THE RELATIONSHIP BETWEEN SPOKEN WORD PRODUCTION AND COMPREHENSION

ISBN-10: 90-9020741-4
ISBN-13: 978-90-9020741-4

Cover Illustration: Adolph Fotografie, Duisburg, Germany
Cover Design: Ponsen & Looijen BV, Wageningen
Printed and bound by Ponsen & Looijen BV, Wageningen

THE RELATIONSHIP BETWEEN SPOKEN WORD PRODUCTION AND COMPREHENSION

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 26 juni 2006
des namiddags om 1.30 uur precies

door

REBECCA ÖZDEMIR

geboren op 1 augustus 1979
te Duisburg, Duitsland

ACKNOWLEDGEMENTS

I was always looking forward to writing the acknowledgements, because this means that the thesis is finally done. But now that the situation is here it is much harder than I thought.

First of all, of course, I am truly greatful to my two supervisors, Pim Levelt and Ardi Roelofs. They gave me all the support I needed from my first day at the institute until the last letter of this thesis. I could always count on Ardi when I was having problems or needed input to get further with my work. And, Ardi, you were such a great help with submitting and revising papers during the last months when my time was limited and my head was full of other things. I don't know what I would have done without your great help. But most importantly, Pim's and Ardi's understanding and support in personally hard times was real and their optimism always helped me feel better and get back to work soon. I really could not have whished for better supervisors. Thank you so much.

A big thank you also goes to the whole Utterance Encoding Group. Although more and more people left during the last 3 years, I enjoyed the time in the group and the lunches with the girls in the beginning of my time at the institute. I would especially like to thank Joana Cholin for introducing me to many people and the habits at the institute, and Heidrun Bien for being my buddy until there was no one left on the floor but the two of us ... and sorry that I left you all alone on the floor so often during the last months. You became a friend and I was always happy to see you and steel cookies, fruit or cappuccino pulver from your office. I also have to really thank Evelyn Giering for organising everything with the university and the printer in the last phase of this thesis when things had to go quick and I was not able to do it. I could not have made it without her.

I also would like to thank Anne Cutler and James McQueen for giving me a second home in the Comprehension Group. It was really needed for the topic to get input from experts in comprehension, and this was the place to go. I was always welcome to join the meetings, learn more about recognition processes, and present my work and get really sophisticated and helpful feedback. Thanks to all the members of the group for their comments and suggestions. Furthermore I would like to thank James for being there to talk about ideas, data and models in small extra-meetings.

I would like to say thank you also to the people from the technical group, who made work at this institute really a pleasure. A especially great thank you goes to Ad Verbund, for organising the technical stuff to give me the opportunity to work when my hands were not usable.

I would also like to thank the student assistants for lending me their beautifully clear Dutch voices, which made it possible to get the effects. I am also thankful that Lisa van den Berg, who helped me cutting stimuli for hours. Since my Dutch is still not as good as it should be, I don't know what I would have done without Annelie Tuinman, who helped me translating my summary to Dutch. The thesis would probably not have been ready for years.

The first thank you not concerning work goes to Leah Roberts, who made my life in Nijmegen in between the weekends in Duisburg a nice time by keeping me alive with broccoli pasta and watching stupid sitcoms and Emergency Room in the evenings. You were a really good housemate and became a good friend ... and in the end my proofreader and my paranimfe!

Die Personen allerdings, denen ich am meisten danken möchte, sind meine Familie und meine Freunde. Und dies ist auch der schwierigste Teil des "Danke Sagens", denn man kann schwer in Worte fassen, was sie mir bedeuten. Sie waren (und sind es immernoch) jederzeit für mich da in harten Zeiten, wir hatten viel Spass in guten Zeiten, und ich kann jederzeit auf sie zählen, egal zu welcher Tages- oder Nachtzeit. Diese Dissertation wäre nicht möglich gewesen ohne die Unterstützung und Hilfe von Mama, Papa, Rapha, Oma und Opa. Ein ganz besonders grosses Danke geht an meine Schwester Sue, die meine beste Freundin ist, und auf die ich immer zählen kann. Auch vor ein paar Wochen, als ich eine zweite Paranimfe gesucht habe...die sie jetzt ist! Ein grosses Dankeschön geht auch an meine Schwiegerfamilie und besonders meine Schwiegermutter, die auf mein kleines Mädchen aufgepasst hat und mir alle Hausarbeit abgenommen hat, als die letzten Schliffe dieser Arbeit etwas Ruhe und Zeit erforderten. Und schliesslich, das Leben wäre wesentlich weniger lustig und schön ohne die besten Freunde auf der Welt, Gabi & Sascha, Sandra & Hasan, und Mareen & Fabian. Vielen Vielen Dank, dass es euch alle gibt!

Die letzte Person, der ich danken möchte, ist gleichzeitig auch die Wichtigste, Ümit. Du warst meine erste Liebe als wir noch in der Schule waren, wurdest mein Ehemann, und der Vater unserer kleinen Tochter Leyla. Ich danke Dir von ganzem Herzen für deine Liebe, Unterstützung, und deinen Glauben an mich. Du und die Kleine seid mein Leben, und Ihr bedeutet mir alles. Ich liebe Dich!

Rebecca.

CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

INTRODUCTION

CHAPTER I

In psycholinguistics, speech production and speech comprehension have both been intensively investigated in the past. This has led to the development of detailed theories and computationally implemented models of spoken word recognition, such as SHORTLIST (Norris, 1994) and TRACE (McClelland & Elman, 1986), and of spoken word production, such as DSMSG (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) and WEAVER++ (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997, 2004a). The relationship between the recognition and production of spoken words, however, has received surprisingly little attention (see Roelofs, 2003, for a review). This is particularly striking, since both lines of research distinguish between levels of phonological features, segments, and lexical forms. Moreover, in both research domains, grammatical and conceptual levels of processing are assumed. These commonalities lead to the question of how the systems for speech comprehension and production are related.

The evidence available in the literature concerning this question comes from different sources. Researchers have examined this question by looking at impairments, at chronometrical data, and in the last decade also at neuroimaging data. They have tried to identify brain areas which are responsible for either speech production or comprehension, or for both modalities. Looking at impairments, they have tried to find patients who show an impairment in one modality but not in the other (see Martin & Saffran, 2002, for a review), which would suggest a separation of processing structures. The healthy brain was investigated with neuroimaging techniques (fMRI, PET), where researchers compared activation in the brain in response to a comprehension or a production task. Activation in the same areas was taken as indicating shared processing systems, whereas a separation of activated brain structures was taken as evidence for a separation of processing structures (see Coleman, 1998; Indefrey & Levelt, 2004, for reviews). Chronometrical evidence exists in the form of studies investigating the relationship between comprehension and production directly, or in the form of comparisons of studies from the literature investigating either production or comprehension (see Monsell, 1987; Roelofs, 2003, for reviews). The available evidence, however, is rare and inconclusive with respect to the question of whether the processing systems for speech production and comprehension are shared or separated. There is also no consensus as to what levels the systems are connected at and how they influence each other. The aim of the present thesis is to shed light on these issues with a chronometrical approach.

THE WORKING MODEL

A working model is needed, since it is hard to conduct good experiments without the predictions of a model. Whereas most implemented models concentrate either on spoken word recognition (e.g., Norris, 1994; McClelland & Elman, 1986) or on spoken word production (e.g., Dell et al., 1997), the WEAVER++ model has implemented claims about both production and its relation to comprehension (Levelt et al., 1999; Roelofs, 1997, 2003, 2004a). Word planning is implemented as a staged process moving from conceptual preparation (including the conceptual identification of a pictured object in picture naming) via lemma retrieval (recovering the word as a syntactic entity, including its syntactic properties such as its grammatical gender, which is crucial for the use of the word in phrases and sentences) to word-form encoding, as illustrated in Figure 1.1. WEAVER++ assumes two different lexical levels, namely levels of lemmas and lexical forms. Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification. In the model, concepts and lemmas are shared between production and comprehension, whereas there are separate input and output representations of word forms.

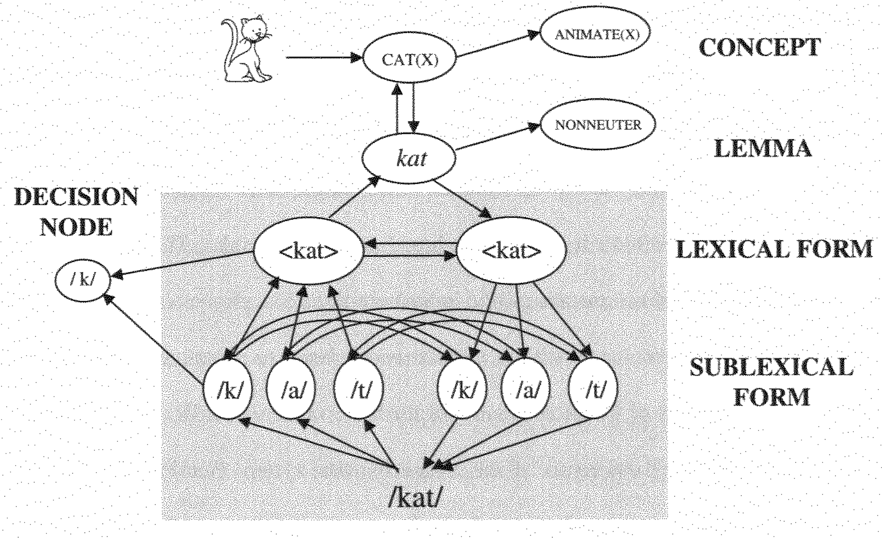


Figure 1.1 *The* WEAVER++ *model combined with the* MERGE *model for performing phoneme monitoring tasks. Word planning is implemented as a staged process moving from conceptual preparation, via lemma retrieval, to word-form encoding (right side). Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification (left side). Decision nodes are used in performing metalinguistic judgement tasks like phoneme monitoring.*

WEAVER++ implements a number of specific claims about how the spoken word production and comprehension systems interact. One of the key observations from the research on spoken word recognition is that as speech unfolds, multiple word candidates become partially activated and compete for selection. For example, the initial speech fragment /kæ/ activates a cohort of words including *cat*, *can*, *captain* and *captive* (e.g., Marslen-Wilson & Welsh, 1978; Zwitserlood, 1989). This multiple activation concerns not only the forms but also the syntactic properties and meanings of the words. Computationally implemented models of spoken word recognition, such as SHORTLIST (Norris, 1994) and TRACE (McClelland & Elman, 1986), all instantiate this insight in one form or another. WEAVER++ also assumes that heard speech activates such cohorts of word candidates, both at the lemma level and at the level of lexical forms (Roelofs, 1997). Moreover, the model assumes sublexical connections between the comprehension and production systems.

MERGE (Norris et al. 2000) is a model that is based on the SHORTLIST model for spoken word recognition (Norris, 1994). It is a module based on decision nodes. These nodes form a specific mechanism to perform metalinguistic judgement tasks like phoneme monitoring. The WEAVER++ model was combined with the MERGE model, because phoneme monitoring will be used in the experiments of Chapters 3 and 4 of the thesis. Decision nodes are connected to the phonemic and lexical form layers of the comprehension system (see left side of Figure 1.1). Imagine the following situation, a participant hears a word (e.g., *cat*) via headphones and has to press a button when a certain phoneme (e.g., *t*) is present in the word. The phonemes are decoded and as soon as the *t* is heard the phoneme node sends activation to the decision node. The lexical form node of *cat* sends activation to the decision node as well. As soon as the decision node reaches a certain threshold the phoneme is detected and the response can be made.

The working model for this thesis is thus a combination of established models for speech processing, WEAVER++ and MERGE. Moreover, it makes testable predictions concerning the relationship between speech comprehension and production.

The comprehension system is not only activated by listening to an interlocutor's utterance. It is also activated by the speaker listening to his own speech. Moreover, comprehension-based accounts of self-monitoring of speech assume that the comprehension system is also used to monitor an inner representation of the utterance, the speech plan. According to the perceptual-loop theory (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999), the speech plan is fed into the comprehension system prior to articulation for checking purposes, and is processed there like external speech. This can be achieved by using the sublexical links from production to comprehension assumed in the working model (see Figure 1.1). Imagine an internal phoneme monitoring task, in which participants have to press a button when the picture name contains a prespecified phoneme. After seeing the picture, the production process traverses from conceptual identification via lemma retrieval to phonological encoding. The speech plan is then fed into the comprehension system. Internal phoneme monitoring would be performed on the representations in the comprehension system like the standard phoneme monitoring task (described above). This would not require additional processing routines or structures.

There is a second type of monitoring account, namely a production-based account. According to such a view (e.g., Laver, 1973; Schlenk, Huber, & Wilmes, 1987) there are several special purpose monitors, which are directly attached to the production processing stages. Each of these monitors checks only one function, for example phoneme selection. An internal phoneme monitoring task under this view would be performed differently. The monitoring device (e.g., the decision nodes) would be directly connected to the production phoneme nodes. The response of the decision nodes would thus directly depend on activation from production processes.

The crucial difference between the two types of monitoring accounts is that the perceptual-loop theory assumes separate form representations for perception and production processes.

The production-based monitor on the other hand can also work in an interactive speech processing model, with form representations which are shared by both speech modalities. In contrast to the perceptual-loop theory, the production-based monitor does not predict perception specific processing steps and mechanisms. This difference can be used to distinguish between the monitoring accounts and to find evidence for the sharing or separation of representations for speech production and comprehension.

OUTLINE OF THE THESIS

The guiding research questions of this thesis are the following:

1. Do speech perception and production influence each other?
2. If so, at what levels of processing do they influence each other?
3. Are there separate representations for perception and production at the form levels of processing?

These questions were investigated in three series of experiments (Chapters 2-4). All series examined the form levels of processing. Chapters 2 and 3 addressed the connecting links between the sublexical and lexical form levels of production and comprehension. The question was whether and how ongoing processing in one modality influences the performance of a task in the other modality. Chapter 4 investigated whether there are shared or separate representations for production and comprehension at the level of phonological representations.

In Chapter 2, the influence of ongoing perception processing on a typical production task was examined. The question as to whether there are sublexical and lexical form links running from comprehension to production, as assumed in the model (see Figure 1.1), was tested. A picture-word interference study was conducted in which participants had to name pictures while simultaneously listening to auditorily presented distractor words. These distractor words were phonologically related to the picture name either at the beginning of the word (sharing onset and nucleus) or at the end of the word (sharing nucleus and coda). Earlier research has shown that phonological overlap between the spoken distractor and the picture name influences performance (e.g., Schriefers & Teruel, 1999; Meyer & Schriefers, 1991). The question was whether there would be a difference in effects between cohort and rhyme overlap, which would indicate a lexical contribution. To determine the exact locus of the effect, two other tasks were introduced (gender decision and animacy decision). Thereby influences from higher processing levels were controlled. To look at the timing of the influence, three SOAs were tested. The distractor was presented 150 msec before picture presentation, 150 msec after picture presentation, or distractor and picture were presented simultaneously.

Chapter 3 investigated the sublexical links in the other direction, namely from speech production to speech comprehension. Participants performed a standard phoneme monitoring task. This task was adapted to fit the purposes of investigating the influence of ongoing speech production planning on this speech comprehension task. Picture primes were shown on the screen before presenting a spoken word over headphones. The question was whether the encoding of a picture name containing the target phoneme immediately before hearing the auditorily presented target would speed up phoneme recognition in the comprehension system.

Chapter 4 addressed the question of whether the influences examined in the first two experimental series arise because of connections between the systems or whether it is rather one shared system serving both modalities. To answer this question, an extensive self-monitoring experiment was conducted. Participants performed a phoneme monitoring task again, but this time not on external speech signals but on their inner speech plan. Pictures were presented and participants were asked to press a button when the picture name contained a prespecified phoneme. The question was whether it is possible to find comprehension specific effects. The uniqueness point effect (Frauenfelder, Segui, & Dijkstra, 1990) was chosen, because it is found in standard phoneme monitoring. The uniqueness point is the segment in a word where it diverges from all other words in the language starting with the same initial phoneme sequence. Frauenfelder et al. (1990) and several other comprehension studies have shown that phonemes after the uniqueness point are identified faster than phonemes before the uniqueness point. Since the uniqueness point does not play a role in speech production it is interesting to test whether it shows up in a self-monitoring task. The perceptual-loop theory predicts such a comprehension-specific effect in production monitoring due to the comprehension-specific processing of the speech plan after it has been fed into the comprehension system. The production-internal monitor on the other hand is bound to production-internal processing, which should not show an effect of uniqueness.

Chapter 5 gives an overview of the results found in Chapters 2-4 and discusses them in light of the working model and other speech processing models.

THE LOCUS OF PHONOLOGICAL FACILITATION FROM SPOKEN DISTRACTORS IN PICTURE NAMING

CHAPTER 2

REBECCA ÖZDEMIR, ARDI ROELOFS, & WILLEM J. M. LEVELT[1]

ABSTRACT

Two experiments are reported that aimed to determine the functional locus of phonological facilitation from spoken distractors in picture naming. Dutch participants named pictures, determined whether the pictured entities were animate or not, and determined the grammatical gender of the picture names. Trials were blocked by task. In Experiment 1, spoken distractor words were presented that shared the beginning or the end of the phonological form of the picture names, or they were unrelated. Distractor onset was 150 msec before or after picture onset, or there was no onset difference. Phonological relatedness speeded up picture naming, but it had no effect on the animacy decision or gender decision. With distractor preexposure, the facilitation was larger for end- than for begin-relatedness in picture naming. In Experiment 2, spoken distractors were presented that were begin-related or unrelated word fragments or whole words. A phonological facilitation effect was obtained in picture naming and gender decision, but not in animacy decision. The facilitation was larger for the fragments than for the words in picture naming but not in gender decision. These results suggest that, for the most part, the phonological facilitation effect arises at the levels of lexical and sublexical forms in spoken word planning, possibly with a small contribution from the lemma level.

---

[1] A version of this chapter is submitted for publication.

INTRODUCTION

The production of spoken words and their recognition have been intensively investigated in psycholinguistics during the past three decades. The psycholinguistic research, focusing independently either on speech recognition or production, has lead to the development of detailed computationally implemented models of spoken word recognition, such as SHORTLIST (Norris, 1994) and TRACE (McClelland & Elman, 1986), and of spoken word production, such as DSMSG (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) and WEAVER++ (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997, 2004a). However, the relationship between the recognition and production of spoken words has received surprisingly little attention (see Roelofs, 2003, for a review). Yet, an examination of the recognition and production literatures reveals that both lines of research distinguish between levels of phonological features, segments, and lexical forms. Moreover, in both research domains, grammatical and conceptual levels of processing are assumed. This raises the question of how the recognition and production systems are related.

Whereas most implemented models concentrate either on spoken word recognition (e.g., Norris, 1994; McClelland & Elman, 1986) or on spoken word production (e.g., Dell et al., 1997), the WEAVER++ model has implemented claims about both production and its relation with comprehension (Levelt et al., 1999; Roelofs, 1997, 2003, 2004a). Word planning is implemented as a staged process moving from conceptual preparation (including the conceptual identification of a pictured object in picture naming) via lemma retrieval (recovering the word as syntactic entity, including its syntactic properties such as grammatical gender, which is crucial for the use of the word in phrases and sentences) to word-form encoding, as illustrated in Figure 2.1. WEAVER++ assumes two different lexical levels, namely the levels of lemmas and lexical forms. Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification. In the model, concepts and lemmas are shared between production and comprehension, whereas there are separate input and output representations of word forms.

WEAVER++ implements a number of specific claims about how the spoken word production and comprehension systems interact. One of the key observations from the research on spoken word recognition is that as speech unfolds, multiple word candidates become partially activated and compete for selection. For example, the initial speech fragment /kæ/ activates a cohort of words including *cat*, *can*, *captain* and *captive* (e.g., Marslen-Wilson & Welsh, 1978; Zwitserlood, 1989). The multiple activation concerns not only the forms but also the syntactic properties and meanings of the words. Computationally implemented models of spoken word recognition, such as SHORTLIST (Norris, 1994) and TRACE (McClelland & Elman, 1986), all instantiate this insight in one form or another. WEAVER++ also assumes that heard speech activates such cohorts of word candidates, both at the lemma level and at the level of lexical forms (Roelofs, 1997). Moreover, the model assumes sublexical connections between the comprehension and production systems. Evidence for the existence of both lexical and sublexical form connections comes from picture-word interference experiments (Meyer & Schriefers, 1991; Roelofs, 2002; Schriefers & Teruel, 1999; Starreveld, 2000).

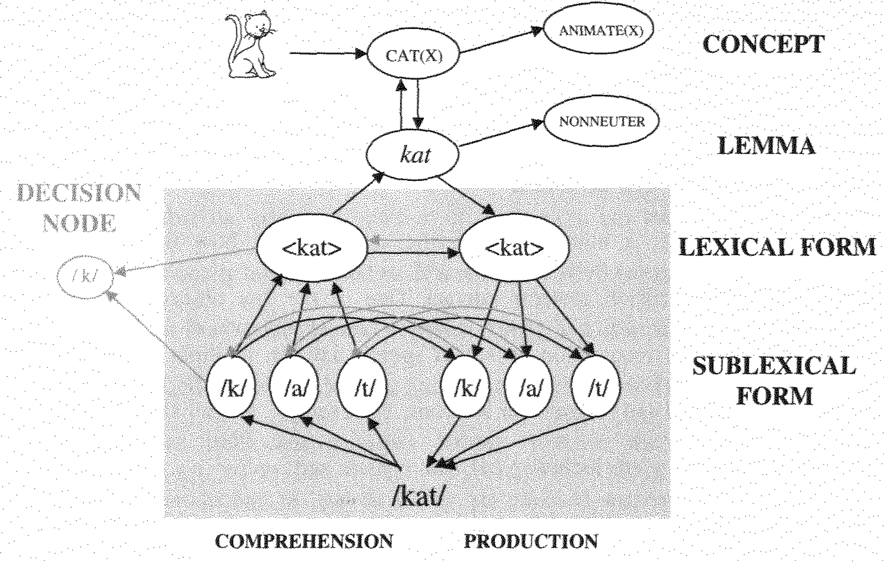


Figure 2.1 *A fragment of the* WEAVER++ *model. Word planning is implemented as a staged process moving from conceptual preparation, via lemma retrieval, to word-form encoding (right side). Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification (left side). There are connecting links from the comprehension to the production side.*

Meyer and Schriefers (1991) conducted a series of picture-word interference experiments in which participants had to name pictured objects while simultaneously trying to ignore spoken distractor words. The spoken distractor words shared the beginning or the end of the phonological form of the picture names, or they were unrelated. The distractor onset was 300 or 150 msec before picture onset, it was 150 msec after picture onset, or there was no onset difference. Both begin- and end-related distractors speeded up picture naming compared to unrelated distractors. The onset of the facilitation effect differed between begin- and end-relatedness. This difference may be due to seriality in word-form encoding for production , as Meyer & Schriefers (1991) explained it. It may, however, also be a lexical cohort effect (Roelofs, 1997). Starreveld (2000) also asked participants to name pictures while trying to ignore spoken distractors. The distractors were begin-related words or their initial fragments (i.e., the part phonologically overlapping with the picture name), or they were phonologically unrelated words or fragments. Starreveld observed that fragments yielded more phonological facilitation than whole words, which he also interpreted as a lexical cohort effect. The picture name and the begin-related words are cohort competitors, whereas the picture name and the fragment are not. The cohort competition may have diminished the facilitation effect. Schriefers and Teruel (1999) and Roelofs (2002) observed that both nonword and word distractors speed up conceptually driven word production, which suggests that the phonological facilitation is at least partly mediated by sublexical connections.

