental session lasted about 15 minutes.

## Design and analysis

Reduced targets were presented to half of the participants and canonical targets to the other half of the participants. Click-responses and eye movements were the dependent variables. For the click-responses we calculated the percentage of correct clicks to the target and the percentage of incorrect clicks to the competitors and the distractor. Participants made no errors in any of the experiments. Statistical analyses on the errors were therefore not carried out.

For the eye-movement data we discarded blinks and saccades. In order to assess the effect of the wider discourse context on the actual recognition of reduced and canonical forms, we analyzed our data from 200 ms onwards because of estimates that it takes 200 ms to program and launch a saccadic eye movement (e.g., Matin, Shao, & Boff, 1993). Thus before 200 ms after word onset fixations are unlikely to be driven by acoustic information from the critical target word. As Figure 2 illustrates, fixations to the competitors in the wider discourse condition converged with the distractor at around 1000 ms after word onset. We therefore choose to statistically analyze proportion of fixations during the 200-1000 ms time window after the acoustic onset of the target word.

The dependent variable was the proportion of fixations to the target word. For the analysis we transformed the fixation proportion with the empirical logit function (cf. Barr, 2008). We tested whether target fixations were influenced by the presence versus absence of wider discourse information using linear mixed effects models (Baayen, Davidson, & Bates, 2008), with participants and items as random effects and in which Discourse Information was coded as a numeric contrast (-0.5 and 0.5, cf. Barr, 2008). The sentential context only condition was coded as -0.5 and the wider discourse condition as 0.5. The amount of support provided by the wider discourse context—as obtained in the pretest—was used as a covariate. We estimated *p*-values by using Markov Chain Monte Carlo simulations (Baayen et al., 2008).

# RESULTS

Figure 2 shows the proportions of fixations over time from -200 ms until 1200 ms after target word onset for both conditions in strongly and weakly supportive discourse contexts. We plotted the two competitors together by taking the average instead of each of them separately because the competitors did not differ significantly from each other.

We first analyzed strongly and weakly supportive discourse contexts together to examine whether listeners benefit in general from the presence of the discourse context. In the 200-1000 ms time window, we found a main effect of Discourse Information for the reduced targets ($\beta_{Discourse\ Information} = 1.08$, $p_{MCMC} < 0.0001$) and for the canonical targets ($\beta_{Discourse\ Information} = 0.99$, $p_{MCMC} = 0.0001$). The positive beta indicates more looks to both types of targets when the discourse context was present than when it was absent.

Next, we added Degree of Support as a covariate to the target analysis on the reduced forms. This analysis again showed a main effect of Discourse Information

($\beta_{Discourse\ Information}$ = 1.08; $p_{MCMC} < 0.001$), a main effect of the Degree of Support ($\beta_{Degree\ of\ Support}$ = 0.45; $p_{MCMC} < 0.01$), and an interaction effect of Discourse Information by Degree of Support ($\beta_{Discourse\ Information\ *\ Degree\ of\ Support}$ = 0.59; $p_{MCMC} < 0.001$). This interaction shows that the ratings—indicating how supportive the discourse contexts were—influenced target fixations only when the discourse contexts were actually presented. This shows there was nothing inherently different between the sentences that happened to occur in strongly versus weakly supportive discourse contexts. The positive beta-weight of the interaction shows that the presence of a strongly supportive discourse context (2A) aided word recognition for reduced forms more than the presence of a weakly supportive discourse context (2B).

Degree of Support was also added as a covariate to the target analysis on the canonical forms. This analysis showed only a main effect of Discourse Information ($\beta_{Discourse\ Information}$ = 0.99, $p_{MCMC}$ = 0.0001). Neither a main effect of Degree of Support nor an interaction between Discourse Information and Degree of Support was found ($p_{MCMC} > 0.1$). For the canonical forms, the benefit provided by the presence of a discourse context was therefore independent of how supportive the context actually was (see Figure 2C and 2D).

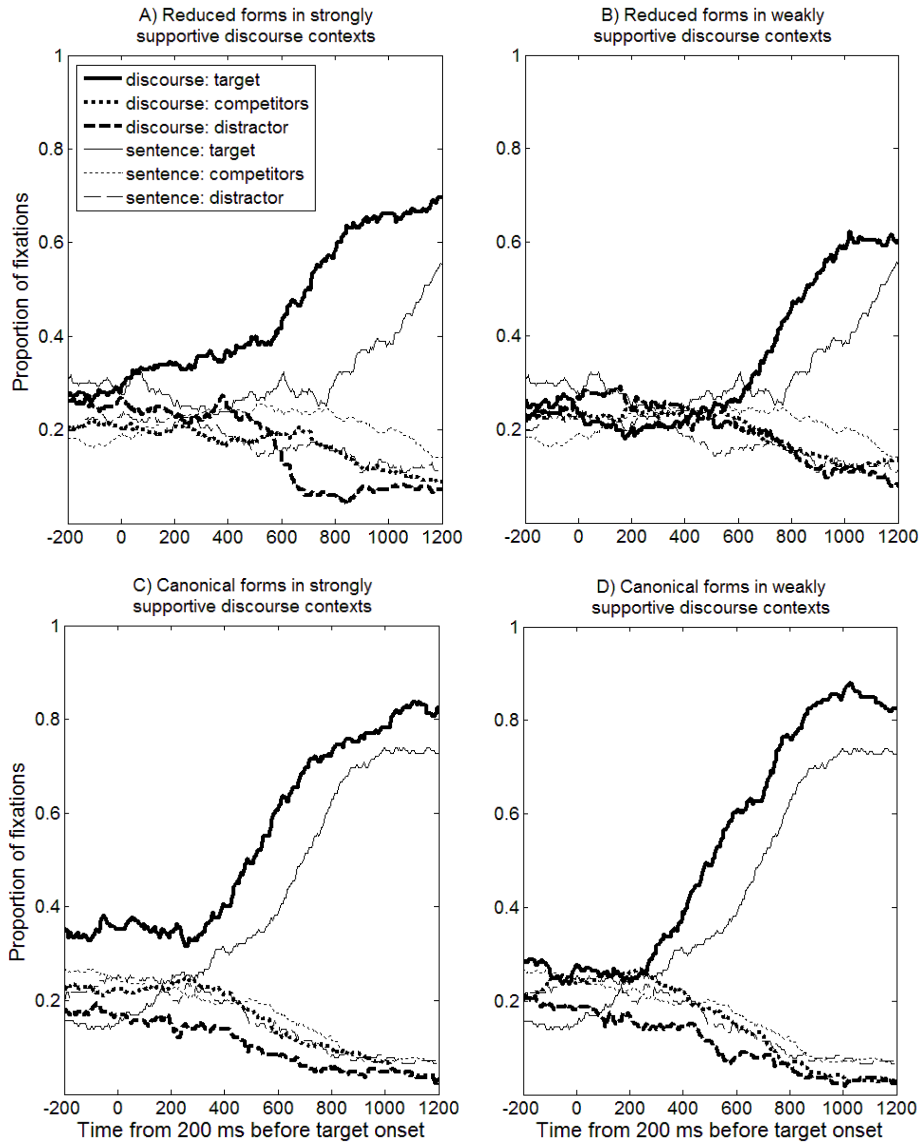



Figure 2: *Proportion of fixations over time from 200 ms before till 1200 ms after target word onset (ms) to targets, averaged competitors, and distractors for the Discourse condition (bold lines) and the Sentence only condition (thin lines) in strongly and weakly supportive discourse contexts. (A) Reduced forms in strongly supportive discourse contexts, (B) Reduced forms in weakly supportive discourse contexts, (C) Canonical forms in strongly supportive discourse contexts, and (D) Canonical forms in weakly supportive discourse contexts.*

## DISCUSSION

Figure 2 shows that participants used the discourse contexts to anticipate up-coming target words.[4] Our primary interest in the present study, however, was not the effect of wider discourse context on anticipation of the target. We were interested in the effect of wider discourse context on the online *recognition* of reduced and canonical forms. Experiment 1 showed that the presence of discourse information facilitates the recognition of canonical and reduced forms. For reduced forms only, the size of this effect was contingent upon the degree of support provided by the discourse context. Strongly supportive discourse contexts helped listeners more to recognize reduced targets than weakly supportive discourse contexts. In contrast, strongly and weakly supportive discourse contexts improved the recognition of canonical targets to a similar degree. Thus, strongly supportive discourse contexts are especially important for the recognition of reduced forms.

Additionally we see that the presence of a discourse context is beneficial in all conditions, from the "easiest" condition (canonical forms in supportive contexts) to the most difficult condition (reduced forms in less supportive contexts). However, as discussed in the Introduction, there is a potential caveat in attributing these benefits to the given discourse information. Previous research has shown that exposure to a speaker's voice is a helpful source in the recognition and adaptation to carefully pronounced canonical forms. Thus Experiment 2 was conducted to measure to what extent the context effects in Experiment 1 were not in fact effects of speaker adaptation.

In Experiment 2 we presented the same target sentences as in Experiment 1, but now they were preceded by the control contexts from our pretest. These control

[4] It takes about 200 ms to program or launch a saccadic eye movement in these types of experiments (see Altmann & Kamide, 2004 for further discussion). Fixations before 200 ms *after* the onset of the critical word are thus unlikely to be driven by acoustic information from the critical word.

contexts provided the voice of the same speaker as the one in the target sentence but no matching discourse information.

# EXPERIMENT 2

# METHOD

## Participants

Forty-eight participants from the Max Planck Institute's subject pool, undergraduates at the Radboud University Nijmegen, were paid to participate in this experiment. They did not participate in the pretest or in Experiment 1 of this chapter. They also did not take part in any of the experiments reported in Chapter 2 or 3 of this thesis. All participants were native speakers of Dutch and had normal or corrected-to-normal vision. None of the participants reported any (history of) hearing problems.

## Materials and procedure

Experiment 2 used the same target sentences as in Experiment 1, but the target sentences were preceded by the control discourse contexts of the pretest. The control discourse contexts were randomly selected from the corpus, consisted of a minimum duration of 5 seconds, and contained the same speaker as the one who spoke in the target sentence. Hence, the control discourse contexts provided speaker information but no matching discourse. All the control discourse contexts ended at an utterance boundary. Participants listened to the target sentence (sentential context only condition) or to the target sentence and the additional 'discourse' context (wider 'discourse' condition). The procedure was identical to Experiment 1.

### Design and analysis

Reduced targets were presented to half of the participants, whereas canonical targets were presented to the other half of the participants. The analyses were similar to Experiment 1.

## RESULTS and DISCUSSION

Figure 3 presents the proportions of fixations over time from -200 ms until 1200 ms after target word onset for both conditions. As in Experiment 1, we plotted the two competitors together by taking the average instead of each of them separately because the competitors did not differ significantly from each other. Given the absence of an effect of degree of support of the contexts on target fixations (see below), we did not plot the fixation proportion separately for strongly and weakly supportive discourse contexts.

In the 200-1000 ms time window we first analyzed whether participants pay more attention to the targets in one of the conditions. We found that listeners looked more often to the reduced targets ($\beta_{\text{'Discourse' Information}} = 0.91$, $p_{MCMC} < 0.001$) and to the canonical targets ($\beta_{\text{'Discourse' Information}} = 0.91$, $p_{MCMC} < 0.001$) when additional speaker information was present than when it was absent. This result demonstrates that hearing more of the same speaker facilitates the recognition of reduced and canonical targets.