Although the existing results suggest that there are lexical effects (Meyer & Schriefers, 1991; Starreveld, 2000) and sublexical effects (Roelofs, 2002; Schriefers & Teruel, 1999), it is unclear what the locus of the lexical effects is. In a model like WEAVER++ that distinguishes between lemmas and lexical forms, the lexical effects may arise at either level (cf. Roelofs, 1997). We report two experiments that aimed to determine the locus of the phonological facilitation effect, thereby simultaneously providing evidence on the relationship between the speech production and comprehension systems.

To determine the locus of phonological facilitation, participants performed three tasks, which differed in the processing levels engaged: determining whether pictured entities are animate or not, determining the grammatical gender of the picture names (neuter or nonneuter in Dutch), and naming the pictures. The animacy decision task requires conceptual level processing and picture naming requires conceptual, lemma, and word-form level processing. No consensus exists in the literature as to whether grammatical gender is available from a noun's lemma, as held by Levelt et al. (1999; Roelofs, 1992), or from the lexical form of a noun, as held by Caramazza (1997). Thus, gender decision may require lemma level processing or form level processing.

The two decision tasks were included for the following reasons. Evidence for phonological priming effects on conceptually mediated processing has been reported by Zwitserlood (1989; but see Janse & Quené, 1999, 2004). In her experiments, participants listened to spoken prime words (e.g., KAPITEIN, "captain") or fragments of these words (e.g., KAPIT). Participants were asked to make lexical decisions to written probes that were presented at the offset of the spoken primes. In the fragment condition (e.g., KAPIT), lexical decision times were reduced for probes that were semantically related to the word whose fragment was presented (e.g., *kapitein*, "captain") as well as for cohort competitors of this word (e.g., *kapitaal*, "capital"). For example, KAPIT facilitated responses to *schip* ("ship", semantically related to *kapitein*, "captain") as well as to *geld* ("money", semantically related to *kapitaal*, "capital"). However, when the prime was the complete word KAPITEIN ("captain"), only responses to *schip* ("ship") were facilitated while responses to *geld* ("money") were not. Thus, as soon as a segment of the fragment indicated that there was only one appropriate candidate, facilitation was observed for this candidate only. The phonological effect found in picture naming experiments (e.g., Meyer & Schriefers, 1991) could thus at least partly be mediated by conceptual level processing. To assess the presence of a conceptual contribution to the phonological facilitation effect in picture naming, the animacy decision task was included.

Evidence for the influence of phonological overlap on lemma level processing has recently been provided by Starreveld and La Heij (2004). Their participants performed a gender decision task on picture names, whereby the distractors consisted of the visual presentation of a first letter prime (e.g., K) of the target picture name (*kat* – "cat"). Their results showed that gender decision latencies were speeded up significantly by presentation of the first letter of the picture name in comparison to a neutral baseline. This effect, however, was much smaller than the effect of first letter primes in a picture naming task. This suggests that part of the phonological facilitation found in picture naming is mediated by the lemma level. We included the gender decision task to have a measure of lemma level activation induced by phonological overlap.

The presence or absence of phonological effects in different tasks should reveal the contribution of the different production processing levels to the phonological facilitation effect usually found in picture naming. It is important to stress, however, that both studies mentioned above (Zwitserlood, 1989; Starreveld & La Heij, 2004) have crucial differences to our study. The semantic effect reported by Zwitserlood (1989) in a comprehension experiment was based on written word targets while our experiments examined the influence on picture targets. The grammatical effect on gender decision found by Starreveld and La Heij (2004) was obtained in the context of written distractor letters, while our experiments examined the influence of spoken distractors.

EXPERIMENT 1

In the first experiment, spoken distractor words were presented that shared the beginning or the end of the phonological form of the picture names, or they were unrelated in form. The distractor onset was 150 msec before the picture appeared, coincided with picture presentation onset, or followed it after 150 msec. This is referred to as the stimulus onset asynchrony (SOA). In the end condition, picture onset was relative to the onset of the shared part of the distractor (cf. Meyer & Schriefers, 1991). Consequently, the time interval between picture onset and the onset of the overlapping segments of the distractor was the same for the begin and end conditions.

If phonological relatedness speeds up picture naming but has no effect on animacy and gender decisions, this would suggest that the facilitation happens during form processing rather than in grammatical or conceptual processing. Moreover, a difference in facilitation between begin (cohort) and end (rhyme) overlap in picture naming, but not in the other two tasks, would suggest that there are both lexical and sublexical form connections between the spoken word recognition and production systems. Less facilitation for begin- than for end-related distractors would suggest that picture naming is hampered more by cohort than by rhyme competitors, suggesting a lexical form effect. Research in the comprehension literature has shown that lexical activation differs between cohort and rhyme competitors (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), in that cohort competitors have the tendency to be more strongly activated than rhyme competitors.

METHOD

Participants. 72 native speakers of Dutch (10 male, 62 female; mean age: 21) participated in the experiment (24 speakers per stimulus onset asynchrony). They were undergraduate students of Radboud University Nijmegen. They had normal or corrected-to-normal hearing and vision. They were paid for their participation.

Materials and Design. The materials were selected such that they met the requirements of all three tasks. There were 32 pictures, all with monosyllabic names, 16 animate and 16 inanimate ones, and 16 taking the Dutch determiner *de* (nonneuter) and 16 taking *het* (neuter).

Each picture was combined with four spoken distractor words, which were begin related, end related, begin unrelated, or end unrelated (e.g., *kat*: KAS, MAT, VOEG, LANS). The full set of materials can be found in Appendix 2-A. In the begin related condition, the picture name and the distractor shared the syllable onset and nucleus. In the end related condition, the nucleus and coda were shared. The remaining segments in the syllables had at least two different phonological features between target and distractor. For two targets it was not possible to find an end related distractor with a mismatch consisting of more than one feature while simultaneously meeting all other constraints (*beer* – PEER; *been* – PEEN). For one other target it was not possible to find an end related distractor totally mismatching at the onset (*speer* – SFEER). The two unrelated conditions were created by recombining the begin related and end related distractors with the pictures. For one target, the same item was erroneously used as begin related and begin unrelated distractor (*kalf* – KAM). For some items the recombination yielded distractors in the unrelated condition which share one segment (e.g., *kat* – LANS, *sok* –SFEER, *voet* – MAT). All distractors were semantically unrelated to the target. They were matched for animacy, gender, frequency (CELEX database, Baayen, Piepenbrock, & Gulikers, 1995), and number of syllables. Moreover, the syllable structure was matched whenever possible after meeting all other constraints.

The experiment was run using a 3 x 2 x 2 within-participants design with the factors *task* (picture naming, gender decision, animate decision), *relatedness* (related, unrelated) and *position* (begin, end). *SOA* (stimulus onset asynchrony) was varied between participants. SOA had three levels: –150, 0 and +150 msec. The order of the tasks was counterbalanced using a Latin square design.

Procedure. Participants were tested individually in a quiet room. They sat in front of a computer screen wearing closed headphones. Before the experiment, participants were familiarized with the pictures and their names. Before each task block, participants received written instructions about the task and were asked to respond as fast and accurately as possible. Dutch nouns with neuter gender take the definite article *het* and nouns with nonneuter gender take the article *de*. In the gender decision task, one button indicated the *de*-response (nonneuter) and the other the *het*-response (neuter). In the animate decision task, the buttons represented *animate* and *inanimate*. Each task block started with a few practice trials and consisted of 4 series of test trials. Each picture appeared once in each series of stimuli. There was a short break between the series. An experimental session lasted about 40 minutes.

Each trial had the following structure. A picture was presented and stayed on the screen for 1000 msec. Depending on the SOA, the distractor appeared 150 msec before the picture appeared, coincided with picture presentation, or followed it after 150 msec. As indicated, in the end condition, picture onset was relative to the beginning of the vowel of the distractor. A new target appeared either after the response (voice key trigger or button press) or after a timeout period of 1500 msec.

The experiment was controlled by the NESU (Nijmegen Experimental Setup) software developed at the Max Planck Institute. The picture naming session was recorded on DAT. The microphone was connected to an electronic voice key, which in turn was connected to a computer that recorded the response times. For the two decision tasks, a push-button box with two keys was used to register response latencies. Response latencies were written to hard disk after each trial. During the experiment the experimenter sat in the same room and noted down hesitations, wrong responses, or failures of the voice key.

Analysis. Trials on which participants stuttered, used names other than the intended ones, corrected themselves afterwards, smacked their lips, or made other noises triggering the voice key were excluded from the analysis of the picture naming latencies. In the two decision tasks wrong button presses were excluded from the analysis. The response latencies were analyzed using repeated measurement analyses of variance with the factors task, relatedness, position, and SOA. SOA was tested between participants and the other factors were tested within participants. Interactions were further explored by *t*-tests. In all analyses, an alpha level of .05 was adopted.

RESULTS AND DISCUSSION

Table 2.1 and Figure 2.2 show that the relatedness effects differed depending on the task, position, and SOA. While the two decision tasks showed no phonological facilitation effect, there were facilitation effects in the picture naming task.

There were no significant effects in the error analysis, except for a main effect of task, $F_1(2,138) = 77.14$, $MSE = 6$, $p < .001$, $F_2(2,186) = 77.34$, $MSE = 4$, $p < .001$. The statistical analysis of the response latencies yielded main effects of task, $F_1(2,138) = 410.78$, $MSE = 13436$, $p < .001$, $F_2(2,186) = 1095.32$, $MSE = 6811$, $p < .001$, position, $F_1(1,69) = 47.78$, $MSE = 773$, $p < .001$, $F_2(1,93) = 46.17$, $MSE = 1237$, $p < .001$, and relatedness, $F_1(1,69) = 54.53$, $MSE = 693$, $p < .001$, $F_2(1,93) = 36.75$, $MSE = 1808$, $p < .001$.

Table 2.1 *Mean Response Latencies (in Milliseconds), Their Standard Deviations, and the Error Percentages per SOA, Task, and Distractor Condition in Experiment 1.*

	SOA								
	−150			0			+150		
Distractor	Mean	SD	E%	Mean	SD	E%	Mean	SD	E%
Picture Naming									
begin related	699	142	6.3	686	149	9.1	661	156	8.6
begin unrelated	712	146	8.2	740	163	8.1	713	186	12.9
end related	669	145	6.5	677	138	9.0	677	153	11.2
end unrelated	700	148	7.4	707	158	9.1	722	180	12.9
Gender Decision									
begin related	810	187	8.2	811	180	7.7	843	218	8.7
begin unrelated	825	181	7.3	815	180	7.8	829	206	9.4
end related	798	180	6.9	796	187	6.1	824	194	10.3
end unrelated	797	170	7.2	800	174	8.1	841	213	9.2
Animate Decision									
begin related	536	124	2.2	554	123	1.4	547	136	1.8
begin unrelated	543	128	1.7	552	119	2.1	553	133	1.8
end related	527	128	1.8	540	122	1.8	535	128	2.3
end unrelated	522	127	2.3	534	120	1.8	536	141	3.3

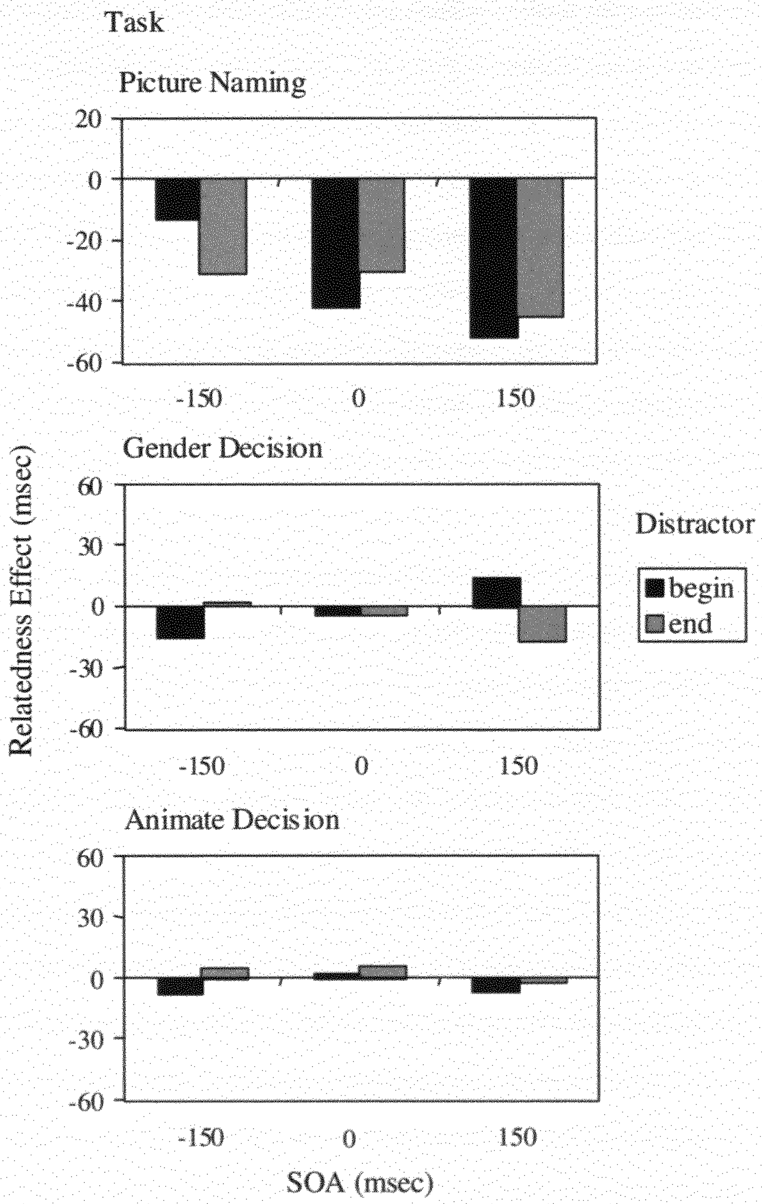


Figure 2.2. *The phonological relatedness effect in milliseconds (response times for the related minus unrelated condition) for the begin and end conditions per task and SOA in Experiment 1. Grey bars show effects for the end related condition, black bars show effects for the begin related condition.*

There were interactions between position and SOA, $F_1(2,69) = 9.6$, $p < .001$, $F_2(2,93) = 6.09$, $p = .003$, of task and relatedness, $F_1(2,138) = 40.32$, $MSE = 659$, $p < .001$, $F_2(2,186) = 43.86$, $MSE = 850$, $p < .001$, and triple interactions between task, relatedness, and SOA, $F_1(4,138) = 2.68$, $MSE = 1768$, $p = .03$, $F_2(4,186) = 3.53$, $MSE = 2998$, $p = .008$, and between task, position, and SOA, which was marginally significant in the item analysis, $F_1(4,138) = 4.89$, $MSE = 2740$, $p = .001$, $F_2(4,186) = 2.31$, $MSE = 2248$, $p = .06$. There was also a four-way interaction between task, relatedness, position, and SOA (significant in the participant analysis and marginally significant in the item analysis), $F_1(4,138) = 4.99$, $MSE = 2173$, $p = .001$, $F_2(4,186) = 2.31$, $MSE = 2198$, $p = .06$. No other interactions were significant.

Picture naming. Figure 2.2 shows that the facilitation effect for the end overlap condition was fairly constant across SOAs. However, in the begin overlap condition the effect increased from SOA = –150 to SOA = +150 msec.

The statistical analysis yielded main effects of position, $F_1(1,69) = 12.49$, $MSE = 780$, $p = .001$, $F_2(1,93) = 14.32$, $MSE = 1148$, $p < .001$, and of relatedness, $F_1(1,69) = 137.58$, $MSE = 652$, $p < .001$, $F_2(1,93) = 73.4$, $MSE = 1840$, $p < .001$, but not of SOA, $F_1(2,69) < 1$, $MSE = 20097$, $p = .81$, $F_2(2,93) < 1$, $MSE = 13317$, $p = .41$. There were interactions between position and SOA, $F_1(2,69) = 14.26$, $MSE = 11132$, $p < .001$, $F_2(2,93) = 7.47$, $MSE = 8580$, $p = .001$, and between relatedness and SOA, $F_1(2,69) = 6.45$, $MSE = 4209$, $p = .003$, $F_2(2,93) = 3.41$, $MSE = 6280$, $p = .037$. The triple interaction between position, relatedness, and SOA was significant by participants, $F_1(2,69) = 3.65$, $MSE = 1494$, $p = .031$, $F_2(2,93) = 1.6$, $MSE = 2756$, $p = .208$.

At SOA = –150 msec, there was a main effect of relatedness, $F_1(1,23) = 22.21$, $MSE = 520$, $p < .001$, $F_2(1,31) = 8.54$, $MSE = 2120$, $p = .006$, and of position, $F_1(1,23) = 20.69$, $MSE =$

495, $p < .001$, $F_2(1,31) = 24.68$, $MSE = 613$, $p < .001$. Position and relatedness interacted, $F_1(1,23) = 6.06$, $MSE = 307$, $p = .022$, $F_2(1,31) = 5.98$, $MSE = 786$, $p = .020$, showing that the relatedness effect was larger in the end than in the begin condition. Pairwise comparisons revealed a significant relatedness effects in the end condition, $t_1(23) = 4.31$, $p < .001$, $t_2(31) = 4.07$, $p < .001$. The relatedness effect in the begin condition was significant in the participants analysis, $t_1(23) = 3.09$, $p = .005$, $t_2(31) = 1.15$, $p = .261$.

At SOA = 0 msec, there was a main effect of relatedness, $F_1(1,23) = 40.76$, $MSE = 744$, $p < .001$, $F_2(1,31) = 16.44$, $MSE = 2597$, $p < .001$, and of position, $F_1(1,23) = 33.72$, $MSE = 526$, $p < .001$, $F_2(1,31) = 5.56$, $MSE = 2564$, $p = .025$. Relatedness and position did not interact, $F_1(1,23) = 1.55$, $MSE = 529$, $p = .23$, $F_2(1,31) < 1$, $MSE = 1779$, $p = .70$.

At SOA = +150 msec, there was a main effect of relatedness, $F_1(1,23) = 81.25$, $MSE = 693$, $p < .001$, $F_2(1,31) = 44.92$, $MSE = 1911$, $p < .001$, but not of position, $F_1(1,23) = 3.05$, $MSE = 1320$, $p = .09$, $F_2(1,31) < 1$, $MSE = 1445$, $p = .40$. Relatedness and position did not interact, $F_1(1,23) < 1$, $MSE = 392$, $p = .39$, $F_2(1,31) < 1$, $MSE = 2390$, $p = .64$.

Gender decision. The statistical analysis yielded a main effect of position, $F_1(1,69) = 16.82$, $MSE = 633$, $p < .001$, $F_2(1,93) = 8.99$, $MSE = 1665$, $p = .003$, but no effect of SOA, $F_1(2,69) < 1$, $MSE = 35796$, $p = .51$, $F_2(2,93) = 2.15$, $MSE = 9441$, $p = .12$. There was no effect of relatedness, $F_1(1,69) = 1.28$, $MSE = 1042$, $p = .26$, $F_2(1,93) = 4.76$, $MSE = 1245$, $p = .032$. There were no interactions between position and relatedness, $F_1(1,69) < 1$, $MSE = 607$, $p = .39$, $F_2(2,93) < 1$, $MSE = 1409$, $p = .024$, of SOA and position, $F_1(2,69) = 2.6$, $MSE = 1740$, $p = .08$, $F_2(2,93) = 2.06$, $MSE = 3435$, $p = .13$, of SOA and relatedness, $F_1(2,69) < 1$, $MSE = 163$, $p = .86$, $F_2(2,93) < 1$, $MSE = 136$, $p = .90$, nor a triple interaction between position, relatedness, and SOA, $F_1(2,69) = 5.54$, $MSE = 3360$, $p = .006$, $F_2(2,93) = 1.47$, $MSE = 2068$, $p = .24$.

Animate decision. There was a main effect of position, $F_1(1,69) = 36.81$, $MSE = 449$, $p < .001$, $F_2(1,93) = 71.62$, $MSE = 372$, $p < .001$, but not of relatedness, $F_1(1,69) < 1$, $MSE = 319$, $p = .96$, $F_2(1,93) < 1$, $MSE = 423$, $p = .93$, and not of SOA, $F_1(2,69) < 1$, $MSE = 15472$, $p = .75$, $F_2(2,93) = 5.02$, $MSE = 2316$, $p = .009$. There were no interactions between position and relatedness, $F_1(1,69) = 3.21$, $MSE = 272$, $p = .08$, $F_2(1,93) = 2.08$, $MSE = 508$, $p = .15$, between SOA and position, $F_1(2,69) < 1$, $MSE = 25$, $p = .95$, $F_2(2,93) < 1$, $MSE = 17$, $p = .95$, between SOA and relatedness, $F_1(2,69) < 1$, $MSE = 309$, $p = .38$, $F_2(2,93) = 2.65$, $MSE = 1122$, $p = .08$, and also no triple interaction between position, relatedness, and SOA, $F_1(2,69) < 1$, $MSE = 120$, $p = .65$, $F_2(2,93) < 1$, $MSE = 28$, $p = .95$.

In the picture naming task, we found phonological facilitation effects for both begin- and end-related distractors, which is in line with previous results (e.g., Meyer & Schriefers, 1991). The facilitation can be explained in terms of sublexical form connections running from the spoken word recognition system to the spoken word production system (e.g., Roelofs, 1997, 2004a). Segments perceived in the spoken distractor word preactivate the corresponding segments in the production system through these sublexical links. When the planning of the picture name reaches the sublexical output form level, the selection of segments occurs faster because of the preactivation. This speed up should occur regardless of the serial position of the segments in the target word, as observed.

Importantly, the phonological facilitation effect was modulated by position of overlap (begin vs. end) at the preexposure SOA but not at the later ones. This modulation suggests lexical involvement. It is likely that the lexical representation of a distractor presented at a preexposure SOA is already strongly activated when the production planning process reaches the lexical form level (because we used short monosyllabic words). A strongly activated lexical candidate will suppress all other lexical candidates within its cohort (Norris, 1994) and will dominate the activation that is sent to the production system via the connecting links (see Figure 2.1). When the production system now wants to select one of these lexical candidates from the same cohort (which is the case in the begin-related condition), selection for production will take longer than when the picture name is not a member of the distractor cohort and therefore has no such strong competitor (which holds for the end-related condition). At later SOAs, the distractor will not be so strongly activated when the production process reaches the lexical level, because there will have been less time for it to be fully recognised and to increase its activation to a high amount. Consequently, the picture names will not be in the context of a highly activated lexical competitor. This would explain why the phonological facilitation effect differs between the begin- and end-relatedness conditions at the preexposure SOA but not at the later ones.

Whereas begin-related distractors yielded less facilitation than end-related distractors in the present experiment, Meyer and Schriefers (1991) found the opposite pattern at the –150 msec preexposure SOA. They obtained stronger facilitation for begin than for end overlapping distractors. It is clear from earlier research (e.g., Allopenna, Magnuson & Tanenhaus, 1998) that cohort members are more activated than rhyme members of a word. This, however, can lead to facilitation of or to interference with production, depending on several factors like cohort size, number of overlapping segments, and so forth (cf. Roelofs, 1997). Consequently, a difference in effects for cohorts and rhymes is predicted, but not the direction of this difference. Looking at the results of Meyer and Schriefers (1991) and the results of the present experiment, it seems that the difference between cohorts and rhymes that actually occurs in an experiment can vary depending on the material set. Regardless of the direction of the effect, it depends on lexical status (cohort vs. rhyme competitor), which is the key point here.

Phonological relatedness speeded up picture naming, but it had no effect on the animacy and gender decisions. This suggests that the facilitation happened during form processing rather than grammatical or conceptual processing. This raises the question as to why Zwitserlood (1989) obtained facilitation effects in a lexical decision task and Starreveld and La Heij (2004) obtained effects in gender decision. Whereas we used animacy decisions to pictures, Zwitserlood (1989) used lexical decisions to written words, which may involve different processing steps. Important for present purposes is that facilitation is not obtained in responding to pictures. Whereas we used spoken words, Starreveld and La Heij (2004) used written letters as primes. It may be that spoken words are less effective as primes than single written letters, because spoken words also contain mismatching speech segments whereas single letters do not. In the next experiment, we used fragment distractors as well, further examining the lexical modulation found at the preexposure SOA.

EXPERIMENT 2

The smaller phonological facilitation effect for begin- than for end-relatedness at the preexposure SOA suggests that picture naming was hampered more by cohort than by rhyme competitors, a lexical effect. If this is the case, phonological facilitation should be smaller when the distractor is a cohort competitor (i.e., a word having both matching and mismatching segments) than when it is not (i.e., a fragment having the matching segments but lacking the mismatching ones), as observed by Starreveld (2000) for picture naming. We tested this prediction in Experiment 2 for all three tasks (picture naming, gender decision, animate decision), using SOA = –150 msec only.

This design, using whole words and word fragments, gives us the opportunity to replicate the effect recently found by Starreveld and La Heij (2004). They observed that presenting the first letter of the picture name speeds up picture naming as well as gender decision, although the effect in gender decision was much smaller than in picture naming. Our results in Experiment 1 showed no facilitation of phonological overlap in gender decision, which could of course be different in the case of matching segments only. This is tested in Experiment 2.

METHOD

Participants. 24 new participants from the MPI participant pool took part in the experiment (6 male, 18 female; mean age: 21). They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, Design, Procedure and Analysis. This was the same as in Experiment 1, except that only the materials of the begin-related and -unrelated word conditions were used (e.g., *kat*: KAS, VOEG). To create the word fragments (*kat*: KA, VOE), the mismatching segments were removed from the speech signal under both visual and auditory control using a speech editor. The manipulation of word versus fragment is referred to as *type*. Only the critical SOA = -150 msec was tested.

RESULTS AND DISCUSSION

Table 2.2 gives the mean latencies, standard deviations, and error percentages per task and distractor condition for Experiment 2. Figure 2.3 shows the relatedness effects for the different tasks and conditions. The table and figure show that there is phonological facilitation in picture naming and gender decision, while there is no effect in the animate decision task.

There were no significant effects in the error analysis. The statistical analysis of the response times yielded main effects of task, $F_1(2,46) = 116.50$, $MSE = 14530$, $p < .001$, $F_2(2,62) = 559.00$, $MSE = 3821$, $p < .001$, type, $F_1(1,23) = 32.71$, $MSE = 816$, $p < .001$, $F_2(1,31) = 95.79$, $MSE = 498$, $p < .001$, and relatedness, $F_1(1,23) = 63.80$, $MSE = 590$, $p < .001$, $F_2(1,31) = 36.26$, $MSE = 1117$, $p < .001$. There were interactions between task and relatedness, $F_1(2,46) = 7.20$, $MSE = 562$, $p = .002$, $F_2(2,62) = 11.85$, $MSE = 826$, $p < .001$, task and type, $F_1(2,46) = 10.65$, $MSE = 618$, $p < .001$, $F_2(2,62) = 9.59$, $MSE = 430$, $p < .001$. There was no interaction between relatedness and type, $F_1(1,23) < 1$, $MSE = 883$, $p = .371$, $F_2(1,31) = 1.11$, $MSE = 973$, $p = .30$. The triple interaction between task, relatedness, and type was significant, $F_1(2,46) = 4.15$, $MSE = 539$, $p = .022$, $F_2(2,62) = 6.46$, $MSE = 464$, $p = .003$.

Picture naming. Figure 2.3 shows that the phonological facilitation effect was larger for the fragment condition than for the word condition. The statistical analysis yielded main effects of type, $F_1(1,23) = 24.24$, $MSE = 662$, $p < .001$, $F_2(1,31) = 45.23$, $MSE = 451$, $p < .001$, and of relatedness, $F_1(1,23) = 65.26$, $MSE = 516$, $p < .001$, $F_2(1,31) = 35.30$, $MSE = 1491$, $p < .001$. There was also an interaction between type and relatedness, $F_1(1,23) = 16.25$, $MSE = 303$, $p = .001$, $F_2(1,31) = 13.05$, $MSE = 569$, $p = .001$, showing that the facilitation was larger for the fragments than for the words.

Table 2.2 *Mean Response Latencies (in Milliseconds), Their Standard Deviations, and the Error Percentages per Task and Distractor Condition in Experiment 2.*

	Task								
	Picture Naming			Gender Decision			Animate Decision		
Distractor	Mean	SD	E%	Mean	SD	E%	Mean	SD	E %
related word	626	148	6.0	751	220	9.6	478	117	3.4
unrelated word	650	159	6.8	768	219	10.9	486	113	2.9
related fragent	586	132	5.5	719	202	9.5	472	122	4.3
unrelated fragment	638	147	6.9	732	194	9.6	475	118	3.5

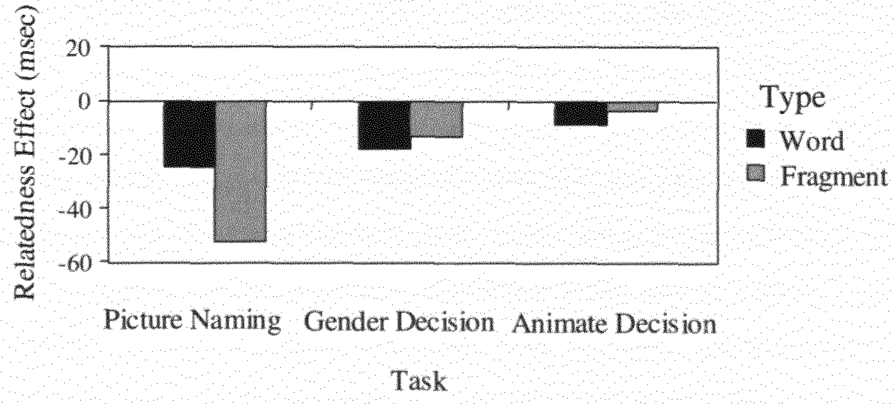


Figure 2.3 *The phonological relatedness effect in milliseconds (response times for the related minus unrelated condition) for the word and fragment conditions per task and SOA in Experiment 2.*

Pairwise comparisons revealed significant relatedness effects in both the word condition, $t_1(23) = 4.47$, $p < .001$, $t_2(31) = 3.46$, $p = .002$, and the fragment condition, $t_1(23) = 8.04$, $p < .001$, $t_2(31) = 6.44$, $p < .001$.

Gender decision. The statistical analysis yielded main effects of type, $F_1(1,23) = 40.44$, $MSE = 690$, $p < .001$, $F_2(1,31) = 50.01$, $MSE = 637$, $p < .001$, and of relatedness, $F_1(1,23) = 4.32$, $MSE = 1283$, $p = .049$, $F_2(1,31) = 7.55$, $MSE = 854$, $p = .010$. There was no interaction between type and relatedness, $F_1(1,23) < 1$, $MSE = 1312$, $p = .746$, $F_2(1,31) < 1$, $MSE = 929$, $p = .60$.

Animate decision. The statistical analysis yielded a main effect of type, $F_1(1,23) = 4.95$, $MSE = 362$, $p = .036$, $F_2(1,31) = 11.31$, $MSE = 269$, $p = .002$. There was no main effect of relatedness, $F_1(1,23) = 2.47$, MSE = 252, $p = .130$, $F_2(1,31) = 3.65$, $MSE = 382$, $p = .07$, and no interaction between type and relatedness, $F_1(1,23) < 1$, $MSE = 344$, $p = .544$, $F_2(1,31) < 1$, $MSE = 379$, $p = .63$.

We observed that phonological facilitation in picture naming is larger when the distractor is not a cohort competitor (a fragment having no mismatching segments) than when it is a cohort competitor (a word having mismatching segments at the end), replicating earlier findings in the literature (Starreveld, 2000). This suggests that there is a lexical contribution to the phonological facilitation effect. Although in this experiment there was also a facilitation effect in the gender decision task, the modulation of the effect by the type of distractor (fragment versus word) was again exclusively found in the picture naming task, suggesting a form level locus of the lexical contribution.

Interestingly, phonological facilitation significantly speeded up the gender decisions in this second experiment. Numerically, however, the effect sizes for gender decision in the two experiments were very close: 15 msec (Experiment 1, begin-overlap condition) and 17 msec (Experiment 2, word condition). It may have turned out significant in Experiment 2 because the second condition (fragment) also showed the effect, whereas the end-overlap condition in Experiment 1 did not. Statistical comparisons showed that, in the present experiment, the effect in picture naming was significantly larger than the effect in gender decision, $F_1(2,46) = 10.65$, $MSE = 2471$, $p < .001$, $F_2(2,60) = 11.85$, $MSE = 3305$, $p < .001$, $t_1(23) = 2.91$, $p = .008$, $t_2(31) = 2.88$, $p = .007$.

The fragment conditions in the present experiment replicate what Starreveld and La Heij (2004) observed for written distractors in the picture naming and the gender decision tasks. Our results with the spoken distractors also showed facilitation effects in both cases, with the effect in picture naming being larger than the effect in gender decision. But if we take the whole word conditions into account, whether one uses written or spoken distractors seems to make a difference. In the experiments of Starreveld and La Heij (2004), letters and whole words yielded the same amount of phonological facilitation in picture naming, which is different from what Starreveld (2000) observed for picture naming in the context of spoken distractors. In our experiment, fragments yielded larger facilitation than whole words did in the picture naming task as well, pointing to a difference in effect between written and spoken distractors. The second important point to mention is that the modulation of the facilitation by the type of the distractor (fragment versus word) was not found in the gender decision task. This result suggests that the locus of the lexical modulation found in the picture naming task is not at the lemma level, but rather at a lexical form processing level.

GENERAL DISCUSSION

Two experiments were reported that aimed to determine the locus of the phonological facilitation effect typically found in picture-word interference studies. Dutch participants named pictures, determined whether the pictured entities were animate or not, and determined the grammatical gender of the picture names. Trials were blocked by task.

In Experiment 1, spoken distractor words were presented that shared the beginning or the end of the phonological form of the picture names. The distractor onset was 150 msec before or after picture onset, or there was no onset difference. Phonological relatedness speeded up picture naming, but it had no effect on the animacy decisions and gender decisions. Facilitation was obtained for both begin- and end-overlap in picture naming, suggesting sublexical involvement. Moreover, begin-overlap yielded a smaller effect than end-overlap at the distractor preexposure SOA, suggesting lexical involvement.

Experiment 2 further examined the locus of this modulation of the phonological facilitation effect. The begin-related and -unrelated distractor words of Experiment 1 were used. In addition, initial fragments of these words were presented as distractors. Gender decisions were speeded up by both words and fragments in this experiment. Moreover, a much larger phonological effect was obtained in picture naming. Most importantly, the modulation of the phonological effect by lexical status (word vs. fragment) was obtained exclusively in picture naming. This supports our assumption that the modulation occurs at the level of lexical forms. To conclude, although there were small effects in the gender decision task suggesting facilitatory influences at the lemma level as well, the results of the two experiments suggest that the main locus of phonological facilitation is the level of lexical and sublexical forms in spoken word planning.

Influences of Speech Planning on Speech Recognition

CHAPTER 3

REBECCA ÖZDEMIR, ARDI ROELOFS, & WILLEM J. M. LEVELT[2]

ABSTRACT

The perceptual-loop theory of self-monitoring of speech production (e.g., Levelt, 1989) holds that speakers monitor their internal and external speech via the speech comprehension system. This view predicts phonological influences from spoken word planning on speech recognition, which was tested in five experiments. Participants were shown pictures while hearing a tone or spoken word presented 300 or 600 msec after picture onset. When a spoken word was presented, participants had to indicate whether it contained a pre-specified phoneme; when the tone was presented, they had to indicate whether the picture name contained the phoneme (Experiment 1) or they had to name the picture (Experiment 2). The phoneme monitoring latencies for the spoken words were shorter when the picture name contained the pre-specified phoneme compared to when it did not. This priming effect was obtained only at the SOA of 600 msec. Priming was also obtained when the phoneme was part of spoken nonwords (Experiment 3). However, no priming was obtained when the overlap concerned phonological features rather than full phonemes (Experiment 4) and when the picture required no response (Experiment 5). These results suggest that internal phonological links exist from spoken word planning to speech recognition, as entailed by the perceptual-loop theory of self-monitoring.

---

[2] A version of this chapter is submitted for publication

INTRODUCTION

In psycholinguistics, speech production and speech comprehension are often treated as if they were separate processes and independent from each other. However, in everyday communication, speech production and perception often seem to happen simultaneously. Speakers regularly seem to plan their turn in a conversation while simultaneously listening to an interlocutor (e.g., Levelt, 1989). Moreover, speakers seem to monitor their own speech for errors while simultaneously planning an upcoming utterance. This raises the question of how ongoing speech recognition processes influence the planning of speech and vice versa.

Evidence from picture-word interference studies suggests that hearing spoken words influences naming pictures (e.g., Schriefers, Meyer, & Levelt, 1990; Meyer & Schriefers, 1991; Starreveld, 2000). When participants have to name pictured objects, for example a pictured cat, the time it takes to name the object is shortened when they hear a phonologically related spoken word (e.g., CAP) compared to an unrelated word (e.g., TREE). Moreover, both spoken words and their initial fragments speed up picture naming (Özdemir, Roelofs, & Levelt, submitted, Chapter 2; Starreveld, 2000). The priming effect from words and fragments suggests the existence of phonological links running from speech perception to speech planning.

It is far less clear, however, whether similar phonological links from speech planning to speech perception exist. Clearly, speakers listen to their own speech, but this does not require internal phonological links between speech planning and recognition. Production and perception are linked via the overt speech signal in that situation. Still, a prominent theory of self-monitoring holds that speakers monitor both their external and internal speech, and that both are accomplished via the speech comprehension system (Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999). An internal route for self-monitoring of speech implies the existence of phonological links from speech planning to speech recognition.

Research conducted by Wheeldon and Levelt (1995) provided evidence that phonological representations underlie the self-monitoring of internal speech. Their participants were native speakers of Dutch who spoke English fluently. They had to monitor for target phonemes in the Dutch translation equivalent of visually presented English words. For example, they had to indicate by means of a button press (yes/no) whether the phoneme /n/ is part of the Dutch translation equivalent of the English word WAITER. The Dutch word is *kelner*, which has /n/ as the onset of the second syllable, so requiring a positive response. All Dutch target words were disyllabic. There is evidence that phonological word representations are planned from the beginning of a word to its end (e.g., Meyer & Schriefers, 1991). To examine the time course of phonological encoding, Wheeldon and Levelt (1995) manipulated the serial position of the critical phonemes in the Dutch words. The target phoneme could be the onset or coda of the first syllable, or the onset or coda of the second syllable. Monitoring latencies increased with the serial position of the segments within the word. In order to experimentally verify whether phonological rather than phonetic representations were monitored, participants had to perform the phoneme monitoring task while simultaneously counting aloud, which is known to suppress the maintenance of phonetic representations. The monitoring latencies were longer with the counting task, but the seriality effect was replicated. This suggests that self-monitoring involves a phonological rather than a phonetic representation. Recently, Wheeldon and Morgan (2002) replicated the seriality effect in internal phoneme monitoring in English.

The research by Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) concerned monitoring of internal speech, suggesting that internal self-monitoring involves phonological representations. However, their research leaves open the question of whether self-monitoring is achieved via the speech comprehension system, as the perceptual-loop theory assumes (Levelt, 1989; Levelt et al., 1999), or via production-internal mechanisms (e.g., Laver, 1973, 1980). Self-monitoring of internal speech via the speech comprehension system requires internal phonological links between the production and the comprehension systems. For example, in computationally modeling self-monitoring, Roelofs (2004a) assumed that phonological representations are generated incrementally during planning the production of words and that these phonological representations are sent to the speech comprehension system as they become available over time. The existence of internal phonological links between speech planning and speech comprehending entails phonological influences of speech planning on the comprehension of *external* speech.

In the present article, we report a series of experiments that examined the presence and nature of these influences. In the experiments, participants were shown pictures and they heard a tone or spoken word presented shortly after picture onset. When a spoken word was presented, participants had to indicate whether it contained a pre-specified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme (in the first experiment) or they had to name the picture (in the other experiments). We measured the phoneme monitoring latencies for the spoken words when the picture name contained the pre-specified phoneme (henceforth the *match* condition) and when it did not (the *nonmatch* condition). Faster responses in the match than in the nonmatch condition would suggest facilitation of the recognition of the target phoneme in the spoken word. Such a phonological priming effect would be evidence for the existence of phonological links running from speech planning to speech recognition, affecting the recognition of external speech.

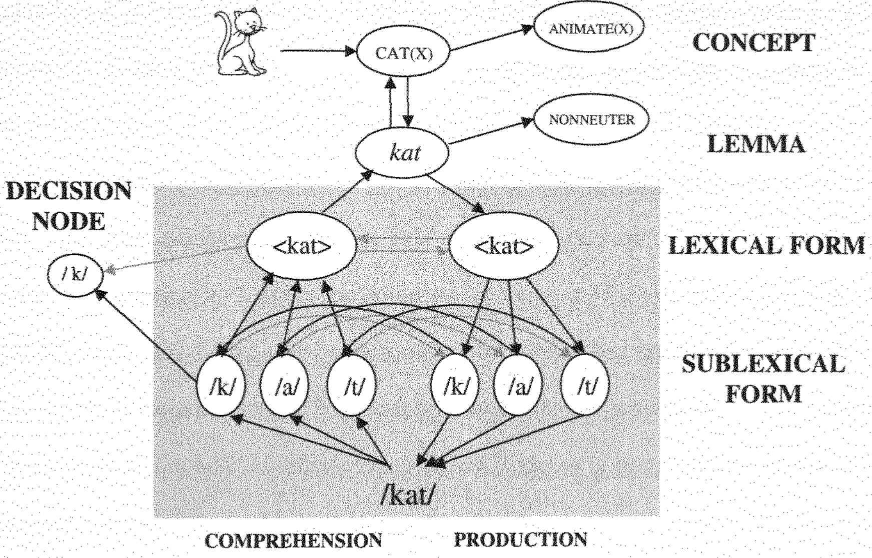


Figure 3.1 *Working model for picture-primed phoneme monitoring, combining the WEAVER++ model of spoken word planning (Levelt et al., 1999; Roelofs, 1997, 2004a) and the MERGE model of phoneme monitoring (Norris et al., 2000; Norris, 1994). The links from production to comprehension are dynamic in that they are established as a result of encoding a phonological word (Roelofs, 2004a). For convenience, the dynamic links are shown as network links.*

As a working model for our research, we used a combination of the WEAVER++ model of spoken word planning (Levelt et al., 1999; Roelofs, 1997, 2004a) and the MERGE model of phoneme monitoring during speech perception (Norris et al., 2000; Norris, 1994). Both models assume that processing occurs by feedforward spreading of activation in a form network. In WEAVER++, activation spreads from morphemes to phonemes in phonological encoding for production. In MERGE, activation spreads from phonemes to morphemes and phoneme decision nodes in speech recognition. Figure 3.1 illustrates the combination of the two models.

There are three network strata in WEAVER++, which are shown in Figure 3.1. A conceptual stratum represents the concepts of words as nodes (e.g., CAT(X)) and links in a semantic network. A syntactic stratum contains lemma nodes for words, such as *cat*, which are connected to nodes for their syntactic class (noun). A word-form stratum represents the morphemes (<cat>) and phonemes (/k/, /æ/, and /t/) of words. Information needed for word production is retrieved from the network by spreading activation. For example, a perceived object (e.g., a cat) activates the corresponding concept node (i.e., CAT(X)). Next, CAT(X) sends activation to its lemma, morpheme, and phonemes. The MERGE model describes how the monitoring for phonemes in external speech is accomplished. The model assumes that phoneme decision nodes are connected to the phonemic layer of a comprehension network. The connections from morphemes to the decision nodes are not always established, but are built on the fly, in response to task demands. Combining WEAVER++ and MERGE makes explicit how word planning in response to a pictured object (e.g., a cat) influences speech recognition processes, as indexed by phoneme monitoring latencies.

Assume a participant is shown a pictured cat followed by the spoken word CAP. The instruction is to monitor for the /k/ in the spoken word. Activation from CAT(X) will spread to the /k/ in the speech production network and from there to the /k/ in the speech recognition network and to the phoneme decision node. The pictured cat will prime the phoneme decision when the interval between picture onset and spoken word onset is not too short, such that the picture pre-activates the target phoneme. Evidence in the literature suggests that phonological activation in picture naming starts around 400 msec after picture onset (e.g., Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991). Thus, an SOA of 300 msec should be too short for obtaining a phonological priming effect, whereas an interval of 600 msec should yield a priming effect. The predicted effect of SOA is tested in Experiment 1.

Our working model predicts that it should not matter how the phonological activation in the production network comes about. That is, as long as a participant needs to plan the picture name, phonological activation should occur. Thus, phonological priming should occur regardless of whether the response to the picture is phoneme monitoring (tested in Experiment 1) or naming (tested in Experiment 2), but not when it is passive viewing (Experiment 5). Moreover, phonological priming should occur regardless of whether the target phoneme is part of a spoken word (tested in Experiments 1 and 2) or a spoken nonword (tested in Experiment 3). However, given that phonemes rather than their features are linked between speech production and recognition in the model, no phonological priming should occur when the target phoneme and the pre-activated phoneme only share most of their phonological features, such as /b/ and /p/ (tested in Experiment 4).

EXPERIMENT 1

The first experiment examined whether phonemes are pre-activated in the speech perception system as a result of speech planning processes. Participants were presented with pictures and a tone or a spoken word presented 300 or 600 msec after picture onset. These SOAs correspond to points in time before and after phonological information is activated in the production system, respectively. When a spoken word was presented, participants had to indicate whether it contained a pre-specified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme.