Next, Degree of Support was added as a covariate to the target analysis on the reduced and the canonical forms. We found no significant interaction between Speaker Information and Degree of Support for reduced forms ($\beta_{\text{'Discourse' Information * Degree of Support}} = 0.20$, $p_{MCMC} > 0.1$) nor for canonical forms ($\beta_{\text{'Discourse' Information * Degree of Support}} = -0.02$, $p_{MCMC} > 0.5$).

In sum, the results of Experiment 2 showed that the effects of speaker adaptation were similar for canonical and reduced forms, and were not influenced by the degree of support provided by the context. This indicates that the benefits for the wider discourse condition consist of two effects: speaker adaptation and discourse information. Experiment 2 shows that speaker adaptation benefits the recognition of reduced and canonical forms to the same degree.

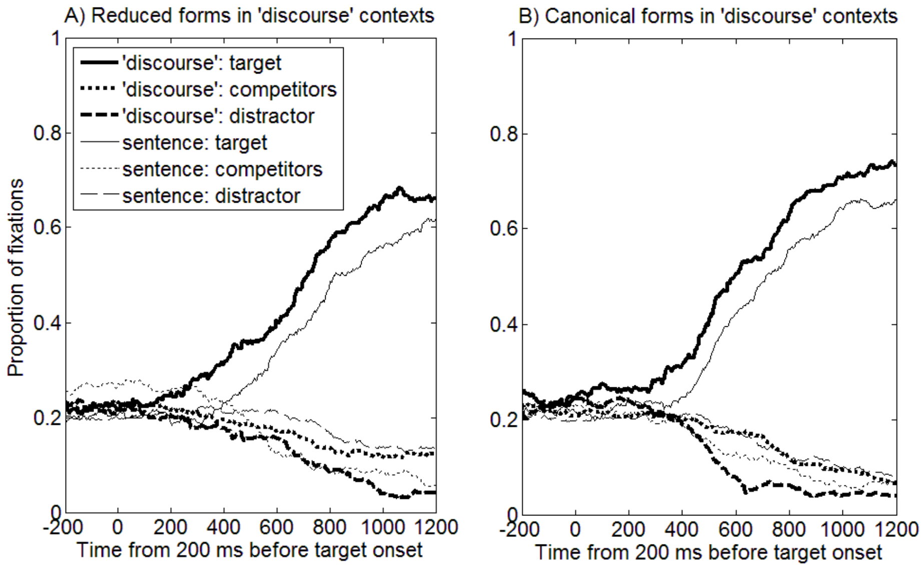



Figure 3: *Proportion of fixations over time from 200 ms before till 1200 ms after target word onset (ms) to targets, averaged competitors, and distractors for the 'Discourse' condition (bold lines) and the Sentence only condition (thin lines). (A) Reduced forms and (B) Canonical forms.*

## GENERAL DISCUSSION

We investigated the extent to which wider discourse context contributes to the recognition of reduced and canonical forms. Perhaps unsurprisingly, Experiment 1 showed that target recognition of both canonical and reduced forms improved when listeners were exposed to discourse information. This result, nevertheless, extends the findings of Ernestus et al. (2002). Ernestus and colleagues found that sentential context helps the recognition of reduced forms; here we have shown that wider discourse information helps even more. More importantly, however, we observed that strongly supportive contexts help the recognition of reduced forms more than weakly supportive contexts, a pattern that was not observed for canonical forms. For canonical forms, the degree of support by wider discourse context did not influence the efficiency of word recognition.

Experiment 2 revealed that the benefits in Experiment 1 are composed of two separate effects: a basic effect of speaker adaptation, which is similar for reduced and canonical forms, and discourse information, which differentially affects canonical and reduced forms. When comparing the results for canonical forms between Experiment 2 and Experiment 1, it is noteworthy that the magnitude of the effects is similar, about 1 logit unit. This suggests that the benefits for canonical forms seem to be largely due to speaker adaptation effects. That is, when there is a clear bottom-up signal—as is the case for our canonical forms—discourse information plays maximally a minor role in spoken-word recognition. The situation is quite different for reduced forms. Speaker information helps to recognize these forms too, but a strongly supportive context exerts an additional advantage for recognition.

Our data suggest that the presence of reduced forms in weakly supportive contexts increases the likelihood that word recognition will fail. This then offers an explanation of why speakers are more likely to use reduced forms in high predictability contexts than in contexts that are less predictable (e.g., Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Lindblom, 1990). Lieberman (1963), for example, showed that words are more carefully pronounced in unpredictable contexts than in predictable contexts (proverbs and adages). Words were generally shorter when they occurred in a highly predictable context than in an unpredictable context. This result is also in line with the so-called Probabilistic Reduction Hypothesis: words are more often reduced when the context is highly predictable (Jurafsky et al., 2001).

Similarly, Lindblom (1990) argues in his Hypo- and Hyperspeech (H&H) theory that speakers accommodate to a certain degree to listeners' communicative needs. In the H&H theory, speech production is characterized as an acoustic continuum to balance the speaker's aim to be understood and to minimize the speaker's effort by controlling the degree of reduction (hyper- and hypo- articulation) depending on the communicative context. If the listener is able to understand the message, the speaker may produce reduced speech (hypospeech), but if the listener appears to be unable to understand the message, the speaker is forced to use clear speech (hyperspeech). It should be noted, however, that our data do not allow us to conclude that speakers indeed reduce more if the discourse context strongly supports a given word. After all, our pre-test did not show that the reduced forms happened to occur in more supportive contexts than the canonical forms. Note that our study was not designed to test whether reduction is more likely if the discourse context is strongly supportive for a given word. Our results thus suggest a need for future research to explore the conditions in which words are likely to be reduced. What our data do show, however, is that with an acoustic form which is fully-realized (as with the canonical forms); a wider discourse context has little additional influence on the *recognition* of spoken words. Most interestingly, we observe a rather different pattern for reduced forms. In contrast to canonical forms, during the acoustic unfolding of the critical reduced target word, the strongly supportive wider discourse context yields an additional benefit for spoken word recognition.

In conclusion, the present study has used natural samples from a spontaneous speech corpus to investigate the extent to which wider discourse context helps the online recognition of spoken words. Our data demonstrate the importance of wider discourse context for the successful recognition of reduced forms during casual speech. A strong contextual match with the wider discourse is more important for the recognition of reduced than canonical pronunciations of spoken words in natural, communicative settings.

# SHADOWING REDUCED SPEECH AND ALIGNMENT

**CHAPTER 5**

This chapter is a slightly revised version of Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *Journal of the Acoustical Society of America, 128*(1), EL32-37.

## ABSTRACT

This study examined whether listeners align to reduced speech. Participants were asked to shadow sentences from a casual speech corpus containing canonical and reduced targets. Participants' productions showed alignment: durations of canonical targets were longer than durations of reduced targets; and participants often imitated the segment types (canonical versus reduced) in both targets. The effect sizes were similar to previous work on alignment. In addition, shadowed productions were overall longer in duration than the original stimuli and this effect was larger for reduced than canonical targets. A possible explanation for this finding is that listeners reconstruct canonical forms from reduced forms.

# INTRODUCTION

Although speech production is highly variable, listeners are most of the time able to understand what a speaker intended to say. Recently, more attention has been paid to the nature of the connection between production and perception. Pickering and Garrod (2004), for example, have argued that people align unconsciously and spontaneously to the person to whom they are speaking. Interlocutors tend to converge on a common speaking style in natural conversations (see Giles, Coupland, & Coupland, 1991, for a review). Characteristic of such natural conversations is that words are often reduced (Johnson, 2004). Such reductions may deviate from their citation form by multiple segments (e.g., [pjutər] for the Dutch word 'computer' [kɔmpjuːtər]). The present study examines whether listeners align their productions when listening to reduced speech.

Two main lines of research have investigated this production-perception link. One type of research mainly uses the shadowing task, in which participants are asked to listen and quickly repeat a speech stimulus. The type of material used in this task is typically careful speech read from a previously prepared script. Porter and Castellanos (1980), for example, used the shadowing task to measure the latency between stimulus and response onsets. In a simple version of this task, participants shadowed an extended /a/ from a model speaker and always had to switch to /ba/. In a choice version of this task, participants again shadowed the long vowel /a/, but had to switch to an unexpected CV. In both tasks, participants shadowed the targets surprisingly quickly. Porter and Castellanos argue that listeners perceive the articulations of a speaker, so that perception delivers - as a byproduct - a blueprint for production.

Fowler, Brown, Sabadini and Weihing (2003) also used the simple and choice task to investigate *what* exactly is imitated. Stop consonants were presented with short and long voice onset times (VOTs). The results showed that listeners produced longer VOTs in their shadowing responses to long VOT stimuli. This supports the idea that perceived gestures guide participants' responses and that alignment may occur at the phonetic level. However, Mitterer and Ernestus (2008) argue that a phonological approach can account for the findings of Fowler and colleagues. In their study, two variants of the Dutch /r/ were presented: uvular /r/ and alveolar /r/. These phonemes represent different gestures, but are mapped onto a similar phonological representation. The results from a shadowing task showed that participants hardly imitated the two types of /r/-stimuli but responded with their preferred variant. Further, no latency costs were found if there was a gestural mismatch between the stimulus and the response. In the same experiment, stops without or with six or twelve prevoicing cycles were presented. The gestural account predicts that the degree of prevoicing should be imitated, whereas the phonological account predicts that only the phonologically relevant presence of prevoicing should be shadowed while the amount of prevoicing, which is phonologically irrelevant in Dutch, should be ignored. The results supported the prediction of the phonological account.

In the context of the debate on the nature of lexical representations, several studies tested imitation in shadowing isolated words (e.g., Goldinger, 1998). In these studies imitation was assessed using the AXB task. In this task, listeners hear three versions of the same word and are asked to judge whether the production of stimulus A or B by a given participant is more similar to that of the model talker, X. The two stimuli from the participant were a pre-experimental baseline recording and a shadowing response. Goldinger (1998) found that listeners judged the shadowing responses to be more similar to the model talker than the productions in the baseline recording, indicating that listeners imitate the speech they hear. This study therefore provides evidence for a link between perception and production of lexical items.

A second line of research investigates whether alignment between speaker and listener also occurs in more natural communicative situations. Pickering and Garrod (2004), for example, focused on a natural form of language: the dialogue. They argue that "interlocutors align their linguistic representations at many levels ranging from the phonological to the syntactic to the semantic. This interactive alignment process is automatic and only depends on simple priming mechanisms that operate at the different levels, together with an assumption of parity of representation for production and comprehension" (p. 188). Evidence for the interactive alignment model focuses mainly on lexical and syntactic levels. Interlocutors will use the same words and syntactic structures (e.g., Branigan, Pickering, Cleland, 2000). Pardo (2006) examined alignment on a phonetic level. Different talkers had to produce similar lexical items before, during and after a conversational exchange using a map task. A different set of participants then performed the AXB task, and they judged later realizations as more similar. This indicates that participants perceived increased similarity in pronunciation between talkers over the course of the conversational interaction.

In sum, the second research line shows that convergence not only occurs in laboratory settings, but also in more natural settings. One of the main differences between the two settings is that the speech in laboratory settings is often carefully pronounced, whereas the speech we are exposed to in our daily encounters is full of reductions. Segments or even whole syllables may be deleted and/or changed into different sounds. Listeners are, however, able to understand conversational speech with ease despite these reductions. It is yet unknown whether people imitate exactly what they perceive if speech is reduced.

The present study takes an intermediate position between the two research lines: using the shadowing task to investigate the perception and the subsequent production of *conversational* speech. Participants were asked to repeat back sentences extracted from a spontaneous speech corpus. Each sentence contained one target word. Crucially, half of the target words were produced in their citation forms whereas the other half were reduced forms. If production and perception are strongly linked at the phonetic level, participants should produce exact copies of the reduced forms (e.g., listening to the Dutch pronunciation [pjutər] should produce [pjutər]). If the connection between production and perception is weak, listeners should produce similar renditions of the target words regardless of the input form (e.g., listening to the Dutch pronunciation [pjutər] may produce [kɔmpjutər]). As dependent variables, we use target word duration and the realization of the target words' segments rather than global measures of similarity as measured with an AXB task.

## METHOD

### Participants

Sixteen members of the Max Planck Institute's subject pool were paid to participate. All participants were native speakers of Dutch and reported no (history of) hearing or speech impairments. None of the participants took part in any of the experiments reported in the previous chapters of this thesis.

### Materials

Sixty-four sentences were extracted from the spontaneous speech subcorpora of the Spoken Dutch Corpus (Oostdijk, 2004). This corpus contains approximately 900 hours of speech of standard Dutch (circa 9 million words) of which 225 hours are spontaneous, real-life conversations. All recordings have been aligned with orthographic transcriptions. We searched the corpus for recordings of mid- to high-frequency words in full or in reduced form. Recordings with background noise or overlapping speech were excluded. The test materials were composed of 64 target sentences uttered by 59 different speakers. Each stimulus sentence contained one target word. Half of the target words was produced canonically (e.g., [bənedə] for *beneden* 'downwards') and the other half was produced in a reduced way (e.g., [məneə]). The same target words were used as in the experiments in Chapter 2 and 3 of this thesis (except Experiment 3.3). The average duration of both target types are presented in Table 1. Canonical targets (*M* = 490 ms; range = 329-773 ms) were significantly longer than reduced targets (*M* = 364 ms; range = 195-588 ms; $\beta_{Word\ Form}$ = -125.5, $p$ = 0.0001). Note that the context for a canonical target was never identical to that of a reduced target because they occurred in different natural corpus utterances.

### Procedure

Participants were tested individually, seated in a sound-attenuated booth in front of a computer screen. Stimuli were presented over headphones at a comfortable listening level. Participants received written instructions on the screen. They had to perform a shadowing task. They were instructed to listen to Dutch sentences and asked to repeat back the sentence as fast as possible. If they were not able to repeat back the whole fragment, they were requested to report individual words. Participants could listen to each sentence only once. Their responses were recorded digitally. The next trial initiated after 1,5 times the total duration of the fragment. For example, if the duration of the sentence was four seconds, participants had six seconds to repeat this particular sentence. A visual warning signal (a cross) appeared when the next trial initiated. Participants were presented with the 64 experimental items. The order of the items was randomized, so that each participant received a different order of presentation.

### Design and analysis

The dependent measures were error rate, duration of the shadowed target responses (for correct responses only) and type of segmental response (canonical versus reduced) to the original stimuli. For all statistical analyses, we used linear mixed effects models (Baayen, Davidson, & Bates, 2008), with participants and items as random effects. Word form was coded as a numeric contrast (-0.5 and 0.5), in which canonical forms were coded as -0.5 and reduced forms as 0.5. A logistic linking function was used for the error pattern.

# RESULTS

## Error rate

Errors consisted of target misidentifications (e.g., shadowing *presentatie* 'presentation' as a response to the stimulus *prestatie* 'performance') or no target response at all. Six participants were omitted from the final analysis because they made more than 25% of errors in the reduced form condition. The 10 remaining participants made on average 2.8% errors (0.9/32) in the canonical form condition and 22% errors in the reduced form condition (7.1/32). The statistical analysis revealed that this differences was significant ($\beta_{Word\ Form} = 2.66$, $p < 0.0001$). The positive beta indicates that participants made more errors in shadowing reduced targets than in shadowing canonical targets.

## Duration alignment

Table 1 presents the average duration of the shadowed target responses. All erroneous responses were excluded from the analysis. The duration of participants' shadowed responses to the canonical forms ($M$ = 501 ms; range = 344-673 ms) were significantly longer than to the reduced targets ($M$ = 480 ms; range = 294-731 ms; $\beta_{Word\ Form} = 24.7$, $p = 0.0001$).

A comparison between the average duration of the canonical targets and the corresponding shadowed responses showed a significant difference ($\beta_{Stim/Resp} = 7.7$, $p < 0.01$), indicating that the shadowed responses were longer than the presented canonical stimuli. A similar statistical difference was found for the average duration of the reduced targets and their shadowed responses ($\beta_{Stim/Resp} = 78$, $p = 0.0001$). Importantly, this effect was much larger for the reduced targets than for the canonical targets, and a combined analysis showed a significant interaction effect ($\beta_{Stim/Resp\ *\ Word\ Form} = 71$, $p = 0.0001$).

## Alignment to segment realizations

As a next step, we examined specific participant responses to the canonical and the reduced stimuli. The first author transcribed the target words. Spectrograms were made with the software package PRAAT (Boersma, 2001) to observe each target word in auditory and visual form. We examined whether a canonical or a reduced segment in the original stimuli remained a canonical or reduced segment in participants' responses. For example, the reduced form [məneə] consists of two reduced segments (the [m] and the [d]) and four canonical segments (the [ə], the [n], the [e], and the [ə]), whereas the canonical form [bənedə] only consists of canonical segments. We calculated how often listeners produced these segments in their original form or in another form (i.e., canonical or reduced, see Table 1).

The results show that participants produced in 88% of the cases a canonical realization and in 12% of the cases a reduced realization when listening to canonical targets. The canonical segments in the reduced targets also often remained intact (93% of the cases). Importantly, however, participants produced a canonical segment 68% of the time when reduced segments of reduced targets were presented. We used a mixed effect logistic regression model to test whether a reduced response was more likely if the stimulus was reduced as well. This was the case ($\beta_{Word\ Form}$ = -1.92, $p <$ 0.0001).[5]

[5] No differences in the extent of alignment were found between the first half and the second half of the experiment, indicating that participants did not align more to the speech stimuli over the course of the experiment.

Table 1: *Segmental responses split by stimulus and response type.*

| Stimulus type | Canonical target *(mean duration: 490 ms)* | | Reduced target *(mean duration: 364 ms)* | |
|---|---|---|---|---|
| Response duration | 501 ms | | 480 ms | |
| | target phoneme realized as | | target phoneme realized as | |
| | Canonical | Reduced | Canonical | Reduced |
| Response phoneme realized as | % | % | % | % |
| Canonical | 88 (2017) | n/a | 93 (1050) | 68 (493) |
| Reduced | 12 (280) | n/a | 7 (78) | 32 (230) |

*Note: Frequencies between parentheses.*

## DISCUSSION

We examined whether listeners align their productions when listening to reduced speech. In a shadowing task participants had to repeat sentences from a casual speech corpus containing canonical and reduced forms. The error pattern showed that canonical forms are easier to recognize than reduced forms. This is convergent with previous offline findings (e.g., Kemps et al., 2004).

Our results provide further evidence for alignment between speaker and listener. The duration data showed that participants' responses to canonical targets were longer than to reduced targets, indicating that listeners accommodate to the duration of the original form of the target. The size of the effect was similar to the results reported by Fowler et al. (2003). In Fowler et al. the VOTs in the stimuli were extended by approximately 78% (from 73 to 130 ms). Participants' responses to the extended VOTs were significantly longer than to the original VOTs, but the difference in the stimuli of a factor of about 1.78 was reduced to a difference in the responses of a factor of 1.10 (in Experiment 4A: from 61 to 69 ms; and in Experiment 4B: from 53 to 57 ms). Similarly, in our study the canonical and reduced targets differed by a factor of 1.35, but responses only differed by a factor of 1.04. The shadowed responses were approximately 4% longer for the canonical forms than for the reduced forms. Thus, both studies showed that the amount of alignment between the original extension and the shadowed extension was around 10%.

Another type of evidence consistent with the previous work on alignment comes from the analysis on segment realization. Branigan et al. (2000) showed that the syntactic structure of the confederate strongly influenced the syntactic structure of the participants, especially when participants had to use the same verb. In these cases, participants produced 55% more syntactically equivalent responses than different responses. However, when participants were asked to use a different verb than the confederate, participants produced 26% more syntactically similar responses than dissimilar responses. In a similar way, our results showed that participants produce 25% more canonical segments in response to canonical than reduced segments when listening to reduced targets. This demonstrates that the degree of alignment in our study is of a similar size as previous work on alignment of syntax.

However, our findings also show that the shadowed target responses were overall significantly longer than the duration of the original target stimuli. Critically, this effect was much bigger for the reduced targets than for the canonical targets, indicating that participants' productions show a bias towards the canonical forms. Apparently, people imitate canonical forms more closely than reduced forms. A possible explanation for the misalignment in reduced speech is that listeners reconstruct canonical forms from their reduced forms (e.g., Kemps, Ernestus, Schreuder, & Baayen, 2004). As a result, much longer responses are produced.

Two earlier studies also found evidence for "online" repair by testing how mispronunciations were shadowed (Marslen-Wilson & Welsh, 1978; Small & Bond, 1986). Misarticulated three-syllable words and words with deleted segments respectively were reconstructed on the fly by participants in a shadowing task. Despite the clear difference between the spontaneous reductions in our study and the artificially created mispronunciations and deletions in those earlier studies, the results converge on the assumption that listeners actively "reconstruct" the input in a shadowing task.

Another indication for reconstruction is that participants' responses often do not mirror the exact reductions that occurred in the original stimuli. The majority of the reduced segments in the stimuli became canonical segments in the responses. Similarly, Gaskell (2003) showed that assimilation (e.g., producing 'leam bacon' for 'lean bacon') is undone prelexically in perception on the basis of fine phonetic detail in the signal and phonological context.

What remains an open question is *to what* people align when they listen to casual utterances from a spontaneous speech corpus. There are two possibilities. First, speech may be perceived along gestural lines (Fowler et al., 2003). Participants' responses are guided by their perception of the speakers' articulatory gestures. A second possibility is that participants do not imitate gestures but rather the speech style. In this case, alignment does not target the exact phonetic properties of the input, but rather more global properties such as speaking rate, pitch range and the amount of hypo- and hyperarticulation. Neither explanation requires conscious effort due to automatic alignment.

In conclusion, our results indicate that the extent of alignment to phonological reductions is similar to the effects found in previous work on phonetic alignment (Fowler et al., 2003) and on syntactic alignment (Branigan et al., 2000). Importantly, however, our findings also suggest that the link between perception and production is weaker for reduced speech because listeners seem to reconstruct canonical forms from their reduced forms. Our study indicates that varying the amount of phonological reductions in the input is a promising avenue to further explore the relation between perception and production.