METHOD

Participants. 42 native speakers of Dutch (5 male, 37 female; mean age: 20 years) from the MPI subject pool took part in the experiment. We tested 28 participants with SOA = 300 msec, and 14 participants with SOA = 600 msec (we predicted a null effect at SOA = 300 msec, hence the larger number of participants). Participants all had normal or corrected-to-normal hearing and vision and they were paid for their participation.

Materials and Design. Participants had to monitor for the phonemes /p/ and /k/ in existing Dutch words. There were 10 items per phoneme, five monosyllabic and five disyllabic words. The target phoneme was always in initial position. Each word was preceded by a picture prime presented either 300 or 600 milliseconds before the onset of the word. There were two critical contexts, called *match* and *nonmatch*. In the match context, participants saw a picture which also had the target phoneme in initial position (e.g., picture: *peer*, word: PAAL; picture: *kist*, word: KAM). In the nonmatch context, the picture name did not start with the target phoneme (e.g., *peer*: KAM, *kist*: PAAL) nor did it contain this phoneme in any other position of its name. The pictures were recombined in such a way that the /k/ match pictures were the pictures for the nonmatch context of the target phoneme /p/ and vice versa. The full set of the test materials can be found in Appendix 3-A.

The picture names of the two contexts were matched for frequency (*p*-pictures: mean = 728 per 42 million; *k*-pictures: mean = 700 per 42 million, CELEX lexical database, Baayen, Piepenbrock, & Gullikers, 1995) and for number of segments (*p*-pictures: mean = 4.6 segments; *k*-pictures: mean = 4.3 segments). The target words were matched for these factors as well (*p*-words: mean frequency = 506, mean length = 4.6; *k*-words: mean frequency = 494, mean length = 4.6). The experiment included several filler trials to make sure that participants could not anticipate picture-word combinations. Several of these filler trials did not contain the target phonemes. In total, there were 25 % go and 75 % no-go trials. Overall, 16 % of the trials were critical trials (8 % match and 8 % nonmatch). Moreover, to make sure that the participants paid attention to the picture name and phonologically encoded it, we included an internal monitoring condition. In one third of the trials, participants heard a tone instead of a word after the presentation of the picture. The SOA for the tone was the same as for the words. In this case, participants had to monitor for the specified phoneme in the picture name.

The experiment was run with the crossed factors *phoneme* (/k/, /p/) and *context* (match, nonmatch). *SOA* (300 msec, 600 msec) was tested between participants. Monitoring trials were blocked by phoneme. The order of blocking was counterbalanced across participants. The order of items within a block was randomized.

Procedure. Participants were tested individually in a quiet room. They sat in front of a computer screen wearing headphones. Before the experiment started they received written instructions to react as fast and as accurately as possible. The participants were instructed to press a response button with their dominant hand when the word began with the target phoneme or, on hearing the tone, when the picture name began with this phoneme. The participants were familiarized with the pictures and their names before the beginning of the experiment.

The structure of an experimental trial was as follows. Participants saw a picture on the screen for 1000 msec. With a certain SOA (300 or 600 msec), they heard either a word or the tone via headphones. The next trial started after the reaction of the participant or after a timeout of 2000 msec. The experiment was controlled by the NESU (Nijmegen Experimental

Setup) software developed at the Max Planck Institute. A push button box with one button was used to register reaction times (RTs), which were written to hard disk after each trial.

The experiment consisted of two blocks of trials (with /p/ and /k/ as target), with every word occurring twice per block (preceded by a picture from the match or the mismatch context) and every picture occurring three times per block (combined with a word containing the target phoneme, a word not containing the target phoneme, or a tone requesting a response to the picture name). There was a short break between the blocks. An experimental session lasted about 30 minutes.

Analysis. Trials on which participants missed the targets or in which RTs exceeded the timeout of 2000 msec were regarded as errors and were excluded from the analysis. Moreover, responses with latencies longer than two standard deviations of the mean per participant and context were excluded from the analysis (SOA = 300 msec: 5.2%, SOA = 600 msec: 5.1%). After this exclusion, repeated measures of variance (ANOVAs) were performed on the data with the crossed factors phoneme and context and the between-participants factor SOA. Interactions were further explored by additional analyses. In all analyses, an alpha level of .05 was adopted.

RESULTS AND DISCUSSION

Figure 3.2 displays the priming effects for the target phonemes computed by subtracting the match context RTs from those of the nonmatch context. The black bars represent the priming effects for the phoneme /k/ and the grey bars represent the effects for the phoneme /p/. The corresponding mean RTs, standard deviations, and error percentages are given in Table 3.1.

There is a facilitation effect for /k/ (82 msec) as well as for /p/ (71 msec) at the SOA = 600 msec, while there is no effect for either of the two phonemes at SOA = 300 msec.

There were no significant effects in the error analysis. The statistical analysis of the reaction times yielded a significant main effect of context, $F_1(1,40) = 10.68$, $MSE = 9550$, $p = .002$, $F_2(1,8) = 16.23$, $MSE = 987$, $p = .004$, but no main effect of phoneme, $F_1(1,40) < 1$, $MSE = 9349$, $p = .35$, $F_2(1,8) < 1$, $MSE = 3087$, $p = .67$. There was an interaction between context and SOA, $F_1(1,40) = 12.79$, $MSE = 122146$, $p = .001$, $F_2(1,8) = 31.97$, $MSE = 31544$, $p < .001$, indicating that the phonological priming effect depended on SOA. There were no other interactions.

There were no significant main effects or interactions at SOA = 300 msec. Most importantly, there was no main effect of context, $F_1(1,27) < 1$, $MSE = 7763$, $p = .77$, $F_2(1,4) = 1.88$, $MSE = 693$, $p = .24$. At SOA = 600 msec, however, there was a main effect of phoneme, $F_1(1,13) = 12.65$, $MSE = 13261$, $p = .004$, $F_2(1,4) = 36.13$, $MSE = 1280$, $p = .004$. No other main effects or the interactions reached significance.

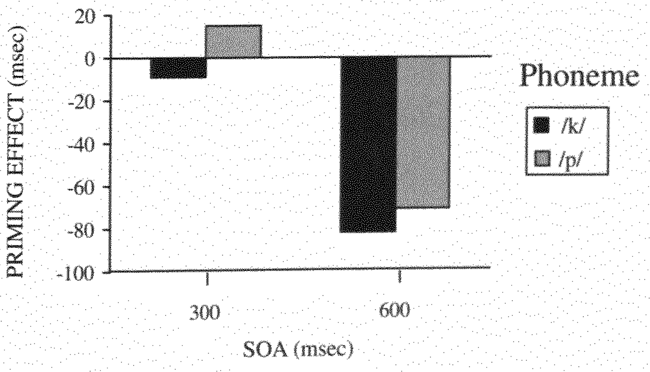


Figure 3.2 *Priming effects (response times in the match condition minus those of the nonmatch condition) observed in Experiment 1. The effects are shown for the SOAs of 300 and 600 msec and the target phonemes /k/ (black bars) and /p/ (grey bars). Negative values indicate facilitation and positive values indicate interference.*

Table 3.1 *Mean Response Times (in Milliseconds), Standard Deviations (in msec), and Error Rates (in %) split up by Phoneme and Context for Experiments 1-5.*

	Target Phoneme								
	/k/			/p/			Total		
Context	Mean	SD	E %	Mean	SD	E %	Mean	SD	E %
	Experiment 1								
	SOA = 300 msec								
match	683	208	0.2	691	221	0.4	687	215	0.3
nonmatch	692	177	0.7	676	224	0.6	684	201	0.7
	SOA = 600 msec								
match	673	205	1.3	702	235	0.5	688	220	0.9
nonmatch	755	249	1.8	773	277	1.3	764	263	1.6
	Experiment 2								
match	707	181	0.4	733	201	0.4	720	191	0.4
nonmatch	769	233	1.3	789	261	0.6	779	247	1.0
	Experiment 3								
match	727	215	0.4	750	225	0.6	739	220	0.5
nonmatch	772	239	0.8	778	211	1.7	775	225	1.3
	Experiment 4								
match	764	274	1.7	793	235	1.3	779	255	1.5
nonmatch	771	232	0.8	759	236	1.7	765	234	1.3
	Experiment 5								
match	597	188	0.9	624	227	0.4	610	207	1.3
nonmatch	577	181	0	631	163	0.8	604	172	0.8

To summarize, the statistical analysis showed that there were no significant effects found for either of the two phonemes at the early SOA (300 msec), while there were facilitation effects for both phonemes at the late SOA (600 msec). The encoding of a picture name containing the target phoneme thus helped the participants recognize this phoneme in the speech signal. This suggests that there are internal phonological links running from speech planning to speech recognition. It appears to be possible to pre-activate segments in the comprehension system via these links, which improves the performance in a task reflecting speech processing at the phonological level. This pre-activation depends on the availability of the relevant information in the production system, which leads to an absence of phonological priming at an SOA preceding the availability (300 msec) and to a facilitation effect at an SOA following the availability (600 msec).

It is important to exclude the possibility that the priming effect found at the late SOA is due to response preparation rather than due to a pre-activation of the segment in the comprehension system. By including the internal monitoring condition we gave the participants the opportunity to prepare for a response. If participants saw a picture whose name contained the target phoneme, they could in principle already prepare the button-press response. If the tone was heard, participants only had to execute the response. When instead of the tone, a word was presented, the response may already have been prepared. This could have led to a speeding up of the reaction time in the match context.

Moreover, it is important to exclude the possibility that the link between the systems causing the pre-activation had been established because there was internal monitoring involved. One could argue that this link does not exist in normal speech processing without involvement of an explicit internal monitoring task. We tried to rule out the response preparation and the task-dependent links objection by running Experiment 2.

EXPERIMENT 2

The intention of the second experiment was to exclude an interpretation of the facilitation effect found in Experiment 1 in terms of response preparation or a link that is set up only for the purpose of performing the internal phoneme monitoring task. We replaced the internal monitoring task with a simple picture naming task. Each time the participants heard a tone they had to simply name the picture. This situation rules out the contribution of response preparation. Moreover, there is also no need for phonological links serving only an explicit internal monitoring task. If the priming effect remains in the present experiment, this strongly suggests that the effect is not due to response preparation and task-dependent links.

METHOD

Participants. 12 new participants (2 male, 10 female; mean age: 20 years) from the MPI subject pool took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, Design, Procedure, and Analysis. This was the same as in Experiment 1, except that the internal monitoring task was replaced by a picture naming task. We tested SOA = 600 msec only. The analysis was performed exactly as the analysis in Experiment 1 but without the between-participants factor of SOA. The percentage of outlier trials was 5.8.

RESULTS AND DISCUSSION

The mean RTs, standard deviations, and error percentages for Experiment 2 are given in Table 3.1. The table shows that there were priming effects for the phoneme /k/ (62 msec) and the phoneme /p/ (56 msec).

There were no significant main effects or interactions in the error analysis. The analysis of the reaction times yielded a main effect of context, $F_1(1,11) = 8.76$, $MSE = 10035$, $p = .013$, $F_2(1,4) = 14.72$, $MSE = 1749$, $p = .019$. The main effect of phoneme, $F_1(1,11) < 1$, $MSE = 23388$, $p = .51$, $F_2(1,4) = 6.50$, $MSE = 832$, $p = .06$, as well as the interaction between context and phoneme, $F_1(1,11) < 1$, $MSE = 5708$, $p = .77$, $F_2(1,4) < 1$, $MSE = 4819$, $p = .79$, did not reach significance. There was no difference in the magnitude of the priming effects between Experiment 1 (SOA = 600 msec) and Experiment 2, as indicated by the absence of an interaction between context and experiment, $F_1(1,24) < 1$, $MSE = 5933$, $p = .57$, $F_2(1,18) < 1$, $MSE = 2245$, $p = .38$.

Experiment 2 replicated the priming effect found in Experiment 1 with picture naming as a control task. This supports the assumption that phonological links exist between the speech production and comprehension levels. The results of the experiment rule out an explanation in terms of links that are established merely for the purpose of an explicit, internal monitoring task. Moreover, the results rule out a response preparation explanation.

What is still unclear, however, is the exact locus of the priming effect. In our working model (illustrated in Figure 3.1), there are two possible loci for the influence on phoneme monitoring of the information coming from the production system. We assumed that the interaction between the production and comprehension systems was mediated by phonological representations. But the influence from production planning on speech recognition could also come from links between the morphemic levels in production and comprehension (the levels of lexical forms in Figure 3.1). The speech comprehension literature suggests that lexical influences on phoneme monitoring are hard to obtain when the target phoneme precedes the uniqueness point of a word (Connine & Titone 1996; Frauenfelder, Segui, & Dijkstra 1990), which is always the case when the target phoneme is in word-initial position as in our experiment. However, lexical influences with word initial phoneme have been shown before (Rubin, Turvey, & van Gelder, 1976; Cutler, Mehler, Norris, & Segui, 1987). Moreover, in the frequently used task of phoneme monitoring in nonwords, participants are forced to base their phoneme decision on sublexical information, since there is no lexical entry and thus no activation from the lexical level. In our first two experiments, however, the spoken items were existing words. The picture names may have spread activation to the comprehension system via connections at the lexical level. It could thus be the case that connections between the lexical form (morpheme) levels in production and comprehension are at least in part responsible for the priming effect.

EXPERIMENT 3

To assess the lexical and sublexical contributions to the priming effect, we ran a phoneme monitoring experiment with target phonemes embedded in spoken nonwords. The target phonemes were in initial position. When performing a phoneme monitoring task on nonwords, participants cannot benefit from lexical processing in the comprehension system. Moreover, lexical influences are hard to obtain with target phonemes in initial position. A priming effect observed in the case of monitoring for initial phonemes of nonwords should thus reflect the sublexical influence from the production to the comprehension system. The

MERGE model implements the claim that "if the use of lexical knowledge is discouraged, only connections from the prelexical level [to the phoneme decision nodes] will be constructed" (Norris et al. 2000, p. 316). Thus, lexical influences are also not expected in this particular experimental design according to this model.

In the present experiment, all spoken words used in Experiments 1 and 2 were replaced by spoken nonwords, forcing the participants to base their monitoring on sublexical information. Finding the facilitation effect again with the target phonemes in nonwords would strongly suggest that the influence from production to perception occurs at sublexical levels of processing.

METHOD

Participants. 12 new participants (2 male, 10 female; mean age: 21 years) from the MPI subject pool took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, Design, Procedure, and Analysis. This was the same as in Experiment 2. We used the same picture primes as in the first two experiments and created nonwords from the spoken words used in Experiments 1 and 2. In the monosyllabic words, we changed the coda (e.g., PAAL was replaced by PAAG). For the disyllabic words, we changed the second syllable (e.g., KASTEEL was replaced by KATROOG). The segmental structure was kept constant. The spoken items were newly recorded to avoid splicing artifacts. Again, only SOA = 600 msec was tested. The analysis of the RTs and errors was performed exactly like the analysis in Experiment 2. The percentage of outlier trials was 5.8.

RESULTS AND DISCUSSION

The mean RTs, standard deviations and error percentages for Experiment 3 are given in Table 3.1. The table shows that there is still a facilitation effect for /k/ as well as for /p/. There were no significant main effects or interactions in the error analysis. The analysis of the RTs showed a main effect of context, $F_1(1,11) = 5.67$, $MSE = 6435$, $p = .036$, $F_2(1,4) = 8.47$, $MSE = 2910$, $p = .044$. There was no significant main effect of phoneme, $F_1(1,11) < 1$, $MSE = 14855$, $p = .61$, $F_2(1,4) < 1$, $MSE = 3838$, $p = .77$, nor was there an interaction between phoneme and context, $F_1(1,11) < 1$, $MSE = 1287$, $p = .44$, $F_2(1,4) = 1.47$, $MSE = 3646$, $p = .29$. The priming effect was numerically larger in Experiment 2 (59 msec) than in Experiment 3 (36 msec). This suggests that there is some lexical influence (at any rate, the lexical effect was not expected to be big, because the target phonemes were all word initial). However, there was no statistical difference in effect of context between Experiments 2 and 3, as indicated by the absence of an interaction between context and experiment, $F_1(1,22) < 1$, $MSE = 4208$, $p = .42$, $F_2(1,18) < 1$, $MSE = 3391$, $p = .72$. Thus, the priming effect seems for the most part due to sublexical connections. To summarize, in this third experiment with phoneme targets in nonwords, the facilitation effect was again obtained, suggesting that the locus of the effect is at the sublexical level of processing. If the facilitation effect shown in Experiment 2 is due to the sublexical connections between the production and comprehension systems, it should also appear when nonwords are used. This was indeed the case.

Our working model assumes that there are phonological links between the production and comprehension systems. The evidence from Experiments 1-3 leaves open the question of whether these links are between phonemes or between their phonological features. To investigate this, we ran Experiment 4.

EXPERIMENT 4

The fourth experiment examined whether the influence between production and comprehension is phonemic in nature or whether the influence occurs at the level of phonological features. If it is the case that phonological features are involved, then it should also be possible to pre-activate a certain phoneme by another phoneme that shares most but not all of its phonological features. Although the priming effect would probably be smaller than by priming the full feature set of a phoneme, it should be observable if it is feature priming that has contributed significantly to the priming effects in Experiments 1-3.

METHOD

Participants. 12 new participants (5 male, 7 female; mean age: 23 years) from the MPI subject pool took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, Design, Procedure and Analysis. The target phonemes were part of the existing words used in Experiments 1 and 2. We replaced the picture primes for the phoneme /p/ by pictures whose names had an initial phoneme differing in one phonological feature. For the target phoneme /p/ (voiceless plosive), we selected pictures whose names started with /b/ (voiced plosive). For the target phoneme /k/ (voiceless plosive), we selected pictures whose names started with / / (voiced fricative). The phonemes /k/ and / / differ in two phonological features. The experimental procedure and analysis were the same as in Experiments 2 and 3. The percentage of outlier trials was 5.2.

RESULTS AND DISCUSSION

The mean RTs, standard deviation, and error percentages for Experiment 4 are given in Table 3.1. There were no response time effects in the experiment. Most importantly, there was no main effect of context, $F_1(1,11) < 1$, $MSE = 6143$, $p = .60$, $F_2(1,4) < 1$, $MSE = 6802$, $p = .99$. If anything, the response times were numerically larger in the match than in the nonmatch condition, exactly the opposite of what was obtained in Experiments 1-3. There were also no significant effects in the error analysis.

The absence of a priming effect suggests that the links from production to comprehension concern connections between phonemes rather than their features. The absence of a priming effect in the present experiment cannot be due to a lack of experimental power, because the power of the experiment was the same as that of Experiments 1-3, where priming effects were obtained. Moreover, the response times were numerically larger in the match than in the nonmatch condition. So, it was not the case that there was an effect that was too small to be detected. If anything, there was an effect in the wrong direction. To conclude, there is no priming effect when the prime and target phonemes share most but not all of their phonological features. The special status of phonemic identity suggests that the priming effect is mediated by phonemic representations, as assumed by our working model.

EXPERIMENT 5

In Experiments 1-4, participants had to respond to the picture primes by monitoring for phonemes in the picture names (Experiment 1) or by naming the pictures (Experiments 2-4). The fifth experiment examined whether priming is obtained when participants passively view the pictures. The WEAVER++ model (Levelt et al., 1999) implements the claim that only selected lemmas spread activation to the phonological level. Thus, the model predicts that no priming should be obtained with passive viewing only.

METHOD

Participants. 20 new participants (5 male, 15 female; mean age: 21 years) from the MPI subject pool took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, Design, Procedure, and Analysis. We used the same materials and design as in Experiments 1 and 2. In the present experiment, participants only had to respond if they heard the pre-specified phoneme in the presented spoken word. There were no tones and there was no additional task to force them to do anything with the pictures. We tested SOA = 600 msec only. After the experiment, the participants were given a recognition test for the pictures. This was a paper sheet with the 20 test pictures on it plus 20 pictures that were not in the experiment. The participants had to indicate which pictures they had seen. This was done to verify that the participants looked at the screen during the experiment. The mean recognition accuracy in the picture recognition test after the experiment was 95.25 %. The analysis of the response times and errors in the experiment was performed exactly like the analysis in Experiment 1 but without the between participants factor of SOA. The percentage of outlier trials was 4.6.

RESULTS AND DISCUSSION

The mean RTs, the standard deviations, and the error percentages for Experiment 5 are given in Table 3.1. The table shows that there is no facilitation effect for either the phoneme /k/ or for /p/ . If anything, RTs are longer in the match than the nonmatch condition.

There were no significant main effects or interactions in the error analysis. The analysis of the reaction times yielded no effect of context, $F_1(1,19) < 1$, $MSE = 2050$, $p = .53$, $F_2(1,9) < 1$, $MSE = 1854$, $p = .37$. There was, however, a main effect of phoneme, $F_1(1,19) = 9.69$, $MSE = 3431$, $p = .006$, $F_2(1,9) = 37.02$, $MSE = 783$, $p < .001$. There was no interaction between context and phoneme, $F_1(1,19) = 4.87$, $MSE = 754$, $p = .040$, $F_2(1,9) < 1$, $MSE = 1402$, $p = .91$. Pairwise comparisons between matching and nonmatching contexts yielded neither a significant effect for phoneme /k/, $t_1(19) = 1.64$, $p = .12$, $t_2(9) < 1$, $p = .37$, nor for phoneme /p/, $t_1(19) < 1$, $p = .55$, $t_2(9) < 1$, $p = .59$.

In summary, no priming was obtained when participants only passively viewed the pictures. This finding corresponds to the prediction by WEAVER++, in which only selected lemmas spread activation to the phonological level.

GENERAL DISCUSSION

We reported five experiments that examined the influence of spoken word planning on speech comprehension. Participants were presented with pictures and a tone or spoken word presented 300 or 600 msec after picture onset. When a spoken word was presented, participants had to indicate whether it contained a pre-specified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme (Experiment 1) or to name the picture (Experiment 2). The phoneme monitoring latencies for the spoken words were shorter when the picture name contained the pre-specified phoneme compared to when it did not. The facilitation was only obtained at the SOA of 600 msec. The facilitation was also obtained for spoken nonwords (Experiment 3). Facilitation was only obtained when the full phoneme was primed rather than most of its phonological features (Experiment 4). No facilitation was obtained when the picture required no response (Experiment 5). These results suggest that there are sublexical form connections running from the word production system to the word comprehension system, as stipulated by the perceptual-loop theory of self-monitoring.

The results of Experiment 5 suggest that passive picture viewing does not lead to sufficient phonological activation to cause a significant facilitation effect in an external phoneme monitoring task. Although participants were paying attention to the pictures, as indicated by the recognition scores obtained after the experimental session of Experiment 5, the attention given to the pictures was apparently not enough to cause the priming effect. We discuss two possible reasons for why this was the case.

First, it could be that there is discrete information flow within the production system, with only selected lemmas activating the corresponding word forms, as implemented in WEAVER++ (Levelt et al., 1999). With passive viewing, the task does not require the selection of lemmas. Consequently, word forms do not become activated, and picture-induced priming of external phoneme monitoring does not occur, as observed in Experiment 5. In contrast, in monitoring for phonemes in internal speech (Experiment 1) or in picture naming (Experiments 2 and 3), a lemma needs to be selected in order to accomplish the task. Consequently, picture-induced priming of external phoneme monitoring should occur, as observed in Experiments 1-3.