# SUMMARY AND CONCLUSIONS

**CHAPTER 6**

## Summary of results

The aim of this thesis was to examine how listeners process strongly reduced forms in casual speech. Such reductions, which occur often in natural conversations, may deviate from their canonical form by multiple segments. For example, the word *beneden* /bənedə/ 'downwards' may be realized as [məneə] by a Dutch speaker. We selected ecologically valid samples from a spontaneous speech corpus to approximate the real-life situation as closely as possible. The experiments in Chapter 2 addressed the question whether spoken word recognition during casual speech differs from spoken word recognition during laboratory speech. In the studies in Chapter 2, three eye-tracking experiments were conducted to explore this question. Listeners' eye movements were monitored as they listened to sentences with canonical (e.g., [bənedə] for *beneden* 'downwards') and reduced forms (e.g., [məneə]) and saw four printed words on the screen: a target (e.g., *beneden* 'downwards'), a competitor phonologically similar to the canonical form ("canonical form" competitor; e.g., *benadelen* 'to disadvantage'), one phonologically similar to the reduced form ("reduced form" competitor; e.g., *meneer* 'mister'), and an unrelated distractor (e.g., *vakantie* 'holiday').

In the first experiment (Experiment 2.1), listeners heard both reduced and canonical targets in sentential and syllabic contexts selected from the spontaneous speech corpus. In the second experiment (Experiment 2.2), the sentences contained canonical targets only and were extracted from the spontaneous speech corpus or recorded in the laboratory. In the third experiment (Experiment 2.3), canonical targets were intermixed with reduced targets in sentential contexts. Listeners' task was to click with the computer's mouse on the word mentioned in the speech utterance.

The results showed that if participants listened to both reduced and canonical speech (Experiment 2.1 and 2.3), their eye movements were directed to a similar degree to both phonological competitors independent of the exact acoustic target form (canonical versus reduced). However, if people listened to carefully-pronounced canonical forms only (Experiment 2.2), a clear preference for the "canonical form" competitor was found in both the laboratory and the casual speech condition. These data suggest that the listening situation influenced the spoken word recognition process. In other words, speech-intrinsic variation such as reductions affected spoken word recognition. Listeners penalized acoustic mismatches less strongly when listening to reduced speech than when listening to fully- articulated speech. It appears that, when onset information in the signal becomes less reliable, listeners depend less on it. The studies in Chapter 2 therefore provided new evidence for the flexibility of the spoken word recognition system to adjust to speech-intrinsic factors.

The experiments in Chapter 3 were set up to complement and extend the findings of Chapter 2. The question addressed was whether phonological competition during casual speech can ever be influenced by the exact phonetic form of the spoken word. The studies in Chapter 3 used a target-absent variant of the visual world paradigm which maximizes the likelihood of observing competition effects. In such a set-up, the visual display contains a "canonical form" competitor (*benadelen* 'to disadvantage'), a "reduced form" competitor (*meneer* 'mister'), and two phonologically unrelated distractors (*vakantie* 'holiday'; *juweel* 'jewel'). Participants listened to canonical (e.g., [bənedə]) and reduced forms (e.g., [məneə]) of a spoken target word (e.g., *beneden* 'downwards'). If one of the words on the screen was present in the utterance they heard, participants were to click on the word they heard. If none were present in the utterance, they were to click the centre of the screen. Test items were always this latter case, while filler items were the former.

In the first experiment (Experiment 3.1), canonical and reduced forms were presented in isolation. The results showed that, when listening to canonical forms, listeners made few errors and looked more often at the "canonical form" competitor than the "reduced form" competitor. For reduced forms, however, listeners made many errors and the competition pattern changed depending on listeners' performance on the test trials. On trials in which participants made the correct mouse click response, there was no difference in attention between competitors.

It is, however, conceivable that listeners could not exploit the fine phonetic detail in the isolated word forms. In the second experiment (Experiment 3.2), the same target words were therefore presented in sentential contexts. The results showed that in this task situation the input form still affected the eye gaze to different competitors. For canonical forms, the pattern of Experiment 3.1 was replicated: more looks to "canonical form" competitors than to "reduced form" competitors. By contrast, when hearing reduced forms, more looks were found to the "reduced form" competitor than to the "canonical form" competitor during an early time-window (400-800 ms). This result suggests that unrelated words overlapping on phonemes with the reduced form competed early during the recognition of reduced forms. This activation may underlie the delays reported for recognition of reduced forms in prior studies using offline methods. In a later time window (800-1200 ms) however, there was no difference in attention between both competitors. An interpretation for this result is that listeners reconstruct canonical forms from their reduced counterparts, a process which is time-consuming.

The final experiment (Experiment 3.3) examined to what extent fine phonetic detail plays a role in the recognition of strongly reduced forms. To explore this issue, the "surface" segments from reduced forms were replaced with "intended" segments from canonical forms. For example, the /m/ from [məneə] was cross-spliced with the /m/ from the full, intended form /met/. The data revealed, as in Experiment 3.2, that listeners looked more often at the "reduced form" competitor than the "canonical form" competitor in the early time window. However, the late time window showed an influence of the cross-splicing: the late rise of the "canonical form" competitor observed in Experiment 3.2 was absent in this experiment. These findings indicate that there are subtle acoustic differences between the cross-spliced and the original stimuli to which listeners are sensitive. The main finding of these studies is thus that, although reductions initially activate competitors which sound similar to the reduced form (in at least some task situations), listeners nevertheless can exploit fine phonetic detail to reconstruct canonical forms from their reduced forms.

The studies reported in Chapter 4 extended the findings in Chapter 2 and 3 by investigating whether the discourse context affects the recognition of reduced forms differently than the recognition of canonical forms. In particular, we tested whether a strongly supportive discourse context is more important for the recognition of reduced forms than for the recognition of canonical forms. Two eye-tracking experiments were conducted. The same four printed words as in the experiments in Chapter 2 were displayed on the screen (i.e., target: *beneden* 'downwards'; "canonical form" competitor: *benadelen* 'to disadvantage'; "reduced form" competitor: *meneer* 'mister'; and an unrelated distractor: *vakantie* 'holiday'). Participants' task was to click on the word that appeared in the target sentence.

The first experiment (Experiment 4.1) presented canonical and reduced forms in a target sentence alone or with an additional discourse context. The additional contexts were samples which directly preceded the target sentences in the spontaneous speech corpus. The second experiment (Experiment 4.2) presented the same target sentences as in Experiment 4.1, but the additional "discourse" contexts only contained information about the target speaker. These additional contexts were arbitrarily selected samples from the same target speaker. A pretest first examined whether the selected discourse samples provided supportive discourse information and whether these samples contain more useful information than the selected "discourse" (speaker-only) samples. In this pretest, participants were asked to rate how well preceding contexts matched with target sentences. The results of the pretest showed that the selected discourse samples provided informative context for listeners and that these samples matched better with the target sentences than the selected "discourse" samples. The degree of support was used as a covariate to examine whether this affects target recognition as measured by fixation proportions.

The results of Experiment 4.1 showed that the recognition of canonical and reduced forms is facilitated by discourse information. This result goes further than the findings of Experiment 2.1 and 3.2 in which it was shown that listeners benefit from sentential context to recognize reduced forms. Experiment 4.1 showed that the wider discourse context helps even more. Importantly, the degree of support mattered for reduced forms, but not for canonical forms. Strongly supportive discourse contexts helped listeners more to recognize reduced forms than weakly supportive contexts. This was not the case for canonical forms: strongly and weakly supportive discourse contexts helped the recognition of canonical forms to the same extent.

It is possible that the effects found in Experiment 4.1 were due to exposure to a speaker's voice rather than due to discourse information. Experiment 4.2 was therefore set up to examine to what extent the discourse context effects in Experiment 4.1 were in fact effects of speaker adaptation. The results of Experiment 4.2 showed that the benefits of speaker adaptation were similar for canonical and reduced forms. And, importantly, the effects were not influenced by the degree of support provided by the context. This result suggests that the effects in Experiment 4.1 consisted of two effects: discourse information and speaker adaptation. Comparing the results between Experiment 4.1 and 4.2 revealed that for canonical forms the effects are mainly due to speaker adaptation. Thus, when a word is carefully and clearly articulated, discourse information plays a smaller role in spoken word recognition than speaker information. However, for reduced forms, there is an advantage when extra speaker information is presented, but a strongly supportive discourse context exerts an additional advantage for reduced form recognition.

These data indicate that the likelihood that word recognition fails increases when reduced forms are heard in weakly supportive discourse contexts. A strong contextual match with the wider discourse is more important for the recognition of reduced than canonical pronunciations of spoken words in natural, communicative settings.

The experiments in Chapters 2 to 4 employed the visual world paradigm to focus on the question how people recognize reduced forms in casual speech and how it impacts on phonological competition. Chapter 5 reported a shadowing task to investigate the question whether listening to reduced forms affects listeners' own subsequent production. Participants' task was to listen to sentences extracted from a spontaneous speech corpus and to repeat these sentences. The target words in these corpus sentences were canonical and reduced forms. The duration of the shadowed target responses and the type of segmental response (canonical versus reduced) to the original stimuli were measured.

The results of the shadowing task showed that listeners' productions to canonical forms were overall longer than to reduced forms, providing evidence for alignment. However, the data also showed that shadowed responses were overall longer than the original stimuli. Importantly, this effect was larger for reduced than for canonical forms. An explanation for this finding was that listeners reconstruct canonical forms from reduced forms. This result was also supported by the finding that the majority of the reduced segments in the stimuli became canonical segments in the responses. The link between perception and production therefore seems to be weaker for reduced forms than for canonical forms. The data of this chapter converge with the results of Chapter 3 by providing evidence that listeners may use a reconstruction mechanism to recognize strongly reduced forms.

## Conclusions

We can now answer the four research questions that were posed in the Introduction. These were:

1. *Does spoken word recognition during casual speech differ from laboratory speech?*

The results of Chapter 2 show that listeners penalize acoustic mismatches less strongly when hearing reduced than laboratory speech. This demonstrates that listening to reduced forms influences the dynamics of spoken word recognition. This is an important finding because it shows how listeners deal with speech variability across different listening conditions. Spoken word recognition during casual speech thus differs from spoken word recognition during laboratory speech. However, this does not necessarily weaken the established results with laboratory speech because the process underlying speech recognition remains essentially the same. In both listening situations, the spoken word recognition system seems to operate in an optimal way. Thus, studies on casual speech do provide insight into how flexible the spoken word recognition system really is, which laboratory studies alone would not be able to reveal.

Note that popular models of spoken word recognition differ in their assumptions about how mismatches are dealt with. The TRACE model (McClelland & Elman, 1986) consists of an interactive-activation network in which candidate words can be activated by any part of the speech signal. For instance, the model predicts that the onset sounds of the word 'cap' will activate the candidate 'cap' and the candidate 'gap' because of their overlap in features (i.e., /k/ and /g/). In TRACE, there is thus no explicit penalty for a mismatch. However, in Shortlist (Norris, 1994), the activation of a word candidate is determined by the degree to which it matches and mismatches the speech signal. A mismatch between input and canonical form leads to stronger deactivation of a lexical candidate. The current results may be taken to indicate that both models need to be revised because neither is able to adjust the activation strategy. As our results indicate, the spoken word recognition system is flexible in some situations, it tolerates – just as the TRACE model – mismatches. In other situations, however, the system acts like Shortlist and does not tolerate mismatches. An important task for the new generation of spoken word recognition models such as Shortlist B (Norris & McQueen, 2008) is that they need to account for the listeners' apparent flexibility in dealing with good and poor input.

An interesting line for future work is how the interplay between the intelligibility of speech and the amount of noise influence spoken word recognition. For example, how do people understand each other if their speech is highly degraded and they speak in a very noisy situation such as in a bar? This research would be able to reveal how flexible the spoken word recognition system is in response to a constantly changing speech signal and to different listening situations. Such data are bound to be very important for more realistic models of spoken word recognition than the current models that are largely based on laboratory speech in noise-free conditions.

2. *Which phonological competitors take part in the competition process when listeners hear strongly reduced forms?*

The results in Chapter 3 show that (in certain task situations) strongly reduced forms in casual speech can initially activate competitors which are similar to the phonological surface form of the reduction, but that listeners nevertheless can exploit fine phonetic detail to reconstruct canonical forms from reduced forms.

How does this result fit with the theoretical accounts of how listeners recognize pronunciation variants? Two different accounts have been proposed in the literature. On one account, a reconstruction process operates at a prelexical level, which mediates between the speech signal and the lexicon on the basis of fine phonetic detail in the signal, phonological context (e.g., Gaskell, 2003), or by top-down restoration (McClelland & Elman, 1986; Warren, 1970). On the other account, phonological variants are assumed to be stored in the mental lexicon. Two different versions of this account exist. According to the episodic view, the entry for a given word in the mental lexicon consists of detailed and concrete episodic memories of pronunciations of that word that have been encountered previously (e.g., Bybee, 2001; Goldinger, 1998; Hawkins, 2003; Johnson, 1997; Pierrehumbert 2001), whereas the other lexical-storage account argues that different pronunciation variants are stored as abstract phonological forms (e.g., Connine, 2004; McLennan, Luce, & Charles-Luce, 2003; Ranbom & Connine, 2007). A comparison between the models and the current data suggests that a reconstruction mechanism is likely to be involved. There are two pieces of evidence that support this conclusion. First, listeners pay most attention to the competitor that is more similar to the canonical form late in time. Second, fine phonetic detail in the reduced forms influences this late competition pattern. However, it remains an open question whether reduced forms are stored in the mental lexicon or not. Further research is therefore required to examine the exact contributions of the mechanisms which help listeners to recognize reduced forms.

3. *Does discourse context affect the recognition of reduced forms differently than the recognition of canonical forms?*

The results in Chapter 4 show that a strong contextual match with the wider discourse is more important for the recognition of reduced than for the recognition of canonical pronunciations of spoken words in natural, communicative settings. Spoken word recognition is more likely to fail when reduced forms are pronounced in weakly supportive discourse context, which might explain why speakers have a tendency to produce reduced forms more often in high than in low predictability contexts (e.g., Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Lindblom, 1990). For canonical forms, the overall benefits seemed to be largely due to speaker adaptation effects with an only very limited role for discourse context. Further research could usefully be directed at this issue. Another interesting follow-up study would be to investigate the conditions in which reduced forms are more likely to occur.

*4. Do listeners accommodate to reduced forms in their own subsequent production?*

The data in Chapter 5 provided evidence that listeners accommodate to a certain extent to reduced forms in conversational speech (cf. Branigan, Pickering, & Cleland, 2000; Fowler, Brown, Sabadini, & Weihing, 2003). Importantly, however, the results also suggested that the link between perception and production is weaker for reduced forms than for canonical forms because listeners seemed to reconstruct canonical forms from reduced forms. Further exploration of this topic would be interesting to gain more insight into the link between perception and production. For example, do listeners also align to less severe phonological reductions and is the type of instruction (repeat back versus imitate) of influence on the size of alignment?

The final empirical chapter also provided an interesting data point regarding the issue of whether the recognition of a reduced form involves some kind of reconstruction process. The findings from the shadowing task converged with the eye-tracking results of Chapter 3 by providing evidence that listeners may use a reconstruction mechanism to recognize reduced forms.

This thesis took an important step towards bridging the gap between tightly-controlled laboratory studies and real-world speech communication. More specifically, the research presented in this thesis provided new insights into the processing of strongly reduced forms in casual speech. It is important for future psycholinguistic work to pay more attention to the influence of speech reductions on spoken word recognition. Such an increased focus on casual speech phenomena would almost certainly result in more realistic models of spoken word recognition.

# REFERENCES

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Altmann, G.T.M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and visual world. In J. Henderson & F. Ferreira (Eds.) *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press (pp. 347-389).

Altmann, G.T.M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502-518.

Altmann, G.T.M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye-movements and mental representation. *Cognition, 111*, 55-71.

Arai, T. (1999). A case study of spontaneous speech in Japanese. *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, San Francisco (pp. 615-618).

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Barr, D.J. (2008). Analyzing 'visual world' eye tracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*(2), 1001-1024.

Boersma, P. (2001). PRAAT, a system for doing phonetic by computer. *Glot International, 5*(9/10), 341-345.

Bradlow, A.R., Nygaard, L.C., & Pisoni, D.B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61* (2), 206-219.

Branigan, H.P., Pickering, M.J., & Cleland, A.A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13-B25.

Brouwer, S., Mitterer, H., & Huettig, F. (under revision). Speech reductions change the dynamics of spoken word recognition.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

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

Connine, C.M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review, 11*(6), 1084-1089.

Connine, C.M., Blasko, D.G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193-210.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Cutler, A. (1998). The recognition of spoken words with variable representation. *Proceedings of the ESCA Workshop on Sound Pattern of Casual Speech*. Aix-en-Provence: ESCA (pp. 83-92).

Dahan, D., & Tanenhaus, M.K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*, 453-459.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*, 447-456.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface*. Utrecht: LOT.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162-173.

Fowler, C.A., Brown, J., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49,* 396-413.

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

Gaskell, M.G. & Marslen-Wilson, W.D. (1996). Phonological variation and interference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 144-158.

Gaskell, M.G., & Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception andPerformance, 24*(2), 380-396.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press (pp. 1-68).

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

Goldinger, S.D., Luce, P.A., & Pisoni, D. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*(5), 501-518.

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

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

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

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Phonetics, 31*, 373-405.

Huettig, F., & Altmann, G.T.M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.). *The on-line study of sentence comprehension: eyetracking, ERP and beyond.* New York: Psychology Press (pp. 187-207).

Huettig, F., & Altmann, G.T.M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*, B23-B32.

Huettig, F., & Altmann, G.T.M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*(8), 985-1018.

Huettig, F., & Altmann, G.T.M. (in press). Looking at anything that is green when hearing 'frog' - How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology.*

Huettig, F., & McQueen, J.M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*, 460-482.

Huettig, F., & McQueen, J.M. (2009). AM radio noise changes the dynamics of spoken word recognition. Paper presented at the AMLaP 2009 conference in Barcelona, Spain.

Huettig, F., Rommers, J., & Meyer, A.S. (submitted). The visual world paradigm: What it does and doesn't tell us about linguistic processing.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J.W. Mullennix (Eds.). *Talker Variability in Speech Processing*. San Diego: Academic Press (pp. 145-165).

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.). *Casual Speech: Data and Analysis. Proceedings of the 1st session of the 10th International Symposium*. Tokyo, Japan: The National International Institute for Japanese Language (pp. 29-54).

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W.D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. Bybee & Paul Hopper (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins (pp. 229-254).

Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156.

Kamide, Y., Scheepers, C., & Altmann, G.T.M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*, 37-55.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 117-127.

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

Kraljic, T., & Samuel, A.G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51,* 141-178.

Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*(1), 51-81.

Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

Ladefoged, P., & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 27*, 98-104.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172-187.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.). *Speech production and speech modeling.* Dordrecht: Kluwer (pp. 403-439).

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during spoken word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Matin, E., Shao, K.C., & Boff, K.R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics, 53*, 372-380.

McLennan, C., Luce, P.A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 539-553.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McQueen, J.M., & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes, 16*(5/6), 469-490.

McQueen, J.M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology, 60*(5), 661-671.

Mehl, M.R., Vazire, S., Ramírez-Esparza, N., Slatcher, R.B., & Pennebaker, J.W. (2007). Are women really more talkative than men? *Science, 317,* 82.

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in casual and read speech. *Language and Speech, 31*(2), 135-156.

Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica, 63,* 209-229.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics, 65*(6), 956-969.

Mitterer, H., Csépe, V., Honbolygo, F, & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*, 451-479.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition, 109*, 168-173.

Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365-378.

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

Norris, D.G., & McQueen, J.M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.

Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter, 5*, 4-8.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382-2393.

Phaf, R.H., Van der Heijden, A.H.C., & Hudson, P.T.W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273-341.

Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral Brain Sciences, 27*, 169-225.

Pierrehumbert, J. (2001) Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.). *Frequency Effects and the Emergence of Lexical Structure*. Amsterdam: John Benjamins (pp. 137-157).

Pitt, M.A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*, 19-36.

Porter, R., & Castellanos, F. (1980). Speech production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America, 67*, 1349-1356.

Ranbom, L.J., & Connine, C.M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*, 273-298.

Samuel, A. (1996). Phoneme restoration. *Language and Cognitive Processes, 11*, 647-653.

Shockey, L. (2003). Sound patterns of spoken English. Oxford: Blackwell Publishing.

Small, L.H., & Bond, Z.S. (1986). Distortions and deletions: Word-initial consonant specificity in fluent speech. *Perception and Psychophysics, 40*(1), 20-26.

Snoeren, N.D., Gaskell, M.G., & Di Betta, A.M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 542-549.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language*, *27*, 324-340.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49,* 161-167.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32*, 359-372.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension, *Science, 268*, 1632-1634.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 443-467.

Van der Heijden, A.H.C., Hagenaar, R., & Bloem, W. (1984). Two stages in postcategorical filtering and selection. *Memory and Cognition, 12*, 458-469.

Warner, N. (To appear). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. Huffman (Eds.). *Oxford Handbook of Laboratory Phonology.*

Warner, N. (submitted). Reduction. In M. Ewen, C. Hume, & K. Rice (Eds.) *The Blackwell Companion to Phonology*.

Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical representations in Spanish: Evidence from eye movements. In J. Trouvain & W.J. Barry (Eds.) *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*. Dudweiler: Pirrot (pp. 1941-1944).

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken- word processing. *Cognition, 32*, 25-64.