Second, the information flow through the production system could be cascading, but the distance in the lexical network from concepts to word forms might be too long to obtain much activation at the phonological level. For example, Dell (1986) proposed that retrieval of information from the lexical network involves jolting the activation of target nodes at each processing level. These activation jolts will only occur when the task requires the retrieval of lexical information. In the absence of any activation jolts, the network distance between concepts and phonological forms may be too far to obtain much phonological activation from irrelevant pictures. This would explain why priming effects were obtained in Experiments 1-3, where the task required lexical retrieval, but not in Experiment 5, where lexical retrieval was not required.

The results of Experiments 2 and 3 (with picture naming as the control task) show that influences of speech planning on speech recognition are obtained even when there is no explicit self-monitoring task. This suggests that the encoding of a phonological representation for the picture is sufficient to yield the priming effect. This seems to contradict an assumption made for the WEAVER++ model by Roelofs (2004a, 2004b). In WEAVER++, the phonological links from production to comprehension are established as a consequence of constructing a phonological word representation for the picture name. Roelofs (2004a) assumed that the feeding of the phonological representation into the comprehension system is under a speaker's control. Encoded phonological representations are sent to the comprehension system for monitoring purposes. However, the findings of Experiments 2 and 3 demonstrate that an explicit internal monitoring task is not required to obtain the influence from production planning on speech comprehension. It may be the case, though, that the default control setting of the production system is to allow for self-monitoring. This means that a constructed phonological representation is automatically sent to the comprehension system. The priming effects in Experiments 2 and 3 suggest that speakers encoded phonological representations of the picture names on all trials. If phonological representations had only been encoded upon hearing the tone (indicating that the picture had to be named), then no priming effect should have been obtained in the phoneme monitoring task. If encoded phonological representations are automatically sent to the speech comprehension system, even when there is no explicit internal monitoring task, the priming effects on phoneme monitoring with picture naming as control task (Experiments 2 and 3) are explained.

In WEAVER++, phonological word representations are constructed incrementally from the beginning of a word to its end. Correspondingly, the phonological word representation is fed into the comprehension system as it becomes available over time (Roelofs, 2004a). This results in sequential activation of the comprehension system, as is the case with the processing of external speech. The sequential activation of the comprehension system by the production system concurs with the effect of the serial position of target phoneme that was obtained by Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) for the monitoring of internal speech.

To conclude, the reported experiments provide evidence for internal links running from speech planning to speech comprehension, as entailed by the perceptual-loop theory of self-monitoring. Speech planning influences speech recognition in the context of picture naming and internal self-monitoring tasks, but not in the context of passive picture viewing. Moreover, the influence is obtained when the planned phonological representation and the perceived speech share phonemes, but this is not the case when only a subset of a phoneme's phonological features are shared, even when this is the majority of features. These results suggest that the encoding of phonological representations is a precondition for obtaining word planning influences on speech perception.

PERCEPTUAL UNIQUENESS POINT EFFECTS IN MONITORING INTERNAL SPEECH

CHAPTER 4

REBECCA ÖZDEMIR, ARDI ROELOFS, & WILLEM J. M. LEVELT[3]

ABSTRACT

There is disagreement concerning how speakers monitor their internal speech. Production-based accounts assume monitoring mechanisms within the production system (e.g., Laver, 1973), whereas comprehension-based accounts, such as the perceptual-loop theory of Levelt and colleagues (Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999), assume monitoring through the speech comprehension system. Comprehension-based monitoring predicts speech-perception specific effects, like the perceptual uniqueness point effect (Frauenfelder, Segui, & Dijkstra, 1990), in the self-monitoring of internal speech. We ran an extensive experiment testing this hypothesis using internal phoneme monitoring and picture naming as tasks. Our results show an effect of the perceptual uniqueness point of a word in internal phoneme monitoring, but no such effect in picture naming. These results support the perceptual-loop theory and challenge production-based monitoring accounts.

---

[3] A version of this chapter is submitted for publication.

INTRODUCTION

People typically listen to the speech of others. However, the person they listen to most often is probably themselves. Speakers monitor their own speech for errors and appropriateness (e.g., Levelt, 1989). In the literature, different accounts exist of how this monitoring is achieved (see Postma, 2000, and Hartsuiker & Kolk, 2001, for recent reviews). Probably all models of self-monitoring assume external monitoring, whereby the speaker monitors self-generated overt speech. This involves the normal speech comprehension process. Self-monitoring models also agree that, in addition, mechanisms exist for the monitoring of the internal speech plan before it is articulated. However, the models make different claims about the functional locus of the internal monitoring device. One class of model assumes that the internal monitoring device is located inside the production system (e.g., Laver, 1973, 1980; Schlenk, Huber, & Wilmes, 1987). Another class of model assumes that internal monitoring is achieved via the speech comprehension system (e.g., Levelt, 1989).

In production-based accounts, there are typically several special-purpose monitors associated with the planning levels in production. Each monitoring device checks a particular aspect of speech planning, such as lexical selection or phoneme selection (e.g., Laver, 1980). In comprehension-based accounts, the monitoring of internal speech is achieved via the speech comprehension system. Such an account has been developed by Levelt and colleagues (Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999), called the perceptual-loop theory of self-monitoring. According to Levelt et al. (1999), in planning spoken words, phonological representations are constructed incrementally from the beginning of a word to its end. The phonological word representations are fed into the speech comprehension system as they become available over time. This results in sequential activation of the comprehension system, as is the case with the processing of real external speech. The comprehension system is then used to monitor the planned speech.

Evidence that incrementally constructed phonological representations underlie the self-monitoring of internal speech was provided by Wheeldon and Levelt (1995). Their participants were native speakers of Dutch who fluently spoke English. The participants had to monitor for target phonemes in the Dutch translation equivalent of visually presented English words. For example, they had to indicate by means of a button press (yes/no) whether the phoneme /n/ is part of the Dutch translation equivalent of the English word WAITER. The Dutch word is *kelner*, which has /n/ as the onset of the second syllable, so requiring a positive response. All Dutch target words were disyllabic. To assess the time course of monitoring phonological representations, the serial position of the target phonemes in the Dutch words was manipulated. The target phoneme could be the onset or coda of the first syllable, or the onset or coda of the second syllable. Phoneme monitoring latencies increased with the serial position of the segments within the word. In order to verify experimentally whether phonological rather than phonetic representations were monitored, participants had to perform the phoneme monitoring task while simultaneously counting aloud, which is known to suppress the maintenance of phonetic representations. The monitoring latencies were longer with the counting task, but the serial position effect was replicated, suggesting that self-monitoring involves a phonological rather than a phonetic representation. Recently, Wheeldon and Morgan (2002) replicated the seriality effect in internal phoneme monitoring in English using native speakers of English. Participants first had to learn pairs of prompt and response words. Next, they were presented with the prompt and had to monitor for pre-specified phonemes in the response words. The monitoring latency patterns observed by Wheeldon and Levelt (1995) were replicated.

Because self-monitoring is achieved via the speech comprehension system according to the perceptual-loop theory, it predicts perception-specific effects on internal self-monitoring. One such perception-specific effect is the uniqueness point effect (e.g., Frauenfelder, Segui, & Dijkstra 1990; Marslen-Wilson, 1990). The uniqueness point (UP) of a word is defined as the phoneme in the word where it diverges from all other words in the language, going from the beginning of a word to its end. The uniqueness point of words influences the speed of word recognition. For example, Marslen-Wilson (1990) showed that lexical decisions are made faster when the uniqueness point is early in a word compared to when it is late in a word. Moreover, in phoneme monitoring experiments, participants are faster at detecting a target phoneme in a word when the phoneme follows the uniqueness point of a word compared to when the phoneme precedes the uniqueness point (Frauenfelder et al., 1990). The serial distance between the target phoneme and the uniqueness point of the word plays a role as well. If the target phoneme follows the uniqueness point of a word, phoneme monitoring is faster when the distance of the phoneme to the uniqueness point is greater (Frauenfelder et al., 1990). Whereas the uniqueness point of a word affects spoken word recognition, there is no evidence to suggest that it affects spoken word production.

According to the perceptual-loop theory, the uniqueness point of a word should not only affect the monitoring of external speech, but also the monitoring of internal speech. Assume a speaker has to decide whether the name of a pictured cat contains the phoneme /k/. Consider the activation flow for such internal phoneme monitoring in a model that combines WEAVER++ (e.g., Roelofs, 1997, 2003) and MERGE (Norris et al., 2000; Norris, 1994), illustrated in Figure 4.1. According to WEAVER++, phonological word representations for intended words are constructed incrementally from the beginning of a word to its end. The phonological representation is fed into the comprehension system as it becomes available over time. In MERGE (see left side of Figure 4.1), there are sublexical as well as lexical connections to phoneme decision nodes used in performing a phoneme monitoring task. Both the lexical and sublexical links send activation to the phoneme decision nodes as soon as the target phoneme is present in the input to the comprehension network. Because word candidates compete in the model, the amount of activation sent to the decision node depends on the activation level of the target word relative to its competitors. Decision latencies should be shorter when the competition is resolved early (i.e., when the word has an early uniqueness point) than when it is resolved late (i.e., when the uniqueness point is late in the word).

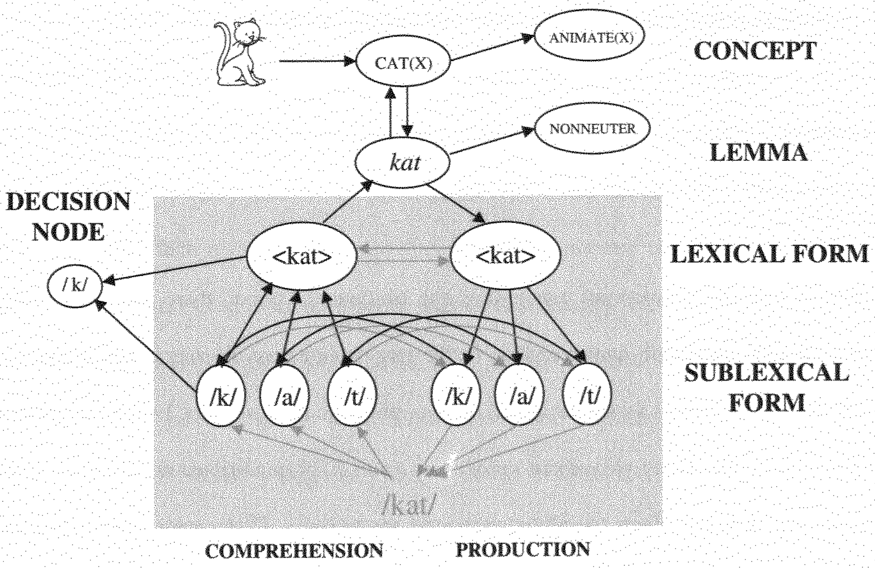


Figure 4.1 *Working model for internal phoneme monitoring, combining the WEAVER++ of spoken word planning (Levelt et al., 1999; Roelofs, 1997, 2004a) and the MERGE model of phoneme monitoring (Norris et al., 2000; Norris, 1994). The links from production to comprehension are dynamic in that they are established as a result of encoding a phonological word (Roelofs, 2004a).*

We report an experiment that was designed to examine whether there are perception-specific effects in the monitoring of internal speech. Participants were presented with pictured objects and they had to indicate whether the picture name contained a pre-specified phoneme. We manipulated the position of the target phonemes relative to the uniqueness point of the picture names. This was done in order to test the critical prediction that monitoring latencies should depend on the distance of the phoneme from the uniqueness point of the picture name. Moreover, in order to provide a replication of the results of Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) using pictures, we manipulated the serial position of the target phoneme within the picture names. Effects of serial position and uniqueness point should be present in self-monitoring of internal speech but not in picture naming. In order to test the latter prediction, participants were also asked to name the pictured objects.

EXPERIMENT

MATERIALS

In order to test the predictions concerning the uniqueness point, there were 30 critical pictures, all with disyllabic names ending in the target phonemes /l/ or /r/. The uniqueness point of the picture names was determined using a phonetic dictionary of Dutch (Heemskerk & Zonneveld, 2000). There were three *uniqueness point* (UP) conditions: no-distance, small-distance, large-distance. Each condition contained ten pictures. Five names ended in /l/ and five names ended in /r/. In the no-distance condition, the uniqueness point of the picture name was the last phoneme itself, /l/ or /r/ (e.g., "ketel"). In the small-distance condition, the picture name became unique at the phoneme before the last one (e.g., "vog*e*l"). In the large-distance condition, the picture name became unique two phonemes before the end of the word (e.g., "za*d*el"). The full set of test materials is presented in Appendix 4-A. The test items in the different conditions were matched for frequency (CELEX database, Baayen, Piepenbrock, & Gulikers, 1995) and number of segments. To avoid any confounds due to the between-items design, the pictures were tested in two pretests with respect to their ease of recognition and differences in articulation (see section *Pretests*).

Additionally, there were 240 filler pictures. Of those picture names, 60 contained the target phonemes (e.g., /l/) in medial or word initial position ("mo*l*en" - windmill; "*l*eraar" - teacher) and 90 filler pictures did not contain the target phoneme. Moreover, we also included 90 pictures with the target phonemes /k/, /n/ and /s/ (15 go- and 15 nogo-trials each). The 45 go-trials varied with respect to the target phoneme's position: 15 word initial, 15 word medial, and 15 word final (five of each phoneme). The picture names in the different serial position conditions were matched on frequency and length. This subset of materials was used to measure the sensitivity of the method of using picture-induced self-monitoring, which has not been used before to study internal monitoring. Wheeldon and Levelt (1995) used a translation paradigm rather than response to pictures to investigate internal monitoring, and Wheeldon and Morgan (2002) used associated word pairs. As indicated, Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) found that detection latencies are influenced by the phoneme's serial position in the word. Phonemes at the beginning of a word are detected faster than word medial phonemes, which are in turn detected faster than word final phonemes. The goal of including the set of 45 phoneme-position trials was to replicate the effect of position found by Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002), to show that our task taps into the internal monitoring process.

PRETESTS

Because both manipulations, UP distance and serial position, were tested between items, there were two other factors which could possibly influence the response latencies. First, it may be that the latencies are influenced by the different amounts of time it takes to recognize the pictures in the different conditions. Second, although there was no overt speech signal present in the monitoring task, it cannot be excluded that the different onsets of the picture names influence the latencies in the picture naming task. To control for these influences, we ran two pretests, delayed naming (to test for onset differences) and picture recognition (to test for differences in ease of recognition).

The pretests and the main experiment were controlled by the NESU (Nijmegen Experimental Setup) software developed at the Max Planck Institute. In the picture naming sessions, a microphone was connected to an electronic voice key, which in turn was connected to a computer that recorded the response times. Response latencies were written to hard disk after each trial. During the experiment the experimenter sat in the same room and noted down hesitations, wrong responses, or failures of the voice key.

Delayed naming. 20 participants (2 male, 18 female; mean age: 21 years) performed a delayed naming task. In this task, participants had to name pictured objects on the screen after receiving a delayed go-signal (a tone). The delay between picture presentation onset and the go-signal gave participants enough time to prepare the response up to but not including its actual articulation. Thus, higher-level processing influences prior to articulation should not influence the latencies. The pictures were presented on the screen for 1000 msec. The delay period between the pictures and the go-signal was 1000 msec. Table 4.1 gives the mean reaction times, error percentages and standard deviations. The statistical analysis of the RTs yielded no significant main effect of UP-distance, $F_1(2, 57) < 1$, $MSE = 8106$, $p = .830$, $F_2(2,27) = 1.90$, $MSE = 3695$, $p = .169$. There was also no main effect of serial position, $F_1(2,57) < 1$, $MSE = 682$, $p = .984$, $F_2(2,42) < 1$, $MSE = 4707$, $p = .380$. There were no effects in the error analyses. To conclude, the critical materials did not differ in ease of articulation onset.

Table 4.1 *Mean Response Latencies (in Milliseconds), their Standard Deviations, and the Error Percentages in the Delayed Naming Pretest.*

UP Distance	Mean	SD	E %
no	611	235	3.7
small	613	222	1.7
large	579	223	1.9
Serial position			
initial	576	222	1.6
medial	588	217	1.4
final	592	236	0.8

Picture recognition. 20 new participants (6 male, 14 female; mean age: 21 years) performed a picture recognition task. A spoken word was presented via headphones. Next, 500 msec after word offset, a picture was presented on the screen for 1000 msec. Participants had to press a green button when word and object referred to the same entity (*yes*-response) and to press a red button when word and picture referred to different entities (*no*-response). Table 4.2 gives the mean reaction times for the *yes*-responses to the critical pictures, error percentages and standard deviations. The statistical analysis of the RTs yielded no significant main effect of UP distance, $F_1(2, 57) < 1$, $MSE = 1426$, $p = .917$, $F_2(2,27) < 1$, $MSE = 2372$, $p = .645$. There was also no main effect of serial position, $F_1(2, 57) < 1$, $MSE = 10086$, $p = .518$, $F_2(2,42) = 2.8$, $MSE = 7517$, $p = .072$. There were no effects in the error analyses. To conclude, the critical pictures did not differ in ease of recognition.

Table 4.2 *Mean Response Latencies (in Milliseconds), their Standard Deviations, and the Error Percentages in the Picture Recognition Pretest.*

UP Distance	Mean	SD	E %
no	535	153	4.0
small	534	169	5.0
large	550	207	5.5
Serial position			
initial	535	170	4.3
medial	557	183	5.0
final	512	176	2.3

MAIN EXPERIMENT

Participants. 32 native speakers of Dutch (9 male, 23 female; mean age: 22 years) participated in the main experiment. None of them had participated in one of the pretests. The participants were undergraduate students of Radboud University Nijmegen. They all had normal or corrected-to-normal vision and were paid for their participation.

Design. The main experiment crossed the factors *task* (picture naming, internal phoneme monitoring) and *UP distance* (no distance, small distance, large distance) within participants.

Also, the factors *task* (picture naming, internal phoneme monitoring) and *serial position* (word initial, word medial, word final) were crossed within participants. Participants first performed the picture naming task and then the phoneme monitoring task.

Procedure. Participants were tested individually in a quiet room. They sat in front of a computer screen wearing headphones. Each participant first had to perform a full session of picture naming. They received written instructions to name the displayed pictures as fast and accurately as possible. The pictures were presented in a randomized order. After the picture naming session, there was a short break. Next, participants received written instructions for the internal phoneme monitoring task. They were asked to press a button with their dominant hand when the name of the pictured object contained the target phoneme. The target phoneme (/l/, /r/, /k/, /n/, and /s/) changed trial-by-trial. The internal phoneme monitoring task consisted of 10 practice trials followed by five experimental blocks containing 54 items each. There were short breaks between the blocks. The whole experimental session lasted about 35 minutes.

Each trial in the picture naming session had the following structure. A picture was presented and stayed on the screen for 1000 msec. A new target picture appeared on the screen either after the response (voice key trigger) or after a time out period of 2000 msec. In the internal phoneme monitoring session the target phoneme was displayed on the screen for 500 msec in the form of the corresponding letter (*l*, *r*, *k*, *n*, or *s*). After 1500 msec the picture appeared and stayed on the screen for 1000 msec. A new target phoneme was displayed either after the response (button press) or after a time out period of 2500 msec. For the internal phoneme monitoring task, a push-button box with one key was used to register response latencies.

Analysis. Trials on which participants stuttered, used names other than those intended, corrected themselves afterwards, smacked their lips, or made other noises triggering the voice key were excluded from the analysis of the picture naming latencies. In the internal phoneme monitoring task wrong button presses and items that were named incorrectly in the picture naming session (on a participant basis) were excluded from the analysis. The response latencies were analyzed using repeated measures analyses of variance with the factors task and UP distance. Interactions were further explored by ANOVAs and *t*-tests. For the serial position manipulation, analyses were performed using the factors task and serial position. All comparisons were tested within participants and between items. In all analyses, an alpha level of .05 was adopted.

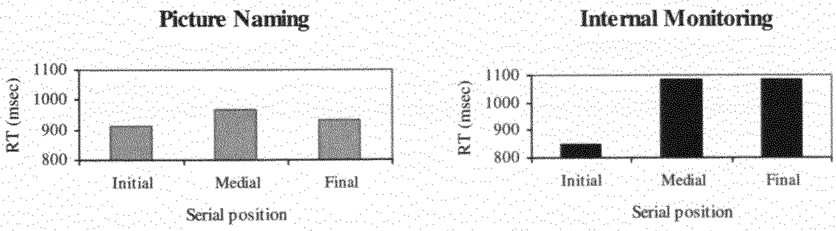
RESULTS AND DISCUSSION

Serial Position. Table 4.3 and Figure 4.2 show the mean latencies of the picture naming and the internal monitoring tasks for the serial position manipulation. There was no difference in the reaction times in picture naming, whereas there were differences in the phoneme monitoring latencies.

For the picture naming task, no main effect of serial position was found, $F_1(2,93) < 1$, $MSE = 12631$, $p < .500$, $F_2(2,42) < 1$, $MSE = 12749$, $p = .450$. However, for the phoneme monitoring task, there was a main effect of serial position, $F_1(2,93) = 22.73$, $MSE = 666690$, $p < .001$, $F_2(2,42) = 28.54$, $MSE = 302959$, $p < .001$.

Table 4.3 *Mean Response Latencies (in Milliseconds), their Standard Deviations, and the Error Percentages per Manipulation (Serial Position, UP Distance) and Task (Internal Monitoring, Picture Naming).*

	Internal Monitoring			Picture Naming		
Serial Position	Mean	SD	E %	Mean	SD	E %
Initial	846	219	2.1	915	247	2.9
Medial	1083	329	7.3	966	270	3.6
Final	1092	303	10.5	931	235	4.6
UP distance						
No	960	250	2.8	827	220	2.4
Small	932	219	4.9	860	242	3.5
Large	898	218	4.4	863	231	2.8


Picture Naming

Internal Monitoring

Figure 4.2 *Mean reaction times in msec per condition observed for the manipulation of the position of the target phoneme within the word for picture naming and internal phoneme monitoring.*

Pairwise comparisons for the monitoring latencies revealed significant differences between initial and medial position, $t_1(31) = 9.36, p < .001, t_2(28) = 6.17, p < .001$, and between initial and final position, $t_1(31) = 10.51, p < .001, t_2(28) = 7.83, p < .001$, but not between medial and final position, $t_1(31) = 1.15, p = .257, t_2(28) < 1, p = .595$. Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) observed a similar pattern. Thus, we replicated the earlier findings in the literature using pictures rather than to-be-translated words or prompt words as target stimuli.

UP distance. Table 4.3 and Figure 4.3 show the means for the two tasks concerning the UP distance manipulation. Again, there seem to be no differences in the naming latencies while the means for the internal monitoring responses seem to differ.