# APPENDIX

Experimental items: canonical and reduced realizations with their "canonical form" and "reduced form" competitors respectively.

| Target word | Canonical form | "Canonical form" competitor | Reduced form | "Reduced form" competitor |
|---|---|---|---|---|
| afspraak 'appointment' | [ɑfspra:k] | [ɑfspre:kə] | [ɑspʋa:] | [ɑspi:ra:tsi:] |
| apparaat 'apparatus' | [ɑpa:ra:t] | [ɑpəri:ti:f] | [ɔpra:t] | [ɔpra:pə] |
| beneden 'downwards' | [bəne:də] | [bəna:de:lə] | [məne:ə] | [məne:r] |
| bijvoorbeeld 'for example' | [bəvo:rbe:lt] | [bəvo:rɛχtə] | [vɔlt] | [vɔlt] |
| computer 'computer' | [kɔmpju:tər] | [kɔmpətɛnt] | [pju:tər] | [pu:tsə] |
| concert 'concert' | [kɔnsɛrt] | [kɔnjɑk] | [kɔsɛr] | [kɔsdba:r] |
| concurrent 'competitor' | [kɔnky:rɛnt] | [kɔnku:r] | [kɔ̃krɛnt] | [kɔnkre:t] |
| constant 'constant' | [kɔnstɑnt] | [kɔnsɛntra:tsi:] | [kɔzən] | [ko:zɛin] |
| cultuur 'culture' | [kʏlty:r] | [kʏltʏs] | [kɔmty:m] | [kɔmst] |
| december 'December' | [de:sɛmbər] | [de:ka:n] | [e:sɛmər] | [e:ta:ʒə] |
| dinsdag 'Tuesday' | [dɪnzdɑχ] | [dɪŋə] | [dɪzɑ] | [di:zɑjn] |

| Target word | Canonical form | "Canonical form" competitor | Reduced form | "Reduced form" competitor |
|---|---|---|---|---|
| directeur 'director' | [di:rɛktø:r] | [di:ri:χe:rə] | [dɪktø:] | [dɪkta:tɔr] |
| kweekschool 'school' | [kʋe:ksχo:l] | [kʋe:kə] | [kʋe:sχo:l] | [kʋe:stə] |
| maandag 'Monday' | [ma:ndɑχ] | [ma:nt] | [ma:nz] | [ma:nza:t] |
| ogenblik 'moment' | [o:χəblɪk] | [o:χkɑs] | [blɪk] | [blɪk] |
| oktober 'October' | [ɔkto:bər] | [ɔkto:pʏs] | [to:ʋər] | [to:vərə] |
| overheid 'government' | [o:vərhɛit] | [o:vərhɛmt] | [o:vərɛi] | [o:vərɛint] |
| parlement 'parliament' | [pɑrləmɛnt] | [pɑrke:rɛ] | [pɑləmɛn] | [pa:lət] |
| plaatsen 'to place' | [pla:tsə] | [pla:tsna:m] | [pla:s] | [pla:se:bo:] |
| positie 'position' | [po:zi:tsi:] | [po:ze:rə] | [psi:tsi:] | [psi:χə] |
| prestatie 'performance' | [prɛsta:tsi:] | [prɛsti:ʒə] | [pəsta:si:] | [pɛsi:mɪst] |
| principe 'principle' | [prɪnsi:pə] | [prɪns] | [pəsi:pə] | [pərso:n] |
| publiek 'audience' | [py:bli:k] | [py:bli:se:rə] | [ʋli:k] | [ʋi:l] |
| redelijk 'reasonable' | [re:dələk] | [re:dərɛi] | [re:lək ] | [re:li:kʋi:] |

| Target word | Canonical form | “Canonical form” competitor | Reduced form | “Reduced form” competitor |
|---|---|---|---|---|
| rekenen ‘to count’ | [re:kənə] | [re:ks] | [re:χən] | [re:χə] |
| rotzooi ‘garbage’ | [rɔtso:j] | [rɔts] | [rɔsɪ:] | [rɔsəχ] |
| standaard ‘default’ | [stɑnda:rt] | [stɑndpla:ts] | [stɑ̃əd] | [stɑŋ] |
| standpunt ‘point of view’ | [stɑntpʏnt] | [stɑntfɑstəχ] | [stɑmpʏ] | [stɑmpɔt] |
| station ‘station’ | [statʃɔn] | [sta:tʏs] | [sa:ʃɔn] | [sa:tɛin] |
| tandarts ‘dentist’ | [tɑndɑrts] | [tɑndpasta:] | [tɑz] | [tɑs] |
| wedstrijd ‘match’ | [ʋɛtstrɛit] | [ʋɛtbu:k] | [ʋɛs] | [ʋɛsp] |
| winter ‘winter’ | [ʋɪntər] | [ʋɪntstɪl] | [ʋɪndə] | [ʋɪndə] |

# SAMENVATTING EN CONCLUSIES

## Samenvatting van de resultaten

Het doel van dit proefschrift was te onderzoeken hoe luisteraars sterk gereduceerde woorden in alledaagse spraak verwerken. Zulke reducties, die vaak in natuurlijke conversaties voorkomen, kunnen met meerdere segmenten afwijken van hun canonieke vorm. Bijvoorbeeld, het woord *beneden* /bənedə/ kan uitgesproken worden als [məneə] door een Nederlandse spreker. Wij selecteerden ecologisch valide fragmenten uit een database met spontane spraak om de dagelijkse taalsituatie zo dicht mogelijk te benaderen. De experimenten in hoofdstuk 2 onderzochten of het woordherkenningsproces in spontane spraak afwijkt van het woordherkenningsproces in zorgvuldig uitgesproken (of laboratorium) spraak. In dit hoofdstuk werden drie oogbewegingsexperimenten uitgevoerd om deze vraag te beantwoorden. De oogbewegingen van luisteraars werden gemeten, terwijl zij luisterden naar zinnen met canonieke (bijvoorbeeld [bənedə] voor *beneden*) en gereduceerde (bijvoorbeeld [məneə]) vormen. Tegelijkertijd werden vier geschreven woorden op het computerscherm afgebeeld: een doelwoord (bijvoorbeeld *beneden*), een concurrent die fonologisch meest gelijk was aan de canonieke vorm ("canonieke vorm" concurrent; bijvoorbeeld *benadelen*), een concurrent die fonologisch meest gelijk was aan de gereduceerde vorm ("gereduceerde vorm" concurrent; bijvoorbeeld *meneer*) en een ongerelateerd woord (bijvoorbeeld *vakantie*)

In het eerste experiment (experiment 2.1) hoorden luisteraars zowel gereduceerde als canonieke doelwoorden in zins- en syllabische contexten. Deze werden uit de database met spontane spraak geselecteerd. In het tweede experiment (experiment 2.2) werden zinnen aangeboden die alleen canonieke doelwoorden bevatten. Deze zinnen werden uit de database geselecteerd of opgenomen in een geluidsdichte cabine. In het derde experiment (experiment 2.3) werden canonieke en gereduceerde doelwoorden in zinscontexten aangeboden. De taak voor de luisteraars was met de computermuis op het geschreven woord te klikken dat in het spraakfragment voorkwam.

De resultaten lieten zien dat, als proefpersonen zowel naar gereduceerde als canonieke vormen luisterden (experimenten 2.1 en 2.3), hun oogbewegingen in dezelfde mate naar de twee fonologische concurrenten keken. Dit was niet afhankelijk van het precieze akoestische doelwoord (canoniek ten opzichte van gereduceerd). Maar als proefpersonen alleen zorgvuldig uitgesproken canonieke vormen hoorden (experiment 2.2) was er een duidelijke voorkeur om naar de "canonieke vorm" concurrent te kijken in zowel de nette spraak als in de spontane spraak conditie. Deze bevindingen laten zien dat de luistersituatie het woordherkenningsproces beïnvloedt. Met andere woorden, spraakintrinsieke variatie, zoals gereduceerde vormen, beïnvloedt het woordherkenningsproces. Luisteraars straffen akoestische incongruenties minder sterk af als er geluisterd wordt naar gereduceerde spraak dan wanneer er geluisterd wordt naar zorgvuldig uitgesproken spraak. Het blijkt dat wanneer het geluidssignaal minder betrouwbaar is, luisteraars er minder afhankelijk van zijn. De experimenten in hoofdstuk 2 verstrekken dus nieuw bewijs dat het woordherkenningsproces zich flexibel aanpast aan spraakintrinsieke factoren.

De experimenten in hoofdstuk 3 werden opgezet om de bevindingen in hoofdstuk 2 aan te vullen en uit te breiden. De vraag die gesteld werd in hoofdstuk 3 was of fonologische competitie in spontane spraak (ooit) beïnvloed kan worden door de precieze akoestische vorm van het gesproken woord. De studies in hoofdstuk 3 maakten gebruik van een variant van het visuele wereld paradigma ('visual world paradigm') waarbij het doelwoord niet visueel werd afgebeeld op het computerscherm, zodat de waarschijnlijkheid om competitie-effecten te vinden toeneemt. In deze opstelling bevat het computerscherm een "canonieke vorm" concurrent (*benadelen*), een "gereduceerde vorm" concurrent (*meneer*) en twee fonologisch ongerelateerde woorden (*vakantie* en *juweel*). Proefpersonen luisterden naar canonieke (bijvoorbeeld [bənedə]) en gereduceerde (bijvoorbeeld [məneə]) vormen. Als een van de woorden op het computerscherm aanwezig was in het spraakfragment, dan moesten proefpersonen op het woord klikken dat ze hoorden (items die afleiden). Als geen van de woorden aanwezig was in het geluidsfragment, dan moesten zij in het midden van het computerscherm klikken (test items)

In het eerste experiment (experiment 3.1) werden canonieke en gereduceerde vormen zonder context gepresenteerd. De resultaten toonden aan dat wanneer er geluisterd werd naar de canonieke vormen, luisteraars slechts enkele fouten maakten en meer naar de "canonieke vorm" concurrent keken dan naar de "gereduceerde vorm" concurrent. Luisteraars maakten echter veel fouten wanneer er naar de gereduceerde vormen geluisterd werd. Bovendien was het competitiepatroon afhankelijk van de prestaties van de proefpersonen op de test items. Als proefpersonen correct klikten (dus in het midden van het scherm), was er geen verschil in aandacht tussen de concurrenten.

Het is mogelijk dat de luisteraars geen gebruik konden maken van de gedetailleerde fonetische informatie, omdat de doelwoorden zonder context werden aangeboden. In het tweede experiment (experiment 3.2) werden de doelwoorden daarom in zinscontexten gepresenteerd. De resultaten lieten zien dat met deze taak (het doelwoord was afwezig op het computerscherm) het oogbewegingspatroon beïnvloedt wordt door de akoestische vorm. Voor de canonieke vormen werd het patroon in experiment 3.2 gerepliceerd: er gingen meer oogbewegingen naar de "canonieke vorm" concurrent dan naar de "gereduceerde vorm" concurrent. Echter, voor gereduceerde vormen werden er meer oogbewegingen gevonden naar de "gereduceerde vorm" concurrent dan naar de "canonieke vorm" concurrent gedurende een vroeg tijdsvenster (400-800 ms). Dit resultaat laat zien dat ongerelateerde woorden, die fonemen delen met de gereduceerde vorm, geactiveerd worden tijdens de herkenning van gereduceerde vormen. Het is mogelijk dat deze activatie ten grondslag ligt aan de vertragingen zoals gerapporteerd in vorige studies die gebruik maakten van offline methoden om het herkenningsproces van gereduceerde vormen in kaart te brengen. In een later tijdsvenster (800-1200 ms) werd er echter geen verschil gevonden in aandacht voor de beide concurrenten. Een mogelijke interpretatie voor dit resultaat is dat luisteraars gereduceerde vormen reconstrueren tot canonieke vormen. Dit is een tijdrovend proces.

Het laatste experiment (experiment 3.3) onderzocht in welke mate gedetailleerde akoestische informatie een rol speelt in de herkenning van sterk gereduceerde vormen. Om dit te onderzoeken werden de daadwerkelijke gerealiseerde segmenten van de gereduceerde vormen vervangen door "bedoelde" segmenten van de canonieke vormen. Bijvoorbeeld de /m/ van [məneə] werd vervangen door de /m/ van de canonieke vorm /met/. De resultaten lieten zien dat, net zoals in experiment 3.2, luisteraars meer naar de "gereduceerde vorm" concurrent keken dan naar de "canonieke vorm" concurrent in het vroege tijdsvenster. Het late tijdsvenster liet echter een invloed zien van de spraakmanipulatie: de late toename van de "canonieke vorm" concurrent, zoals waargenomen in experiment 3.2, was afwezig in dit experiment. Deze bevindingen tonen aan dat er subtiele akoestische verschillen zijn tussen de gemanipuleerde en de originele segmenten waar luisteraars gevoelig voor zijn. De belangrijkste vondst van deze studies is dus dat, alhoewel reducties eerst concurrenten activeren die hetzelfde klinken als gereduceerde vormen (in deze taaksituatie), luisteraars desalniettemin gebruik maken van gedetailleerde akoestische informatie om gereduceerde vormen te reconstrueren tot canonieke vormen.

De experimenten beschreven in hoofdstuk 4 breidden de bevindingen in hoofdstuk 2 en 3 verder uit. In dit hoofdstuk werd onderzocht of de wijdere context ('discourse') de herkenning van gereduceerde vormen anders beïnvloedt dan de herkenning van canonieke vormen. In het bijzonder werd er onderzocht of een sterk ondersteunende discourse context belangrijker is voor de herkenning van gereduceerde vormen dan voor de herkenning van canonieke vormen. Er werden twee oogbewegingexperimenten uitgevoerd. Dezelfde vier geschreven woorden als in de experimenten in hoofdstuk 2 werden op het computerscherm aangeboden (i.e., doelwoord: *beneden*; "canonieke vorm" concurrent: *benadelen*; "gereduceerde vorm" concurrent: *meneer*; en een ongerelateerd woord: *vakantie*). De taak voor de proefpersonen was op het woord te klikken dat in het geluidsfragment voorkwam.

Het eerste experiment (experiment 4.1) bood canonieke en gereduceerde vormen aan in een zinscontext of met een extra discourse context. De extra discourse contexten waren geluidsfragmenten die direct vooraf gingen aan de zinscontexten in de spontane spraak database. Het tweede experiment (experiment 4.2) presenteerde dezelfde zinscontexten als in experiment 4.1, maar de extra "discourse" contexten bevatte alleen informatie over de doelspreker. Deze extra "discourse" contexten werden willekeurig gekozen uit de database, maar bevatte wel de stem van de doelspreker. Een pretest onderzocht eerst of de geselecteerde discourse fragmenten voldoende discourse informatie verstrekten en of deze fragmenten meer nuttige informatie bevatte dan de "discourse" (alleen de spreker was hetzelfde) fragmenten. Proefpersonen werden in deze pretest gevraagd of zij wilden aangeven hoe goed de voorafgaande contexten pasten bij de zinscontexten. De resultaten lieten zien dat de geselecteerde discourse fragmenten informatieve context verstrekten voor luisteraars en dat deze fragmenten beter pasten bij de zinscontexten dan de "discourse" fragmenten. Hoe goed de twee fragmenten bij elkaar pasten werd gebruikt als covariaat om te onderzoeken hoe dit het herkennen van het doelwoord beïnvloedt zoals gemeten in fixatie proportie.

De resultaten van experiment 4.1 lieten zien dat de herkenning van canonieke en gereduceerde vormen wordt versneld door de aanwezigheid van discourse informatie. Dit resultaat gaat verder dan de bevindingen in experiment 2.1 en 3.2 waar werd aangetoond dat luisteraars gebruik maken van de zinscontext om gereduceerde vormen te herkennen. Experiment 4.1 liet zien dat discourse context zelfs meer helpt. Nog belangrijker is dat de mate van discourse ondersteuning van invloed was op de herkenning van gereduceerde vormen, maar niet voor de herkenning van canonieke vormen. Goed passende discourse contexten hielpen luisteraars meer om gereduceerde vormen te herkennen dan slecht passende discourse contexten. Dit was niet het geval voor canonieke vormen: sterke en zwakke discourse contexten hielpen de herkenning van canonieke vormen in dezelfde mate.

Het is mogelijk dat de effecten, zoals gevonden in experiment 4.1, werden veroorzaakt door de blootstelling aan de stem van de doelspreker in plaats van door de discourse informatie. Experiment 4.2 werd daarom opgezet om te onderzoeken in welke mate de discourse context effecten in experiment 4.1 eigenlijk sprekeradaptatie effecten zijn. De resultaten van experiment 4.2 lieten zien dat de effecten van extra sprekerinformatie hetzelfde waren voor canonieke en gereduceerde vormen. De effecten werden bovendien niet beïnvloed door de mate van discourse ondersteuning zoals verstrekt door de extra context. Dit resultaat laat zien dat het effect in experiment 4.1 bestaat uit twee effecten: discourse informatie en sprekerinformatie. Het vergelijken van de resultaten tussen experiment 4.1 en 4.2 toonde aan dat de effecten voor canonieke vormen voornamelijk veroorzaakt zijn door het aanpassen aan de spreker. Dus wanneer een woord zorgvuldig wordt uitgesproken speelt discourse informatie een kleinere rol in het woordherkenningsproces dan sprekerinformatie. Voor gereduceerde vormen daarentegen is er een voordeel wanneer extra sprekerinformatie wordt gepresenteerd, maar een sterk ondersteunende discourse context oefent een extra voordeel uit voor de herkenning van gereduceerde vormen.

Deze bevindingen laten zien dat de waarschijnlijkheid dat het woordherkenningsproces faalt, toeneemt wanneer gereduceerde vormen in zwakke discourse ondersteunende contexten worden uitgesproken. Een sterke contextuele overeenkomst met de discourse is dus belangrijker voor de herkenning van gereduceerde dan canonieke vormen in natuurlijke, communicatieve situaties.

De experimenten in hoofdstuk 2, 3 en 4 maakten gebruik van de visual world paradigm om te onderzoeken hoe mensen gereduceerde vormen in spontane spraak herkennen en hoe dit van invloed is op fonologische competitieprocessen. Hoofdstuk 5 rapporteerde een schaduwtaak ('shadowing task') om te onderzoeken of het luisteren naar gereduceerde vormen van invloed is op luisteraars' eigen, opeenvolgende productie. De taak voor proefpersonen was om te luisteren naar zinnen die uit de spontane spraakdatabase werden gehaald en om deze zinnen vervolgens te herhalen. De doelwoorden in deze zinnen waren canoniek of gereduceerd uitgesproken. De duur van de geschaduwde doelwoorden en de segmentale reacties (canoniek of gereduceerd) op de originele doelwoorden werden gemeten.

De resultaten van deze taak lieten zien dat de producties van luisteraars voor canonieke vormen over het algemeen langer waren dan voor gereduceerde vormen. Dit verstrekt bewijs voor 'alignment'; aanpassing aan het spraaksignaal. Maar de resultaten lieten ook zien dat de geschaduwde reacties over het algemeen langer waren dan de originele doelwoorden. Dit effect was groter voor de gereduceerde dan voor de canonieke vormen. Een verklaring voor deze bevinding is dat luisteraars de gereduceerde vormen reconstrueren tot canonieke vormen. Deze interpretatie wordt ook ondersteund door de bevinding dat het merendeel van de gereduceerde segmenten in de originele doelwoorden canoniek werden in de geschaduwde reacties. De connectie tussen perceptie en productie lijkt daarom zwakker te zijn voor gereduceerde vormen dan voor canonieke vormen. De resultaten van dit hoofdstuk komen overeen met de resultaten van hoofdstuk 3 door aan te tonen dat luisteraars inderdaad gebruik lijken te maken van een reconstructiemechanisme om sterk gereduceerde vormen te herkennen.

## Conclusies

We kunnen nu de vier onderzoeksvragen beantwoorden die in de introductie gesteld werden. Dit waren:

1. *Verschilt het woordherkenningsproces in spontane spraak van het woordherkenningsproces in zorgvuldig uitgesproken spraak?*

De resultaten in hoofdstuk 2 tonen aan dat luisteraars akoestische incongruenties minder sterk afstraffen wanneer zij luisteren naar gereduceerde spraak dan wanneer zij luisteren naar zorgvuldig uitgesproken spraak. Dit laat zien dat het luisteren naar gereduceerde vormen het woordherkenningsproces beïnvloedt. Dit is een belangrijke bevinding aangezien het laat zien hoe luisteraars omgaan met spraakvariatie in verschillende luistersituaties. Het woordherkenningsproces in spontane spraak verschilt dus van het woordherkenningsproces in laboratoriumspraak. Dit betekent echter niet dat de gevestigde resultaten met zorgvuldig uitgesproken spraak afgezwakt moeten worden, omdat het proces onderliggend aan woordherkenning voornamelijk hetzelfde blijft. In beide luistersituaties probeert het woordherkenningsproces op de meest optimale manier te werk te gaan. Dus studies met spontane spraak geven inzicht in de vraag hoe flexibel het woordherkenningsproces is. Dit is een vraag die laboratoriumstudies alleen niet kunnen beantwoorden.

Populaire modellen van het woordherkenningsproces verschillen in hun aannames hoe er omgegaan wordt met incongruenties. Het TRACE model (McClelland & Elman, 1986) bestaat uit een interactief activatienetwerk waarin woordkandidaten geactiveerd kunnen worden door elk deel van het spraaksignaal. Het model voorspelt bijvoorbeeld dat de beginklanken van het woord *krap* de kandidaat *krap* en de kandidaat *grap* activeert door de overeenkomst in kenmerken (i.e., /k/ en /g/). In TRACE is er dus geen expliciete afstraffing voor een incongruentie. In

Shortlist (Norris, 1994) daarentegen wordt de activatie van woordkandidaten bepaald door de mate waarin de kandidaten overeenkomen en niet overeenkomen met het spraaksignaal. Een mismatch tussen het spraaksignaal en de canonieke vorm leidt tot sterke deactivatie van een lexicale kandidaat. De huidige resultaten lijken te laten zien dat beide modellen moeten worden aangepast, aangezien beide modellen niet in staat zijn om de activatiestrategie aan te passen. Zoals onze resultaten laten zien, is het woordherkenningsproces flexibel in sommige situaties, het tolereert - zoals in het TRACE model - incongruenties. In anders situaties echter lijkt het woordherkenningsproces meer op Shortlist en tolereert geen incongruenties. Een belangrijke taak voor de nieuwe generatie woordherkenningsmodellen zoals Shortlist B (Norris & McQueen, 2008) is het integreren van spraakvariatie.