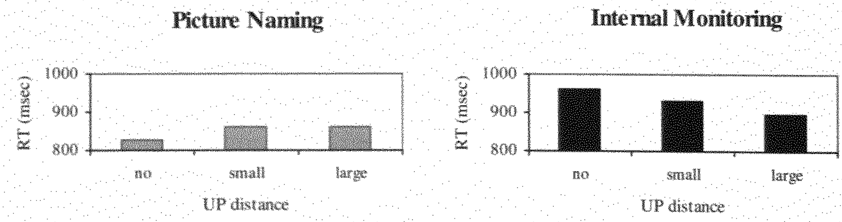


Figure 4.3 *Mean reaction times in msec per condition observed for the manipulation of the distance of the target from the uniqueness point of the word for picture naming and internal phoneme monitoring.*

Overall, the statistical analysis yielded a main effect of task, $F_1(1,31) = 18.42$, $MSE = 17091$, $p < .001$, $F_2(1,27) = 22.45$, $MSE = 3086$, $p < .001$, but no main effect of UP distance, $F_1(2,62) < 1$, $MSE = 5131$, $p = .558$, $F_2(2,26) < 1$, $MSE = 1961$, $p = .894$. There was an interaction between task and UP distance, $F_1(2,62) = 6.86$, $MSE = 6194$, $p = .002$, $F_2(2,26) = 3.45$, $MSE = 10631$, $p = .047$.

For the picture naming task, there was no effect of UP-distance, $F_1(2,93) = 1.84$, $MSE = 98820$, $p = .159$, $F_2(2,26) < 1$, $MSE = 1899$, $p = .881$. Pairwise comparisons revealed no significant differences in any of the comparisons.

For the internal phoneme monitoring task, the analysis yielded a significant effect of UP distance in the participant analysis, $F_1(2,93) = 4.37$, $MSE = 230224$, $p = .013$, $F_2(2,26) = 1.93$, $MSE = 10693$, $p = .165$. Pairwise comparisons revealed a significant difference between the no-distance and large-distance conditions, $t_1(31) = 3.18$, $p < .001$, $t_2(17) = 1.88$, $p < .04$, and a significant difference between the small-distance and large-distance conditions in the participant analysis, $t_1(31) = 1.82$, $p < .04$, $t_2(17) < 1$, $p = .190$. The difference between the no-distance and small-distance conditions was not significant, $t_1(31) = 1.57$, $p = .064$, $t_2(17) = 1.16$, $p = .132$.

GENERAL DISCUSSION

We reported an experiment in which participants performed picture naming and internal phoneme monitoring tasks. The crucial manipulation was the distance between the last phoneme of the picture name and the uniqueness point of the name. There were three conditions: no distance (the last phoneme was the uniqueness point of the word), small distance (one phoneme distance to the uniqueness point), and large distance (two phonemes distance to the uniqueness point). A perception-based monitoring account predicts an effect of uniqueness-point distance for the internal monitoring task but not for the picture naming task, whereas a production-based monitoring account predicts no effects at all. Our results show an effect of UP distance in the internal phoneme monitoring task but not in the picture naming task. These results support the predictions of the perceptual-loop theory.

The critical effects were significant in the analyses by participants, but some of the effects were not significant in the by-item analyses. The absence of some of the effects in the item analysis is probably due to the between-items design. Having only ten items per condition reduces the power in the by-item analysis compared to the by-participant analysis. However, it was not possible to find more items that met the constraints on the experimental materials. The trend found in the by-item analysis, however, goes in the same direction as the effects in the by-participant analysis.

Whereas the results support the predictions by the perceptual-loop theory, they seem to challenge production-internal monitoring accounts. However, we have to be careful in excluding the possibility, that some version of production-internal monitoring may account for the present findings. As Postma (2000) argued, a major problem with evaluating production-internal monitoring accounts is that they are highly underspecified. Whereas computational implementations exist for the perceptual-loop theory (e.g., Hartsuiker & Kolk, 2001: Roelofs, 2004a), this is not the case for production-internal monitoring models. Still, the present findings put important constraints on any such production-internal monitoring account. First, a production-internal monitoring model should exhibit a uniqueness point effect in self-monitoring but not in speech production. Second, the model should exhibit a uniqueness point effect in self-monitoring in the absence of overt speech. That is, the uniqueness point effect should be obtained by monitoring *internal* speech. Our perception-based monitoring model (illustrated in Figure 4.1) exhibits all these properties, and therefore provides an important benchmark for possible future production-internal monitoring accounts of our findings. It would have been nice to show that in a comprehension version of this experiment with these materials the observed uniqueness point effect would show up (perhaps even stronger). We tried to do this, but there were serious problems which made the results unusable. In a comprehension phoneme monitoring experiment you start measuring the RT with the beginning of the critical segment. Since the test materials all had an /l/ or /r/ as critical segment which was preceded by a schwa most of the time, it was not possible to determine the exact starting position of the critical segment in the recordings. Using the best starting positions that could be determined was obviously not good enough to get clear results. However, given the clear evidence from other studies (e.g., Frauenfelder et al., 1990), it would have been nice but not needed to show the effect again.

Vigliocco and Hartsuiker (2002) argued for both production-based and comprehension-based self-monitoring of internal speech. "Inner speech exists: It is phonetic in nature, we are consciously aware of it, and we can inspect it through self-perception. However, it does not continue when we speak aloud, because of articulatory suppression. There is, at the same time, inner monitoring when we speak aloud, but this type of monitoring is internal to the production system; has access to, inter alia, phonological codes; and does not lead to conscious awareness" (pp. 466-467). However, Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002) obtained evidence for internal monitoring of phonological codes in the absence of overt speech, which was replicated in the present experiment. Wheeldon and Levelt (1995) provided evidence that the monitored representations were phonological rather than phonetic, because the RT patterns for the monitoring persisted when the monitoring task was combined with a secondary articulation task. Moreover, evidence by Özdemir, Roelofs, and Levelt (submitted, Chapter 3) suggests that the comprehension system is fed by phonological rather than phonetic representations. Picture primes speed up phoneme monitoring in external spoken words when the picture name contains the target phoneme, but there is no priming when the phoneme in the picture name and the spoken word share most but not all of their phonological features. To conclude, contrary to the claim of Vigliocco and Hartsuiker (2002), the evidence suggests that monitoring of internal speech concerns phonological rather than phonetic representations; the monitoring of phonological representations can be performed in the face of a concurrent articulatory task.

Still, it may be argued that monitoring for errors and monitoring for phonemes in internal speech are accomplished via different mechanisms. Perhaps monitoring for phonemes is accomplished via the speech comprehension system, as suggested by the present experimental results, whereas monitoring for errors is not. Note that this dual-system proposal is theoretically less parsimonious than assuming that monitoring for errors and phonemes are accomplished via the same system. Thus, unless there is specific evidence that demands the division of labor, the single-system view is to be preferred. Moreover, the single-system view is supported by empirical evidence. Dell and Repka (1992) asked participants to detect speech errors in internally and externally generated tongue twisters. They observed that similar types of errors were reported in internal and external speech. However, in internal but not in external speech, the errors occurred predominantly at the beginning of words. This suggests that the participants internally monitored a representation that was incrementally generated from the beginning of a word to its end, and that the internal generation process was stopped after the detection of an error. The seriality effects observed for both internal error monitoring and for phoneme monitoring suggest that the two types of monitoring are accomplished via the same system.

To conclude, we reported evidence that the uniqueness point of words influences phoneme monitoring in internal speech, but it does not influence picture naming. The most parsimonious explanation of this finding is that the monitoring of internal speech is accomplished via the speech-comprehension system, as maintained by the perceptual-loop theory of self-monitoring.

SUMMARY AND CONCLUSIONS

CHAPTER 5

The relationship between speech production and speech comprehension is an interesting and challenging topic for language research. Language users are able to comprehend and produce speech simultaneously at rapid rates. Although this is a fascinating phenomenon, it has not gained much attention in experimental research in the past. For a long time it was assumed that if one was able to find out how speech comprehension works, the problem of how speakers produce speech would be halfway resolved. But it soon became clear that this is not the case and that comprehension and production are crucially different processes. Therefore researchers started to examine both speech modalities independently from each other. They have developed detailed theories and models for production and comprehension, which have been tested extensively using different experimental paradigms and computer simulations. However, the relationship between these two main modalities of speech was rarely of interest.

The topic of this thesis is exactly this underresearched area of language, namely the relationship between speech production and speech comprehension. One aim of the three experimental series was to learn more about how the two modalities influence each other in simultaneous processing. Another aim was to find evidence concerning the question of whether speech comprehension and speech production share representational units or whether each modality has separate representations for its own purposes. The focus of investigation was on the form levels of processing. In this final chapter the results from the three experimental series (Chapters 2-4) are summarized and discussed in the light of different speech processing accounts. The first two experimental series aimed at identifying influences from one modality on the other when both of them are active. The first series investigated the influence of ongoing speech perception on the performance of a speech production task. The second series looked at the other direction of influence, namely from ongoing speech planning to the performance of a speech comprehension task. The third experimental series examined the existence of a comprehension-specific effect in internal monitoring.

The goal of the first experimental series (Chapter 2) was to identify influences that an auditorily presented distractor with phonological overlap (comprehension processing) has on a picture naming task (production performance). It is known from several previous experiments that phonological overlap of distractor and target in the picture-word interference paradigm causes facilitation irrespective of the position of the overlap (e.g. Meyer & Schriefers, 1990; Levelt et al., 1999). From these studies it is not clear, however, where the exact locus of this facilitation effect is. It is likely that the effect arises at phonological processing levels. However, it may also be that the phonological similarity leads to an extra amount of activation at the lexical levels of processing. To localize the phonological facilitation effect exactly, in this series of experiments participants performed a gender decision and an animacy decision task on the same materials. Compared to the picture naming task, these two tasks controlled for higher-level processing influences. If an effect were found in the picture naming task but not in one of the higher-level tasks, the effect can be localized at the form levels of processing.

Each picture in Experiment 1 was combined with four distractors, two related conditions (begin overlap and end overlap) and two unrelated conditions. The related conditions differed in the position of the overlap (begin versus end overlap). A pure sublexical influence based on phonological preactivation should result in the same amount of facilitation for both conditions, whereas different effects would suggest additional lexical influences. Three SOAs were tested (-150, 0, +150) to examine the timing of the effects.

The results showed that there were no effects at any of the SOAs for the gender decision and the animacy decision tasks. There was, however, facilitation for both overlap conditions at all SOAs in the picture naming task. The effect sizes for the two conditions only differed at the earliest SOA, where the effect for the begin overlap condition was smaller than for the end overlap condition. It was argued that this modulation of the effect by the position of overlap was due to the fact that at this early SOA the distractor words were fully recognized and highly active when the picture name entered the stage of word-form encoding. This made the begin-related distractor a serious cohort competitor to the picture name. The encoding process of the picture name was hindered by this competitor, which lead to a smaller facilitation effect in the begin overlap condition than in the end overlap condition. To test whether it was really cohort membership of distractor and target that caused the competition and therefore the decreasing of facilitation in the begin overlap condition, Experiment 2 was conducted.

In Experiment 2 the same pictures were used and combined with the begin-related distractors from Experiment 1. Additionally, they were combined with fragments consisting of only the overlapping segments. If it is true that it was cohort membership that caused competition of the picture name and distractor word, this decrease should not be evident in the fragment condition. The results showed that the fragments caused significantly more facilitation than the whole words. This supports the conclusion that cohort competition is present in the context of begin overlapping words due to the mismatching segments, which decreases the facilitation effect. In Experiment 2, small but significant amounts of facilitation were also found in the gender decision task, suggesting a small contribution to the phonological facilitation effect also from the lemma level. The same amount of facilitation also showed up in the gender decision task at SOA –150 in Experiment 1, but did not turn out to be significant. It is, however, important to note that the modulation by lexical status was only found in the picture naming task. The modulation can therefore be localized at a lexical form level rather than at the lemma level.

From this pattern of results it was argued that the main locus of phonological facilitation effects in the picture-word interference paradigm is at the form level of processing. Although there seems to be a small contribution to the effect caused by lemma level processing, the bulk of the effect is due to form level processing. It was suggested that these effects are the result of activation spreading via two types of connecting links between the comprehension and production form levels of processing. The strong facilitation for both positions of overlap is caused by activation spreading from the comprehension phoneme representations using sublexical links, causing a preactivation in the phoneme representations of the production system. These preactivations lead to faster retrieval of these phonemes in the phonological encoding process. This shows up in the faster response times in the case of phonological overlap compared to unrelated distractors, irrespective of the position of this overlap. The modulation of the effect by the position of overlap (given enough time to fully recognize the distractor as is the case at the earliest SOA) is argued to be due to information sent from the comprehension to the production system via lexical form links. Being phonologically similar at the beginning of the word causes stronger interference in the process of morpheme retrieval for production than being phonologically similar at the end of the word. An important fact to mention here is that this modulation of the effect by the position of overlap (Experiment 1) or lexical status (Experiment 2) is not a higher-level lexical effect because it was not present in the higher-level decision tasks. From this experimental series it was concluded that there are sublexical as well as lexical form links connecting the two systems in the direction from comprehension to production. The sublexical links cause facilitation in the process of phonological encoding by preactivating the respective phonemes in the production system. The lexical links cause interference when the distractor is fully recognized and therefore a competitor to the target picture name in the process of morpheme retrieval. This competition seems to be stronger for cohort overlap than for rhyme overlap.

Although the results fit the predictions of our working model, an interactive activation account in which comprehension and production share form representations and processing pathways would also be able to explain the result pattern found in this experimental series. The distractor would activate its respective phoneme units. Because of the fact that these are the same representational units as the ones used by speech production, an effect of preactivation and therefore facilitation would arise. The phoneme nodes would have a certain activation level already because of the distractor activation and would therefore be easier to retrieve in the production process. However, in such a model where activation flows up and down through the system at any time, one would expect to find an effect in the higher-level tasks. On the basis of these experiments it is clear that the two modalities influence each other in the direction from comprehension to production. However, it is not possible to decide whether representations are shared or separated between speech comprehension and production.

In the second series of experiments (Chapter 3) the other direction of influence was investigated, namely the sublexical influence from production to comprehension representations. The question was whether it would also be possible to preactivate phoneme representations in the comprehension system by ongoing speech planning processes. In contrast to influences from spoken words on speech production, which was examined before, this direction of influence has not been investigated previously. Participants performed a standard phoneme monitoring task, in which they had to press a button when a prespecified phoneme was present in a word that was presented via headphones. This task was adapted to fit the purposes of investigating the influence of ongoing speech planning on this task. Before presenting the auditory word, a picture was displayed on the screen, whose name either also contained the target phoneme (*match* condition) or not (*mismatch* condition). The prediction was that if there are sublexical influences from speech production to speech comprehension there should be an effect of preactivation observable in the match condition compared to the mismatch condition.

Two target phonemes were tested, /p/ and /k/, either embedded in existing words (Experiments 1, 2, 4, and 5) or pseudo-words (Experiment 3). The pictures were presented at two different SOAs, 300 msec and 600 msec before the auditory target. At the 300 msec SOA no effect of the condition was expected to occur because it takes participants 400 to 450 msec to get to the stages of phonological encoding after picture onset (see Levelt et al., 1991). Thus, 300 msec is not enough time to retrieve the phonemes in the production system and therefore also not enough time to see an effect on the comprehension task. The 600 msec interval, on the other hand, is enough time to activate the phonemes in the production system and send this activation to the comprehension system. An effect should therefore be observable at this late SOA. To force participants to encode the picture name instead of just passively looking at the screen, a second task was introduced. In one third of the trials a tone was presented after the prime picture instead of a word. In these trials the participants had to switch the source and either perform their phoneme monitoring on the picture name (Experiment 1), or simply name the picture (Experiments 2 - 4). This additional task forced the participants to pay attention to the pictures and encode their names. Experiment 5 examined what happens without this control task.

In Experiments 1 and 2, facilitation was found for the match condition when the target phoneme was in the context of existing words. However, this facilitation could have been due to a combination of sublexical and lexical influences (see connections to the decision nodes from sublexical as well as lexical levels in Figure 1.1). Experiment 3 was run to show the pure sublexical effect and filter out lexical contributions to the effect. The target phonemes were embedded in non-existing words to encourage the system to base its phoneme detection on sublexical information only. Again, a facilitation effect was found. The lexical influences on the effect from the comprehension side of processing were minimized. The picture names could, however, still have contributed to the effect via lexical connections from production to comprehension, and from there to the decision nodes (see Figure 1.1). The facilitation effect was smaller for the pseudo-words than for the words, suggesting some lexical influence. The statistical difference in effect size between words and pseudo-words was not significant. However, this was not to be expected because the targets were all word initial and it is hard to obtain lexical effects in such a situation.

Knowing that matching phonemes can cause facilitation, the question arose as to whether the unit of preactivation was the phoneme or whether it was even smaller representations like phonological features. Experiment 4 addressed this question. The target phonemes (in the context of existing words) were preceded by pictures whose initial phonemes differed in one or two phonological features with the target phoneme. No preactivation was observed in the two-feature mismatch, nor in the one-feature mismatch condition. There was thus no evidence for influences from production to comprehension at the phonological feature level.

Another interesting question was whether the effect would also arise when participants were just passively viewing the screen, without any intention to encode the picture names. Experiment 5 was run, which had the same material set as Experiment 1, but without the tone condition. No effect was observed in this experiment, suggesting that merely viewing the pictures on the screen without the intention to encode their names is not enough to get speech planning started and therefore also not enough to observe an effect in the comprehension task.

From this experimental series it was concluded that there are dynamic links connecting the sublexical levels of the two modalities in the direction from production to comprehension. To establish these dynamic links, the intention to construct a phonological word representation (internal monitoring and picture naming as control tasks) is needed. When this is not the case, as in passive viewing, the links seem to be absent. The activation spreading via these links caused a preactivation of the phonemes in the comprehension system and therefore shorter latencies in the phoneme monitoring task. It was possible to exclude the lexical influence from the comprehension side of processing by having participants perform the monitoring on non-existing words (Experiment 3). It was, however, not possible to exclude a lexical effect due to an activation spreading of the picture name to the comprehension system via lexical form links. With this design, we were unable to address the question of the existence or nonexistence of lexical form links from production to comprehension. The results of Experiment 4 suggest that the communication between the systems is indeed at the level of phonemes and not at an even lower level like the level of phonological features, supporting the assumptions of WEAVER++, in which phonological word representations mediate between production and comprehension.

Although the working model can explain the results that were found, the results from this series of experiments can also not determine whether there are shared or separate representational units for production and comprehension. An interactive activation account with shared representations could also explain the result pattern. Assuming separate representations, however, also does not necessarily make connecting links between production and comprehension mandatory. The results found in this series of experiments can also be explained by a direct link from the production phoneme units to the phoneme monitoring device (e.g., the decision nodes), causing a preactivation there and therefore the facilitation of the response. The decision nodes would then need less input from the comprehension system to reach their response threshold. Although this would entail extra processing structures and would therefore be less economical, it is a possible alternative.

The third experimental series (Chapter 4) was designed to enable us to decide between the two possible sources of the effects found in the second experimental series (an extra link to decision nodes from the production representations versus monitoring via the comprehension system). The question was addressed using the predictions of two different accounts of self-monitoring of speech. Production-based monitors assume that the monitoring of an inner speech plan is done on production internal representations (reflecting a direct link from production representations to the monitoring device). The perceptual-loop theory, on the other hand, assumes that monitoring makes use of the comprehension system and is performed on comprehension internal representations. The crucial difference is that the perceptual-loop theory predicts comprehension-specific effects in monitoring an inner speech plan, which production-based monitors do not. The uniqueness point effect was chosen as a test case to choose between these two monitoring theories. This would clarify whether the effects found in Chapter 3 were due to activation spreading via links from production to comprehension representations, or to direct links from production representations to the decision nodes.

Two tasks were performed on a set of pictures, picture naming and internal phoneme monitoring. In the picture naming task no uniqueness point effect was expected because it is known to be an effect specific to the serial processing of speech and its recognition processes. For the internal phoneme monitoring task, however, the perceptual-loop theory predicts an effect of uniqueness. The inner speech plan is fed into the comprehension system and is serially processed like external speech. Thus, it was predicted that an effect of being unique and recognized at a certain point would be shown. Production-based monitors do not predict such an effect because the monitoring is based on production internal representations and processes, in which phonemes become available in parallel and uniqueness should play no role. It is important to note that both tasks, picture naming and internal phoneme monitoring, are based on the same activation flow in the system. There was no external speech signal present at any time during the phoneme monitoring experiment. Therefore, a dissociation in effects between the two tasks can only be due to the involvement of different representations (production versus comprehension) and processes (parallel versus serial activation). The results can therefore also be interpreted with respect to the question concerning shared or separate representations for speech production and comprehension.

The target phoneme was always in the last position of the word and varied with respect to its distance from the perceptual uniqueness point of the word. There were three conditions, *no-distance* (target phoneme is the uniqueness point of the word), *small-distance* (target phoneme is one phoneme after the uniqueness point of the word), and *large-distance* (target phoneme is two phonemes after the uniqueness point of the word). The prediction was that if there is involvement of comprehension-specific representations and serial processing in the internal monitoring task, as predicted by the perceptual-loop theory, the detection times for the target phoneme should become faster the greater the distance from the uniqueness point.

A production-based monitor should not show this effect because of the parallel availability of phonemes in the process of phonological encoding.

The results showed that there was no uniqueness point effect in the picture naming task, as predicted by production- as well as comprehension-based accounts. There was, however, an effect of distance from the uniqueness point in the internal phoneme monitoring task. This pattern of results is compatible with the perceptual-loop theory and challenges production-based monitors. The findings suggest that the uniqueness point effect in a production monitoring task is due to the feeding of the speech plan into the comprehension system for monitoring purposes. These results suggest that there have to be separate form representations for the two modalities, with comprehension representations and serial processing being responsible for the uniqueness point effect.

The working model has all the components necessary to explain the performance and results in this last experiment. The MERGE model was combined with the WEAVER++ model and used to explain phoneme monitoring. The monitoring of inner speech is achieved via the links connecting the production and comprehension form levels. The uniqueness point effect arises because of the serial processing of the speech plan inside the comprehension system.

Interactive activation models in which production and comprehension share the same representations and pathways, although able to explain the result patterns of the first two experimental series (Chapter 2 and 3), are challenged by the results of this last experiment (Chapter 4). Neither models assuming parallel activation of phonological content nor models assuming gradient activation of phonological content can explain the presence of the uniqueness point effect in the self-monitoring.

A question that remains open is whether a view expressed by Zwitserlood (1994) would be able to explain the results of this last crucial experiment. She assumes that there are separate feedforward pathways for the two modalities but shared representations. Imagine the activation of the picture name arriving at the phonological stages. Whether they become available in parallel or serially does not matter for the present purposes. The critical point now is that Zwitserlood (1994) assumes that there is no feedback in the system. However, bottom-up serial processing of the phoneme string is needed to obtain an uniqueness point effect. Thus, there has to be a trigger to start the bottom-up processing of the picture name. Since there is no external speech signal present, there is, therefore, no motivation for the system to activate the recognition pathways in this experiment. This alternative model thus seems to be challenged by the results of the last experiment as well.

After reviewing the results of the experimental series presented in this thesis and discussing their implications for existing models of speech processing, I now come back to the main research questions posed at the beginning and to see how the experiments have contributed to the answers to these questions.

Do speech perception and production influence each other?

This question can certainly be answered with *yes*. In every experiment the introduced influence from the other modality (comprehension influence in Chapter 2 and production influence in Chapter 3) had a strong effect on the task performance.

If so, at what levels of processing do they influence each other?

The first two experimental series investigated this question and were able to identify at least two levels of influence. Concerning the comprehension to production direction of influence (Chapter 2), connecting links at the sublexical level and at the lexical form level were found on the picture naming task. In the other direction from production to comprehension (Chapter 3), there was evidence for connecting links at the sublexical level. Evidence for lexical form links from production to comprehension could neither be confirmed nor rejected with this design.

Are there separate representations for perception and production at the form levels of processing?

The last experiment (Chapter 4) showed a comprehension specific effect, the uniqueness point effect, in an internal monitoring task. This suggests that representations outside the production system have to be involved in the performance of this task, leading to the assumption that sublexical representations for production and comprehension are separate.

Although some of these results could also be explained in an interactive activation model assuming shared representations, in such a model one would assume more effects on the higher-level tasks in the first experimental series, for example, due to the information flowing up and down through the system. Also, if one system were serving both modalities there should have been feature priming in Experiment 4 of Chapter 3, since features play a role in speech comprehension. Last but not least, the uniqueness point effect in the self-monitoring task in the main experiment of Chapter 4 challenges interactive activation models assuming shared representations for both modalities. The possibility of a model in which representations are shared and pathways are separated is an interesting one that has to be further specified to make it testable in the future. Taken together, the results of the three experimental series presented in this thesis suggest a speech processing model with separated but closely linked form levels of processing for perception and production like the working model illustrated in Figure 1.1.

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.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database.* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.

Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science, 25*, 663-678.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177-208.

Connine, C. M., Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes, 11*, 635-645.

Cutler, A., Mehler, J., Norris, D., Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*, 141-177.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124-142.

Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 237-262). New York: Plenum Press.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801-838.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes?. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9-37). Berlin: Mouton de Gruyter.

Eimas, P. D., Hornstein, S. B., Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29*, 160-180.

Foss, D. J., Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12*, 1-31.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory?. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113-157.

Heemskerk, J. S. M., & Zonneveld, W. (2000). *Uitspraakwoordenboek*. Utrecht: Het Spektrum.

Janse, E., & Quené, H. (1999). On the suitability of the cross-modal semantic priming task. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1937-1940). San Francisco.

Janse, E., & Quené, H. (2004). On measuring multiple lexical activation using the cross-modal semantic priming technique. In H. Quené & V. van Heuven (Eds.), *On speech and language: Studies for Sieb G. Nooteboom* (pp. 105-122). Utrecht: LOT Occasional Series.

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 281-294). Hillsdale, N.J.: Erlbaum.

Laver, J. D. M. (1973). The detection and correction of slips of tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand.* New York: Academic Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 98,* 13464-13471.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122-142.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills.* New York: Springer.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In: G. Altmann (Ed.). *Cognitive models of speech processing* (pp. 148-172). Cambridge: MIT Press.

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

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

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 1146-1160.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimental elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behaviour, 21*, 578-594.

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

Norris, D., McQueen, J., Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-370.

O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word-production. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 57-73.

Özdemir, R., Roelofs, A., Levelt, W. J. M. (submitted). *Influences of Spoken Word Planning on Speech Recognition.*

Özdemir, R., Roelofs, A., Levelt, W. J. M. (submitted). *The locus of phonological facilitation by spoken distractors in picture naming.*

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 539-557.

Pitt, M. A., Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, 29, 149-188.

Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S. J., & Friston, K. J. (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain, 119*, 919-931.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249-284.

Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 55,* 465-483.

Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 115-158). Berlin: Mouton de Gruyter.

Roelofs, A. (2004a). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review, 111,* 561-572.

Roelofs, A. (2004b). Comprehension-based versus production-internal feedback in planning spoken words: A rejoinder to Rapp and Goldrick (2004). *Psychological Review, 111,* 579-580

Rubin, P., Turvey, M. T., van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in nonwords. *Perception and Psychophysics, 19,* 394-398.

Schlenk, K, Huber, W., & Wilmes, K. (1987). "Prepairs" and repairs: different monitoring functions in aphasic language production. *Brain and Language, 30,* 226-244.

Schriefers, H., Meyer, A., Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29,* 86-102.

Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11,* 17-50.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In: E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17-51). Hillsdale, NJ: Erlbaum.

Starreveld, P. A. (2000). On the interpretation of auditory context effects in word production. *Journal of Memory and Language, 42*, 497–525.

Starreveld, P. A., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes, 19*, 677-711.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin, 128*, 442-272.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34*, 311-334.

Wheeldon, L. R., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes, 17*, 503-535.

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

Appendix 2-A

Materials for Experiments 1 and 2 (Chapter 2)

Picture	Distractor					
	related begin	related end	unrelated begin	unrelated end	related fragment	unrelated fragment
			Animate, Nonneuter			
zwaan (*swan*)	zwaai (*swing*)	traan (*tear*)	halm (*stalk*)	wond (*wound*)	zwaa	ha
hond (*dog*)	homp (*chunk*)	wond (*wound*)	gang (*corridor*)	traan (*tear*)	ho	ga
beer (*bear*)	beek (*brook*)	peer (*pear*)	neut (*drop*)	mand (*basket*)	bee	neu
kat (*cat*)	kas (*cashdesk*)	mat (*mat*)	voeg (*joint*)	lans (*lance*)	ka	voe
gans (*goose*)	gang (*corridor*)	lans (*lance*)	zwaai (*swing*)	keus (*choice*)	ga	zwaa
hand (*hand*)	halm (*stalk*)	mand (*basket*)	beek (*brook*)	peer (*pear*)	ha	bee
neus (*nose*)	neut (*drop*)	keus (*choice*)	homp (*chunk*)	hoed (*hat*)	neu	ho
voet (*foot*)	voeg (*joint*)	hoed (*hat*)	kas (*cashdesk*)	mat (*mat*)	voe	ka

Appendix 2-A (continued)

Materials for Experiments 1 and 2 (Chapter 2)

Picture	Distractor					
	related begin	related end	unrelated begin	unrelated end	related fragment	unrelated fragment
			Animate, Neuter			
hert (*deer*)	helm (*helmet*)	snert (*trash*)	haak (*hook*)	kaap (*cape*)	he	haa
kalf (*calf*)	kam (*comb*)	zalf (*salve*)	kam (*comb*)	snert (*trash*)	ka	ka
lam (*lamb*)	lak (*lacquer*)	dam (*dam*)	beet (*bite*)	blaar (*blister*)	la	bee
paard (*horse*)	paal (*post*)	kaart (*card*)	oost (*east*)	zalf (*salve*)	paa	oo
schaap (*sheep*)	schaal (*scale*)	kaap (*cape*)	helm (*helmet*)	peen (*carrot*)	schaa	he
haar (*hair*)	haak (*hook*)	blaar (*blister*)	schaal (*scale*)	dam (*dam*)	haa	schaa
been (*leg*)	beet (*bite*)	peen (*carrot*)	lak (*lacquer*)	toog (*bar*)	bee	la
oog (*eye*)	oost (*east*)	toog (*bar*)	paal (*post*)	kaart (*card*)	oo	paa

Appendix 2-A (continued)

Materials for Experiments 1 and 2 (Chapter 2)

Picture	Distractor					
	related begin	related end	unrelated begin	unrelated end	related fragment	unrelated fragment
Nonanimate, Nonneuter						
bril (*glasses*)	brink (*farmyard*)	gil (*scream*)	dom (*cathedral*)	last (*load*)	bri	do
rok (*skirt*)	rol (*role*)	stok (*stick*)	speen (*teat*)	trant (*style*)	ro	spee
want (*mitt*)	walm (*smoke*)	trant (*style*)	kalk (*lime*)	stok (*stick*)	wa	ka
sok (*socket*)	som (*sum*)	lok (*lock*)	kolk (*eddy*)	sfeer (*atmosphere*)	so	ko
dolk (*dagger*)	dom (*cathedral*)	wolk (*cloud*)	brink (*farmyard*)	gil (*scream*)	do	bri
speer (*spear*)	speen (*teat*)	sfeer (*atmosphere*)	rol (*role*)	wolk (*cloud*)	spee	ro
kast (*cupboard*)	kalk (*lime*)	last (*load*)	walm (*smoke*)	bom (*bomb*)	ka	wa
kom (*bowl*)	kolk (*eddy*)	bom (*bomb*)	som (*sum*)	lok (*lock*)	ko	so

Appendix 2-A (continued)

Materials for Experiments 1 and 2 (Chapter 2)

Picture	Distractor					
	related begin	related end	unrelated begin	unrelated end	related fragment	unrelated fragment
			Nonanimate, Neuter			
glas *(glass)*	glans *(glow)*	plas *(puddle)*	mep *(clout)*	les *(lesson)*	gla	me
bord *(plate)*	bonk *(lump)*	port *(postage)*	velg *(rim)*	plas *(puddle)*	bo	ve
schip *(ship)*	schil *(rind)*	wip *(seesaw)*	bel *(bell)*	mest *(menure)*	schi	be
vlot *(raft)*	vlok *(flake)*	grot *(cave)*	hal *(hall)*	port *(postage)*	vlo	ha
mes *(knife)*	mep *(clout)*	les *(lesson)*	glans *(glow)*	pret *(fun)*	me	gla
vest *(waistcoat)*	velg *(rim)*	mest *(menure)*	schil *(rind)*	wip *(seesaw)*	ve	schi
bed *(bed)*	bel *(bell)*	pret *(fun)*	bonk *(lump)*	kwark *(curd)*	be	bo
hark *(rake)*	hal *(hall)*	kwark *(curd)*	vlok *(flake)*	grot *(cave)*	ha	vlo

Appendix 3-A

Materials for Experiments 1 – 5 (Chapter 3)

Prime (Picture)		Target (Spoken Item)		
match	nonmatch	word	word (feature mismatch)	nonword
paard (*horse*)	kers (*cherry*)	poets (*trick*)	berg (*mountain*)	poers
peer (*pear*)	kam (*comb*)	paal (*post*)	bot (*bone*)	paag
pijl (*arrow*)	kast (*cupboard*)	puin (*rubble*)	bijl (*axe*)	puig
palm (*palm*)	kom (*bowl*)	pers (*press*)	bord (*plate*)	pert
pen (*pen*)	kaars (*candle*)	puk (*mite*)	bel (*bell*)	pun
puzzel (*puzzle*)	konijn (*rabbit*)	panter (*panther*)	borstel (*hair brush*)	pansug
penseel (*paintbrush*)	kameel (*camel*)	paleis (*palace*)	ballon (*balloon*)	paruif
pincet (*tweezers*)	kassa (*till*)	piraat (*pirat*)	banaan (*banana*)	piluuf
pistool (*pistol*)	ketel (*kettle*)	parfum (*perfume*)	bureau (*desk*)	parsin
passer (*compass*)	kegel (*cone*)	peddel (*paddle*)	bezem (*broom*)	pebbor

Appendix 3-A (continued)

Materials for Experiments 1 - 5 (Chapter 3)

Prime (Picture)		Target (Spoken Item)		
match	nonmatch	word	word (feature mismatch)	nonword
kers *(cherry)*	paard *(horse)*	korf *(basket)*	glas *(glass)*	kols
kam *(comb)*	peer *(pear)*	kist *(box)*	gril *(barbecue)*	kirs
kast *(cupboard)*	pijl *(arrow)*	koets *(coach)*	gans *(goose)*	koerl
kom *(bowl)*	palm *(palm)*	kan *(jug)*	galg *(gallows)*	kal
kaars *(candle)*	pen *(pen)*	kurk *(cork)*	gum *(rubber)*	kurl
konijn *(rabbit)*	puzzel *(puzzle)*	kasteel *(castle)*	gondel *(gondola)*	katroog
kameel *(camel)*	penseel *(paintbrush)*	karaf *(carafe)*	gitaar *(guitar)*	kalis
kassa *(till)*	pincet *(tweezers)*	kachel *(stove)*	geweer *(rifle)*	kafug
ketel *(kettle)*	pistool *(pistol)*	klaver *(clover)*	gewei *(antlers)*	klasuf
kegel *(cone)*	passer *(compass)*	koffer *(suitcase)*	gieter *(watering can)*	konnis

Appendix 4-A

Materials for the Main Experiment (Chapter 4)

	UP Distance		
Target	no	small	large
/l/	ketel (*kettle*)	bijbel (*bible*)	puzzel (*puzzle*)
	appel (*apple*)	wortel (*carrot*)	zadel (*saddle*)
	engel (*angel*)	kachel (*oven*)	stempel (*stamp*)
	tempel (*temple*)	trommel (*drum*)	deksel (*lid*)
	tafel (*table*)	vogel (*bird*)	spiegel (*mirror*)
/r/	kever (*beetle*)	masker (*mask*)	dokter (*doctor*)
	vlieger (*kite*)	boter (*butter*)	spijker (*nail*)
	kikker (*frog*)	koffer (*suitcase*)	wekker (*alarmclock*)
	tijger (*tiger*)	motor (*motorcycle*)	anker (*anchor*)
	halter (*bar-bell*)	vinger (*finger*)	ladder (*ladder*)

Appendix 4-A (continued)

Materials for the Main Experiment (Chapter 4)

	Serial Position		
Target	initial	medial	final
/k/	kanon (*canon*)	beker (*cup*)	asbak (*ash tray*)
	kano (*canoe*)	fakkel (*torch*)	rugzak (*backpack*)
	kegel (*bowling pin*)	stekker (*plug*)	zwempak (*swimsuit*)
	ketting (*chain*)	bliksem (*lightning*)	monnik (*monk*)
	kassa (*cash desk*)	sikkel (*sickle*)	handdoek (*towel*)
/s/	sleutel (*key*)	pistool (*revolver*)	vleermuis (*bat*)
	schommel (*seesaw*)	kussen (*pillow*)	infuus (*infusion*)
	snavel (*pecker*)	passer (*pair of compasses*)	cactus (*cactus*)
	schotel (*satellite dish*)	borstel (*brush*)	vleugels (*wings*)
	sigaar (*cigar*)	hamster (*hamster*)	kompas (*compass*)
/n/	nijlpaard (*hippopotamus*)	gondel (*gondola*)	eekhoorn (*squirrel*)
	neushoorn (*rhinocerus*)	honing (*honey*)	citroen (*lemon*)
	nagel (*finger nail*)	printer (*printer*)	kuiken (*chick*)
	navel (*navel*)	magneet (*magnet*)	druiven (*grapes*)
	nijptang (*nippers*)	panda (*panda bear*)	glijbaan (*slide*)

SAMENVATTING

De relatie tussen spraakproductie en spraakbegrip is een interessant en uitdagend onderwerp voor taalonderzoek. Taalgebruikers zijn in staat spraak tegelijkertijd en in hoog tempo te begrijpen en te produceren. Hoewel dit een fascinerend fenomeen is, heeft het in het verleden niet veel aandacht gekregen in experimenteel onderzoek. Lang werd aangenomen dat het probleem hoe sprekers spraak produceren al voor de helft zou zijn opgelost als men erachter kon komen hoe spraakbegrip werkt. Het werd echter duidelijk dat dit niet het geval is en dat spraakbegrip en spraakproductie zeer verschillende processen zijn. Daarom begonnen onderzoekers beide spraakmodaliteiten los van elkaar te bekijken. Zij hebben gedetailleerde theorieën en modellen voor spraakproductie en spraakbegrip ontwikkeld, die uitgebreid zijn getest met verschillende experimentele technieken en computersimulaties. Echter, de relatie tussen deze twee belangrijkste modaliteiten van spraak was zelden het onderwerp van onderzoek.

Het onderwerp van dit proefschrift is precies dit onderbelichte gebied van taal, namelijk de relatie tussen spraakproductie en spraakbegrip. Eén doel van de drie reeksen experimenten die worden gepresenteerd was om meer te weten te komen over hoe de twee modaliteiten elkaar beïnvloeden tijdens simultane verwerking. Een ander doel was om erachter te komen of spraakbegrip en spraakproductie representaties delen of dat elke modaliteit afzonderlijke representaties heeft voor zijn eigen doel. Het onderzoek was toegespitst op de vormniveaus van spraak. De eerste twee reeksen experimenten richtten zich op het identificeren van invloeden van de ene modaliteit op de andere als beide modaliteiten actief zijn. De eerste reeks experimenten onderzocht de invloed van spraakperceptie op het uitvoeren van een spraakproductietaak. De tweede reeks experimenten onderzocht de andere richting van invloed, namelijk de invloed van spraakplanning op het uitvoeren van een spraakbegripstaak. De derde reeks experimenten onderzocht het bestaan van een begripsspecifiek effect in interne-spraak monitoren, dat licht werpt op de vraag over gedeelde of afzonderlijke representaties.

Het doel van de eerste reeks experimenten (Hoofdstuk 2) was het bepalen van invloeden die een auditief gepresenteerde distractor met fonologische overlap (spraakbegrip) kan hebben op een plaatjesbenoemtaak (spraakproductie). Het is bekend van verschillende eerdere experimenten dat fonologische overlap van distractor en doelwoord facilitatie veroorzaakt in het plaatje-woord interferentieparadigma, ongeacht de positie van de overlap (b.v. Meyer & Schriefers, 1990; Levelt e.a., 1999). Uit deze resultaten wordt echter niet duidelijk waar binnen het woordplanningsproces dit facilitatie-effect ontstaat. Waarschijnlijk ontstaat het effect op fonologisch planningsniveau. Maar het is ook mogelijk dat de fonologische overeenkomst leidt tot een extra hoeveelheid activatie op het niveau van lexicaal plannen. Om het fonologische facilitatie-effect exact te lokaliseren, voerden proefpersonen in deze reeks experimenten een geslachtsbeslissingstaak en een levend/niet-levend beslissingstaak uit met dezelfde materialen. Vergeleken met de plaatjesbenoemtaak controleren deze twee taken voor invloeden van processen die plaatsvinden op hogere niveaus. Dus als er een effect wordt gevonden in de plaatjesbenoemtaak, maar niet in één van de processen die plaatsvinden op hogere niveaus, dan kan het effect worden gelokaliseerd op de vormniveaus van het planningsproces.

Plaatjes werden gecombineerd met begin- en eindgerelateerde distractor woorden en werden aangeboden met drie verschillende tijdsverschillen of 'Stimulus Onset Asynchronies'

(SOA's), namelijk -150, 0, en +150 ms. De resultaten lieten zien dat er helemaal geen effecten zijn voor de geslachtsbeslissingstaak en de levend/niet-levend beslissingstaak. Er waren echter wel facilitatie-effecten voor beide overlapcondities op alle SOA's in de plaatjesbenoemtaak, die in grootte alleen verschilden voor de vroegste SOA. De redenering is dat sublexicale vorminvloeden de facilitatie veroorzaakten en dat lexicale vorminvloeden (zoals cohortlidmaatschap) verantwoordelijk zijn voor het verschil in effectgrootte voor de vroegste SOA.

Experiment 2 testte deze lexicale hypothese. De plaatjes werden gecombineerd met spraakfragmenten die slechts bestonden uit de overlappende segmenten. Deze fragmenten waren zodoende ook cohortleden van de plaatjesnamen, maar geen cohortconcurrenten zoals de begin-gerelateerde woorddistractoren, aangezien de fragmenten de niet-overeenkomende segmenten missen. De resultaten lieten zien dat de fragmenten zelfs meer facilitatie veroorzaakten dan de volledige woorden, wat de conclusie van Experiment 1 bevestigt. In Experiment 2 werden ook kleine, maar significante facilitatie-effecten gevonden in de geslachtsbeslissingstaak, hetgeen een kleine bijdrage aan het fonologische facilitatie-effect vanuit het lemmaniveau suggereert. Het is echter belangrijk te vermelden dat de modulatie door lexicale status uitsluitend werd gevonden in de plaatjesbenoemtaak en daarom kan worden gelokaliseerd op het lexicale vormniveau en niet op het lemmaniveau.

Op basis van dit patroon van resultaten werd geconcludeerd dat de belangrijkste bron van fonologische facilitatie-effecten in het plaatje-woord interferentieparadigma het vormniveau van woordplannen is. Hoewel er een kleine bijdrage lijkt te zijn veroorzaakt door verwerking op het lemmaniveau, is het grootste deel van het effect het gevolg van verwerking op het vormniveau. Geopperd werd dat deze effecten voortvloeien uit het spreiden van activatie via twee soorten verbindingen tussen de vormniveaus van spraakbegrip en spraakproductie. De sterke facilitatie voor beide posities van overlap wordt veroorzaakt door activatie die spreidt van de begripsfoneemrepresentaties via sublexicale schakels, wat weer een preactivatie van de foneemrepresentaties van het productiesysteem veroorzaakt. De modulatie van het effect door de positie van de overlap (gesteld dat er genoeg tijd is om de distractor volledig te herkennen, zoals het geval is bij de vroegste SOA) wordt toegeschreven aan de informatie die wordt gezonden van het begrips- naar het productiesysteem via lexicale vormschakels. De verstoring van het proces van het ophalen van lexicale vormen voor productie is groter wanneer het begin van woorden fonologisch gelijk is, dan wanneer woorden aan het eind fonologisch op elkaar lijken. Het is belangrijk om hier te vermelden dat deze modulatie van het effect door de positie van de overlap (Experiment 1) of lexicale status (Experiment 2) geen lexicaal effect van een hoger niveau is, aangezien de modulatie niet aanwezig was in de beslissingstaken die via hogere niveaus werden uitgevoerd.

Hoewel de resultaten aansluiten bij de voorspellingen van ons werkmodel, is een interactieve-activatie verklaring, waarin spraakbegrip en spraakproductie vormrepresentaties en verwerkingspaden delen, ook in staat om het patroon van resultaten uit de reeks experimenten te verklaren. Op basis van deze experimenten is het duidelijk dat de twee modaliteiten elkaar beïnvloeden in de richting van spraakbegrip naar spraakproductie. Het is echter niet mogelijk om vast te stellen of begrip en productie representaties delen, of dat die onafhankelijk van elkaar zijn.

In een tweede reeks experimenten (Hoofdstuk 3) werd de andere richting van invloed onderzocht, namelijk de sublexicale invloed van spraakproductie op spraakbegrip. De vraag was of het ook mogelijk is om foneemrepresentaties in het begripssysteem te preactiveren door middel van simultane spraakplanningsprocessen. Proefpersonen namen deel aan een standaard foneemmonitoringstaak waarin ze op een knop moesten drukken wanneer een eerder gespecificeerd foneem aanwezig was in een woord dat werd aangeboden via een koptelefoon. De taak werd aangepast aan het doel van het onderzoek dat de invloed van gelijktijdige spraakplanning op de monitoringstaak onderzocht. Voordat het auditieve woord werd aangeboden, werd een plaatje op het scherm getoond waarvan de naam wél het doelfoneem bevatte (de zogenaamde *kloppende* conditie) of niet (de *niet-kloppende* conditie). De voorspelling was dat als er sublexicale invloeden van spraakproductie op spraakbegrip zijn, er een effect van preactivatie te zien zou moeten zijn in de kloppende conditie ten opzichte van de niet-kloppende conditie.