Een interessant idee voor toekomstig onderzoek is hoe de interactie tussen de verstaanbaarheid van spraak en de mate van geluidsachtergrond het woordherkenningsproces beïnvloedt. Hoe verstaan mensen elkaar bijvoorbeeld als hun spraak erg verstoord is en als zij zich op een erg drukke plek bevinden zoals in een bar? Dit soort onderzoek zou in staat moeten zijn om uit te vinden hoe flexibel het woordherkenningsproces is als reactie op het constant veranderende spraaksignaal en op verschillende luistersituaties. Zulke bevindingen zullen erg belangrijk zijn voor de ontwikkeling van meer realistische modellen van het woordherkenningsproces in vergelijking met de huidige modellen die voornamelijk gebaseerd zijn op laboratoriumspraak in lawaaivrije omstandigheden.

2. *Welke fonologische concurrenten nemen deel aan het competitieproces wanneer luisteraars sterk gereduceerde vormen horen?*

De resultaten in hoofdstuk 3 laten zien dat (in bepaalde taaksituaties) sterk gereduceerde vormen in spontane spraak concurrenten kunnen activeren die hetzelfde klinken als de fonologische vorm van de reductie, maar dat luisteraars desalniettemin gebruik maken van gedetailleerde fonetische informatie om gereduceerde vormen te reconstrueren tot canonieke vormen.

Hoe past dit resultaat met de theoretische benaderingen die verklaren hoe luisteraars spraakvariatie herkennen? Twee verschillende benaderingen zijn voorgesteld in de literatuur. De eerste benadering gaat er vanuit dat een reconstructieproces aan het werk is op een prelexicaal niveau, dat medieert tussen het spraaksignaal en het lexicon op basis van gedetailleerde fonetische informatie in het spraaksignaal, de fonologische context (bijv. Gaskell, 2003), of door top-down reconstructie (McClelland & Elman, 1986; Warren, 1970). De andere benadering gaat er vanuit dat elke spraakvariant wordt opgeslagen in het mentale lexicon. Er bestaan twee verschillende versies van deze benadering. Volgens de episodische versie bestaat ieder woord in het mentaal lexicon uit gedetailleerde en concrete episodische uitspraken die eerder zijn waargenomen (bijv. Bybee, 2001; Goldinger, 1998; Hawkins, 2003). De andere versie beargumenteert dat verschillende spraakvariaties opgeslagen zijn als abstracte fonologische vormen (bijv. Connine, 2004; McLennan, Luce, & Charles-Luce, 2003; Ranbom & Connine, 2007). Een vergelijking tussen de modellen en de huidige bevindingen toont aan dat het zeer waarschijnlijk is dat een reconstructiemechanisme betrokken is bij de herkenning van sterk gereduceerde vormen. Er zijn twee bewijzen die deze conclusie ondersteunen. Ten eerste geven luisteraars meer aandacht aan de concurrent die het meest gelijk is aan de canonieke vorm in het late tijdsvenster. Ten tweede wordt het late competitiepatroon beïnvloed door de gedetailleerde fonetische informatie in de gereduceerde vormen. Het blijft echter een open vraag of gereduceerde vormen ook daadwerkelijk opgeslagen zijn in het mentale lexicon. Vervolgonderzoek is vereist om te onderzoeken wat de exacte bijdrage is van de mechanismes die luisteraars helpen om gereduceerde vormen te herkennen.

*3. Beïnvloedt de 'discourse' context de herkenning van gereduceerde vormen anders dan de herkenning van canonieke vormen?*

De resultaten in hoofdstuk 4 lieten zien dat een sterk ondersteunende 'discourse' context belangrijker is voor de herkenning van gereduceerde vormen dan voor de herkenning van canonieke vormen in natuurlijke omstandigheden. De kans is daarom groot dat het woordherkenningsproces faalt wanneer luisteraars gereduceerde vormen in een zwak ondersteunende 'discourse' context horen. Dit zou kunnen verklaren waarom sprekers de neiging hebben om gereduceerde vormen meer in voorspelbare dan onvoorspelbare contexten uit te spreken (bijv. Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Lindblom, 1990). Voor de canonieke vormen lijken de voordelen voornamelijk door extra sprekerinformatie te komen en speelt de 'discourse' context een kleinere rol. Het zou interessant zijn om deze bevinding verder te onderzoeken. Een andere idee voor vervolgstudies is om te onderzoeken onder welke condities gereduceerde vormen het meest voorkomen.

4. *Passen luisteraars zich aan gereduceerde vormen aan in hun eigen opeenvolgende productie?*

De bevindingen in hoofdstuk 5 lieten zien dat luisteraars zich in zekere mate aanpassen aan gereduceerde vormen in spontane spraak (zie ook Branigan, Pickering, & Cleland, 2000; Fowler, Brown, Sabadini, & Weihing, 2003). De resultaten toonden echter ook aan dat de connectie tussen perceptie en productie zwakker is voor gereduceerde vormen dan voor canonieke vormen, omdat luisteraars gereduceerde vormen proberen te reconstrueren tot canonieke vormen. Verdere verkenning van dit onderwerp zou interessant zijn om meer inzicht te krijgen in de connectie tussen perceptie en productie. Zullen luisteraars zich bijvoorbeeld ook aanpassen aan minder sterke fonologische reducties en is het instructietype (herhalen of imiteren) ook van invloed op de mate van aanpassing aan het spraaksignaal?

Het laatste empirische hoofdstuk liet ook zien dat er een reconstructiemechanisme betrokken was bij het herkennen van gereduceerde vormen. De bevindingen van de schaduwtaak kwamen overeen met de oogbewegingsresultaten in hoofdstuk 3. Luisteraars bleken in beide taken gebruik te maken van een reconstructiemechanisme om sterk gereduceerde vormen te herkennen.

Dit proefschrift heeft een belangrijke stap genomen om de kloof tussen gecontroleerde laboratoriumstudies en de natuurlijke, echte communicatie te dichten. Het onderzoek in dit proefschrift heeft nieuwe inzichten gegeven in het verwerken van sterk gereduceerde vormen in spontane spraak. Het is belangrijk voor toekomstig psycholinguïstisch onderzoek om meer aandacht te besteden aan de invloed van spraakreducties op het woordherkenningsproces. Een toename van onderzoek naar spontane spraakfenomenen zal ongetwijfeld in meer realistische woordherkenningsmodellen resulteren.

# CURRICULUM VITAE

Susanne Brouwer was born on October 19, 1981 in Zaanstad, the Netherlands. She studied Psychology and Linguistics at Utrecht University, the Netherlands. In 2004, she received her MSc degree in Psychology. In 2006, she graduated from the research master's program in Linguistics (cum laude). In the same year, she was awarded a 3-year scholarship from the Max Planck Society to do her PhD research at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. She joined the Language Comprehension Group. In 2008, she was awarded a grant from the Hugh Knowles Leadership Fund, which allowed her to spend 6 months as visiting scholar at Northwestern University in Evanston, IL (USA). Currently, she is working as a postdoctoral researcher at the Linguistics Department of Northwestern University, where she joined the Speech Communication Research Group.

# MPI SERIES IN PSYCHOLINGUISTICS

1. *The electrophysiology of speaking. Investigations on the time course of semantic, syntactic and phonological processing*
   Miranda van Turennout

2. *The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsaggital articulography*
   Niels O. Schiller

3. *Lexical access in the production of ellipsis and pronouns*
   Bernadette M. Schmitt

4. *The open-/closed-class distinction in spoken-word recognition*
   Alette Haveman

5. *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach*
   Kay Behnke

6. *Gesture and speech production*
   Jan-Peter de Ruiter

7. *Comparative intonational phonology: English and German*
   Esther Grabe

8. *Finiteness in adult and child German*
   Ingeborg Lasser

9. *Language input for word discovery*
   Joost van de Weijer

10. *Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe*
    James Essegbey

11. *Producing past and plural inflections*
    Dirk Janssen

12. *Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea*
    Anna Margetts

13. *From speech to words*
    Arie van der Lugt

14. *Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language*
    Eva Schultze-Berndt

15. *Interpreting indefinites: An experimental study of children's language comprehension*
    Irene Krämer

16. *Language specific listening: The case of phonetic sequences*
    Andrea Weber

17. *Moving eyes and naming objects*
    Femke van der Meulen

18. *Analogy in morphology: The selection of linking elements in Dutch compounds*
    Andrea Krott

19. *Morphology in speech comprehension*
    Kerstin Mauth

20. *Morphological families in the mental lexicon*
    Nivja H. de Jong

21. *Fixed expressions and the production of idioms*
    Simone A. Sprenger

22. *The grammatical coding of postural semantics in Goemai*
    Birgit Hellwig

23. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies*
    Fermin Moscoso del Prado Martin

24. *Contextual influences on spoken-word processing*
    Daniëlle van den Brink

25. *Perceptual relevance of prevoicing in Dutch*
    Petra M. van Alphen

26. *Syllables in speech production: Effects of syllable preparation and syllable frequency*
    Joana Cholin

27. *Producing complex spoken numerals for time and space*
    Marjolein Meeuwissen

28. *Morphology in auditory lexical processing*
    Rachèl J.J.K. Kemps

29. *At the same time ...: The expression of simultaneity in learner varieties*
    Barbara Schmiedtová

30. *A grammar of Jalonke argument structure*
    Friederike Lüpke

31. *Agrammatic comprehension: An electrophysiological approach*
    Marlies Wassenaar

32. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*
    Frank Seifart

33. *Prosodically-conditioned detail in the recognition of spoken words*
    Anne Pier Salverda

34. *Phonetic and lexical processing in a second language*
    Mirjam Broersma

35. *Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension*
    Oliver Müller

36. *Lexically-guided perceptual learning in speech processing*
    Frank Eisner

37. *Sensitivity to detailed acoustic information in word recognition*
    Keren B. Shatzman

38. *The relationship between spoken word production and comprehension*
    Rebecca Özdemir

39. *Disfluency: Interrupting speech and gesture*
    Mandana Seyfeddinipur

40. *The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation*
    Christiane Dietrich

41. *Cognitive cladistics and the relativity of spatial cognition*
Daniel B.M. Haun

42. *The acquisition of auditory categories*
Martijn Goudbeek

43. *Affix reduction in spoken Dutch*
Mark Pluymaekers

44. *Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence*
Valesca Kooijman

45. *Space and iconicity in German Sign Language (DGS)*
Pamela Perniss

46. *On the production of morphologically complex words with special attention to effects of frequency*
Heidrun Bien

47. *Crosslinguistic influence in first and second languages: Convergence in speech and gesture*
Amanda Brown

48. *The acquisition of verb compounding in Mandarin Chinese*
Jidong Chen

49. *Phoneme inventories and patterns of speech sound perception*
Anita Wagner

50. *Lexical processing of morphologically complex words: An information-theoretical perspective*
Victor Kuperman

51. *A grammar of Savosavo: A Papuan language of the Solomon Islands*
Claudia Wegener

52. *Prosodic structure in speech perception and production*
Claudia Kuzla

53. *The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension*
Sarah Schimke

54. *Studies on intonation and information structure in child and adult German*
Laura de Ruiter

55. *Processing the fine temporal structure of spoken words*
    Eva Reinisch

56. *Semantics and (ir)regular inflection in morphological processing*
    Wieke Tabak

57. *Processing strongly reduced forms in casual speech*
    Susanne Brouwer