Twee doelfonemen werden gebruikt, /p/ en /k/, die ofwel waren ingebed in bestaande woorden (Experimenten 1, 2, 4 en 5) of in pseudo-woorden (Experiment 3). De plaatjes werden aangeboden met twee verschillende SOA's, 300 ms en 600 ms, gebaseerd op de kennis die er is over de tijd die het kost om de naam van een plaatje fonologisch te coderen (zie Levelt e.a., 1991). Om proefpersonen te dwingen op de plaatjes te letten en de namen te coderen in plaats van slechts passief naar het computerscherm te staren, werd een controletaak gebruikt.

Facilitatie werd gevonden voor de kloppende conditie wanneer het doelfoneem in de context van een bestaand woord voorkwam (Experimenten 1 en 2) en in pseudo-woorden (Experiment 3). Het facilitatie-effect was kleiner voor de pseudo-woorden dan voor de woorden, wat wijst op enige lexicale invloed. Er werd geen facilitatie-effect gevonden wanneer alleen de meeste fonologische kenmerken werden geprimed (Experiment 4) of wanneer er geen controle taak was (Experiment 5).

Op basis van de resultaten van deze reeks experimenten werd geconcludeerd dat er dynamische verbindingen bestaan tussen de sublexicale niveaus van de twee modaliteiten in de richting van spraakproductie naar spraakbegrip. Om deze dynamische verbindingen te construeren is de intentie nodig om een fonologische woordrepresentatie te maken (zoals aanwezig in de controle taak). Wanneer dit niet het geval is, zoals tijdens het passief kijken (Experiment 5), lijken de verbindingen afwezig. De activatie die spreid via deze verbindingen veroorzaakte een preactivatie van de fonemen in het begripssysteem en daardoor kortere reactietijden in de foneemmonitoringstaak. Met dit design waren we niet in staat om de vraag betreffende het bestaan dan wel de afwezigheid van lexicale vormverbindingen van productie naar begrip te beantwoorden. De resultaten van Experiment 4 wijzen erop dat de communicatie tussen de systemen inderdaad plaatsvindt op het niveau van fonemen en niet op een lager niveau zoals het niveau van fonologische kenmerken. Dit steunt de aannames van het werkmodel, waarin fonologische woordrepresentaties bemiddelen tussen spraakproductie en spraakbegrip.

Alhoewel het werkmodel de gevonden resultaten kan verklaren, kunnen de resultaten uit deze reeks experimenten geen uitsluitsel geven over de vraag of er gedeelde of gescheiden representaties zijn voor spraakproductie en spraakbegrip. Een interactieve-activatie verklaring met gedeelde representaties kan de resultaten ook verklaren. Zelfs als aangenomen wordt dat er gescheiden representaties zijn, dan is het nog niet noodzakelijkerwijs het geval dat verbindende schakels tussen productie en begrip bestaan. De gevonden resultaten in deze reeks experimenten kunnen ook worden uitgelegd door een directe verbinding van de productiefonemen naar het foneemmonitoringsmechanisme (b.v. de beslissingsknopen), wat preactivatie veroorzaakt en daardoor de facilitatie van de reactie. De beslissingsknopen zouden minder input vanuit het begripssysteem nodig hebben om hun responsdrempel te bereiken. Hoewel dit extra verwerkingsstructuren zou betekenen, en dus minder economisch zou zijn, is het wel een mogelijk alternatief.

De derde reeks experimenten (Hoofdstuk 4) was ontworpen om ons in staat te stellen te kiezen tussen de twee mogelijke bronnen van de gevonden effecten in de tweede reeks experimenten (een extra verbinding naar de beslissingsknopen vanuit de productierepresentaties versus monitoren via het begripssysteem). De vraag werd benaderd door gebruik te maken van de voorspellingen van de twee verschillende verklaringen wat betreft zelf-monitoren van spraak. Een productie-gebaseerde monitoringstheorie gaat ervan uit dat het monitoren van een intern spraakplan wordt gedaan op basis van productie-interne representaties (wat een directe verbinding tussen productierepresentaties en het monitoringsmechanisme weerspiegelt). De perceptuele-loop theorie, aan de andere kant, veronderstelt dat monitoren gebruik maakt van het begripssysteem en wordt uitgevoerd op begrips-interne representaties. Het cruciale verschil is dat de perceptuele-loop theorie begrips-specifieke effecten in het monitoren van een intern spraakplan voorspelt, wat een productie-gebaseerde monitoringstheorie niet doet. Het uniekheidspunteffect is gekozen als test om te kiezen tussen deze twee monitoringstheorieën. Dit zou duidelijk maken of de effecten in Hoofdstuk 3 het gevolg waren van spreidende activatie via verbindingen van spraakproductie naar spraakbegrip, of van directe verbindingen van productie representaties naar beslissingsknopen.

Twee taken werden uitgevoerd op een verzameling plaatjes, plaatjesbenoemen en intern foneemmonitoren. In de plaatjesbenoemtaak werd geen effect van het uniekheidspunt verwacht, omdat het bekend is dat het een effect is dat in het bijzonder optreedt bij seriële spraakverwerking en spraakherkenning. Voor de interne foneemmonitoringstaak voorspelt de perceptuele-loop theorie echter een effect van het uniekheidspunt. Het interne spraakplan wordt doorgegeven aan het begripssysteem en wordt serieel verwerkt zoals externe spraak. Dus, de voorspelling was dat een effect van het uniek zijn en herkend worden op een bepaald moment te zien zou zijn. Een productie-gebaseerde monitoringstheorie voorspelt echter niet zo'n effect, omdat het monitoren is gebaseerd op productie-interne representaties en processen, waarin fonemen parallel beschikbaar komen en het uniekheidspunt geen rol zou moeten spelen. Het is belangrijk te vermelden dat beide taken, plaatjesbenoemen en intern foneemmonitoren, gebaseerd zijn op dezelfde activatiestroom in het systeem. Er was geen extern spraaksignaal aanwezig gedurende het foneemmonitoringsexperiment. Dus, een dissociatie in effecten tussen de twee taken kan alleen het gevolg zijn van de betrokkenheid van verschillende representaties (productie versus begrip) en processen (parallelle versus seriële activatie). De resultaten kunnen daarom ook worden geïnterpreteerd met betrekking tot de vraag of er gedeelde of gescheiden representaties voor spraakproductie en spraakbegrip.

Het doelfoneem was altijd het laatste foneem van een plaatsjesnaam en varieerde in afstand tot het perceptuele uniekheidspunt van het woord. Er waren drie condities, *geen afstand* (het doelfoneem is het uniekheidspunt van het woord), *kleine afstand* (het doelfoneem volgt één foneem na het uniekheidspunt van het woord), en *grote afstand* (het doelfoneem volgt twee fonemen na het uniekheidspunt van het woord). De voorspelling was dat als er een betrokkenheid is van begrips-specifieke representaties en seriële verwerking in de interne monitoringstaak, zoals voorspeld door de perceptuele-loop theorie, de monitoringstijden voor het doelfoneem korter zouden moeten worden wanneer de afstand van het uniekheidspunt groter is (zoals gebleken is bij foneemmonitoren met externe spraak). Productie-gebaseerd monitoren zou dit effect niet moeten laten zien vanwege de parallelle beschikbaarheid van fonemen in het proces van fonologische codering.

De resultaten lieten zien dat er geen uniekheidspunteffect is in de plaatjesbenoemtaak, zoals voorspeld door beide verklaringen. Er was echter wel een effect van het uniekheidspunt in de interne foneemmonitoringstaak. Dit patroon van resultaten ondersteunt de perceptuele-loop theorie en is strijdig met een productie-gebaseerde monitoringstheorie. De bevindingen wijzen erop dat het uniekheidspunteffect in een productiemonitoringstaak veroorzaakt wordt door het doorgeven van het spraakplan naar het begripssysteem voor monitoringsdoeleinden. Hoewel interactieve activatiemodellen in staat zijn om de resultaten van de eerste twee reeksen experimenten te verklaren (Hoofdstuk 2 en 3), worden ze door de resultaten van dit laatste experiment betwist (Hoofdstuk 4).

De gepresenteerde resultaten hebben bijgedragen aan het beantwoorden van de vragen die waren gesteld in de introductie:

Beïnvloeden spraakperceptie en spraakproductie elkaar?

Deze vraag kan zeker worden beantwoord met *ja*. In ieder experiment had de geïntroduceerde invloed vanuit de ene modaliteit (de invloed van begrip in Hoofdstuk 2 en de invloed van productie in Hoofdstuk 3) een sterk effect op de uitvoering van een taak in de andere modaliteit.

Zo ja, op welke niveaus van verwerking beïnvloeden ze elkaar?

Wat betreft de richting van invloed van spraakbegrip naar spraakproductie (Hoofdstuk 2) werden er evidentie voor verbindende schakels op het sublexicale en het lexicale vormniveau gevonden in de plaatjesbenoemtaak. In de andere richting van productie naar begrip (Hoofdstuk 3), waren er aanwijzingen voor verbindende schakels op het sublexicale niveau. Een mogelijke hypothese over lexicale vormverbindingen van spraakproductie naar spraakbegrip kon niet worden bevestigd of verworpen.

Zijn er gescheiden representaties voor perceptie en productie op het vormniveau van verwerking?

Het laatste *experiment* (Hoofdstuk 4) liet een begrips-specifiek effect zien, het uniekheidspunteffect, in een interne foneemmonitoringstaak. Dit duidt erop dat representaties buiten het productiesysteem betrokken moeten zijn bij het uitvoeren van deze taak, wat leidt tot de aanname dat sublexicale representaties voor productie en begrip gescheiden zijn. Al met al wijzen de resultaten van de drie reeksen experimenten gepresenteerd in dit proefschrift op een spraakverwerkingsmodel met gescheiden maar sterk verbonden vormniveaus van verwerking voor spraakperceptie en spraakproductie, zoals in het werkmodel dat is afgebeeld in Figuur 1.1

ZUSAMMENFASSUNG

Die Untersuchung der Beziehung zwischen Sprachproduktion und Spracherkennung ist für die Sprachforschung von grossem Interesse und gleichzeitig eine Herausforderung. Sprecher sind in der Lage, Sprache in hoher Geschwindigkeit zu verstehen und zu produzieren, und das oft sogar zeitgleich. Obwohl dies ein faszinierendes Phenomen ist, hat es in der experimentellen Forschung bis heute nicht viel Beachtung gefunden. Lange Zeit wurde angenommen, dass, wenn man in der Lage wäre zu verstehen, wie die Spracherkennung funktioniert, auch das Rätsel der Sprachproduktion weitgehend gelöst wäre. Aber es wurde bald klar, dass dies nicht der Fall ist und Spracherkennung und Sprachproduktion grundlegend unterschiedliche Prozesse sind. Aus diesem Grund fingen Forscher an, die beiden Modalitäten gesprochener Sprache unabhängig voneinander zu untersuchen. Sie entwickelten detaillierte Theorien und Modelle für Sprachproduktion und Spracherkennung, welche ausgiebig anhand verschiedener experimenteller Paradigmen und mit Hilfe von Computersimulationen getestet wurden. Die Beziehung zwischen den beiden Modalitäten war jedoch selten im Mittelpunkt des Interesses.

Dieses wenig erforschte Gebiet der Sprachforschung, die Beziehung zwischen Spracherkennung und Sprachproduktion, ist das Thema der vorliegenden Dissertation. Ein Ziel der drei präsentierten Experimentserien war, etwas darüber zu lernen, wie die beiden Modalitäten sich in simultaner Verarbeitung gegenseitig beeinflussen. Ein weiteres Ziel war, Evidenz für die Beantwortung der Frage zu finden, ob Spracherkennung und Sprachproduktion die gleichen Repräsentationen innerhalb des Sprachsystem verwenden, oder ob jede Modalität separate Repräsentationen für ihre eigenen Zwecke hat. Der Fokus des

Interesses lag auf den Ebenen der phonologischen Formverarbeitung. Die ersten beiden Experimentserien (Kapitel 2 und 3) untersuchten den Einfluss der einen Modalität auf die andere, in Situationen in denen beide gleichzeitig aktiv sind. Die erste Serie untersuchte den Einfluss von aktiver Spracherkennung auf die Bewältigung einer Sprachproduktionsaufgabe. Die andere Richtung des Einflusses, die einer aktiven Sprachproduktion auf die Bewältigung einer Spracherkennungsaufgabe, war der Fokus der zweiten Experimentserie. In der dritten Experimentserie wurde schliesslich das Auftreten eines Effektes, der spezifisch für Spracherkennungsprozesse ist, in einer Sprachproduktionsaufgabe untersucht. Diese Fragestellung wurde gewählt, um die Frage nach geteilten oder getrennten Repräsentationen zu beantworten. Die Ergebnisse wurden im Hinblick auf verschiedene Sprachverarbeitungsmodelle zusammengefasst und diskutiert.

Das Ziel der ersten Experimentserie (Kapitel 2) war, Einflüsse eines auditiv präsentierten Distraktors mit phonologischer Überlappung (Spracherkennung) auf eine Bildbenennungsaufgabe (Sprachproduktion) zu identifizieren. Es ist aus früheren Studien bekannt, dass phonologische Überlappung von Distraktor und Target im Bild-Wort-Interferenz Paradigma eine Erleichterung der Bildbenennung bewirkt, unabhängig von der Position der Überlappung (e.g. Meyer & Schriefers, 1990; Levelt et al. 1999). Aus diesen Studien wird jedoch nicht klar, wo genau der Lokus der Erleichterung innerhalb des Systems ist. Es ist wahrscheinlich, dass der Lokus des Effekts die phonologischen Verarbeitungsebenen sind. Phonologische Ähnlichkeit könnte jedoch auch eine zusätzliche Aktivierung auf den lexikalischen Verarbeitungsebenen bewirken (z.B. Kohortenaktivierung) und dadurch zu einer Erleichterung der Verarbeitung führen. Um den Lokus des Effekts exakt zu bestimmen, nahmen die Versuchspersonen in dieser Experimentserie zusätzlich zur

Bildbenennung an einer Aufgabe zur Bestimmung des grammatischen Geschlechts, und an einer Bestimmung der Belebtheit teil. In Kombination mit der Bildbenennungsaufgabe kontrollierten diese beiden Aufgaben den Einfluss der höheren Verarbeitungsebenen auf den Effekt. Wenn ein Effekt in der Bildbenennung auftritt, nicht aber in einer der anderen Aufgaben, ist der Lokus des Effekts eine phonologische Formverarbeitungsebene.

Die Bildertargets wurden mit Distraktoren kombiniert, die phonologisch am Beginn oder am Ende des Bildnamens überlappten, und zu drei SOAs präsentiert (-150, 0, +150). Die Ergebnisse zeigten keinerlei Effekte für die Geschlechts- und Belebtheitsbestimmung. Es zeigten sich jedoch Erleichterungseffekte für beide Überlappungsbedingungen an allen SOAs in der Bildbenennungsaufgabe, deren Grösse sich nur an SOA –150 unterschied. Daraus wurde geschlussfolgert, dass sublexikalische Einflüsse die Erleichterung bedingten. Die Unterschiede in der Effektgrösse an der frühesten SOA wurden mit lexikalischen Einflüssen erklärt, die nur auftreten, wenn genügend Zeit für die vollständige Aktivierungsentfaltung des Distraktors gegeben ist.

Diese Schlussfolgerung wurde in Experiment 2 überprüft. Die Bilder wurden zusätzlich mit Fragmentdistraktoren kombiniert, die nur aus den überlappenden Segmenten bestanden. Diese Fragmente waren Kohorten*mitglieder* des Bilnamens, nicht jedoch Kohorten*konkurrenten* wie die beginnüberlappenden Wortdistraktoren, da sie keine unstimmigen Segmente enthielten. Wenn lexikalischer Wettbewerb den Unterschied in der Effektgrösse (in Experiment 1 an SOA –150) bedingt hat, sollten die Fragmente eine stärkere Erleichterung zeigen als die Wörter, da sie die Kohorte aktivieren (was den Zugriff und die Selektion erleichtert), jedoch keine Konkurrenz einführen. Die Fragmentdistraktoren führten zu mehr Erleichterung als die Wortdistraktoren, was die Schlussfolgerung aus Experiment 1 bestätigte. In Experiment 2 wurde zusätzlich ein signifikanter Erleichterungseffekt in der

Geschlechtsbestimmungsaufgabe gefunden, welches auf eine Beteiligung der Lemmaebene am phonologischen Erleichterungseffekt hindeutet. Es ist allerdings wichtig zu begreifen, dass die Modulation hinsichtlich des lexikalischen Status des Distraktors (Fragment versus Wort) nur in der Bildbenennungsaufgabe gefunden wurde.

Diese Ergebnisse der zwei Experimente waren die Grundlage anzunehmen, dass der Hauptlokus des phonologischen Erleichterungseffektes in Bild-Wort-Interferenz Experimenten die Formverarbeitungsebene ist. Auch wenn ein gewisser Anteil des Effektes auf der Lemmaebene aufzutreten scheint (signifikant in Experiment 2), ist der grösste Teil des Effektes eine Folge der Verarbeitung auf Formebenen. Es wurde argumentiert, dass diese Effekte das Ergebnis von Aktivierungen sind, die über zwei Verbindungstypen zwischen den Formverarbeitungsebenen von Spracherkennung und Sprachproduktion weitergegeben werden. Die starke Erleichterung für beide Überlappungspositionen wurde durch Aktivierungsübertragung auf der Phonemebene erzeugt, die eine Voraktivierung der entsprechenden Phonemrepräsentationen im Sprachproduktionssystem erzeugte. Somit konnten diese im Verlauf der Planung des Bildnamens schneller abgerufen werden. Die Modulation des Effekts bezüglich der Position der Überlappung (SOA –150 in Experiment 1) bzw. des lexikalischen Status (Experiment 2) wurde der Informationsübertragung vom Spracherkennungs- zum Sprachproduktionssystem über Verbindungen der lexikalischen Formverarbeitung (Morphemebene) zugeschrieben. Phonologische Ähnlichkeit am Wortanfang führte zu stärkerer Beeinträchtigung durch Konkurrenz während des Abrufens des Bildnamens als phonologische Ähnlichkeit am Wortende (Experiment 1). Diese Beeinträchtigung wurde aufgehoben durch die Entfernung unstimmiger Segmente (Experiment 2). Es ist wichtig zu erwähnen, dass diese Modulationen kein Effekt der höheren lexikalischen Verarbeitungsebenen sind, da sie nicht in den Kontrollaufgaben (Geschlechtsbestimmung, Belebtheitsbestimmung) aufgetreten sind.

Obwohl die Ergebnisse die Vorhersagen des Arbeitsmodelles (Figure 1.1) bestätigen, wäre ein interaktives Modell, in dem Spracherkennung und Sprachproduktion dieselben Repräsentationen und Verarbeitungswege nutzen ebenso in der Lage, die Ergebnisse dieser Experimentserie zu erklären. Die Experimente zeigen, dass eine aktive Spracherkennung die Sprachproduktionsprozesse beeinflusst. Es ist auf dieser Grundlage jedoch nicht möglich, zwischen geteilten oder getrennten Repräsentationen für Spracherkennung und Sprachproduktion zu unterscheiden.

In der zweiten Experimentserie (Kapitel 3) wurde die andere Richtung des Einflusses untersucht, nämlich der sublexikalische Einfluss von der Sprachproduktion auf die Spracherkennung. Die Frage war, ob es genauso möglich ist, Spracherkennungsrepräsentationen von Phonemen vorzuaktivieren durch aktive Sprachproduktionsprozesse. Die Versuchspersonen nahmen an einer Phonemüberwachungsaufgabe teil. Sie sollten einen Knopf drücken, wenn ein vorher spezifiziertes Phonem in einem über Kopfhörer präsentierten Wort vorhanden war. Diese typische Spracherkennungsaufgabe wurde den speziellen Bedürfnissen der Fragestellung. Vor der Wortpräsentation wurde ein Bild auf dem Monitor gezeigt, dessen Bezeichnung entweder ebenfalls das gesuchte Phonem enthielt (*match* Bedingung), oder aber nicht (*mismatch* Bedingung). Die Vorhersage war, dass, wenn es sublexikalische Verbindungen in der Richtung von Sprachproduktion zu Spracherkennung gibt, sich in der match Bedingung im Vergleich zur mismatch Bedingung ein Effekt der Voraktivierung zeigen sollte.

Zwei Targetphoneme wurden getestet, /p/ und /k/, die entweder Teil existierender Worte waren (Experimente 1, 2, 4 und 5), oder aber Teil von Pseudowörtern (Experiment 3). Die

Bilder wurden zu zwei verschiedenen SOAs präsentiert, 300 oder 600 Millisekunden vor dem auditiven Target. Basierend auf dem Wissen über die Zeitspanne des phonologischen Enkodierens (siehe Levelt et al., 1991) wurde ein Effekt für SOA 600 vorhergesagt, nicht jedoch für SOA 300. Um die Versuchspersonen zu ermutigen die Bilder anzusehen und deren Namen zu enkodieren anstatt nur passiv auf den Monitor zu schauen, wurde eine zusätzliche Kontrollaufgaben gestellt.

Es wurden an der SOA 600 sowohl Erleichterungseffekte gefunden in Fällen, in denen das Targetphonem im Kontext existierender Wörter war (Experimente 1 und 2), als auch im Kontext von Pseudowörtern (Experiment 3). Der Erleichterungseffekt war zahlenmässig (jedoch nicht statistisch signifikant) kleiner im Falle von Pseudowörtern, was einen geringen lexikalischen Einfluss suggeriert. Es wurden keine Erleichterungseffekte gefunden wenn nur die Mehrzahl der phonologischen Merkmale anstelle eines kompletten Phonems im Bildnamen enthalten war (Experiment 4), oder aber wenn keine Kontrollaufgabe bezüglich der Bilder vorhanden war (Experiment 5).

Es wurde geschlussfolgert, dass es dynamische Verbindungen gibt, die die sublexikalischen Ebenen der zwei Modalitäten in der Richtung von Sprachproduktion zu Spracherkennung verbinden. Um diese Verbindungen zu errichten ist die Absicht von Nöten, eine phonologische Wortrepräsentation zu konstruieren (wie es durch die Kontrollaufgabe der Fall war). Wenn dies nicht der Fall ist, wie im Fall der passiven Betrachtung der Bilder (Experiment 5), scheinen diese Verbindungen abwesend zu sein. Die Aktivierungsübertragung über diese Verbindungen führte zu einer Voraktivierung der entsprechenden Phoneme im Spracherkennungssystem, was eine schnellere Erkennung dieser Phoneme in der Überwachungsaufgabe zur Folge hatte. Die Ergebnisse von Experiment 4 deuten darauf hin, dass die Kommunikation der beiden Systeme auf der Phonemebene stattfindet und nicht auf einer noch tieferen Ebene, wie der der phonologischen Merkmale. Diese Ergebnisse bestärken die Annahmen von WEAVER++, in dem phonologische Wortrepräsentationen als Einheit der Kommunikation zwischen Sprachproduktion und Spracherkennung dienen. Mit diesem Experimentdesign war es nicht möglich, die Existenz von Verbindungen auf lexikalischer Formebene in der Richtung von Sprachproduktion zu Spracherkennung zu untersuchen.

Auch die Ergebnisse dieser Experimentserie können vom Arbeitsmodell erklärt werden, ohne jedoch eine klare Aussage hinsichtlich der Teilung oder Trennung von Repräsentationseinheiten für Sprachproduktion und Spracherkennung zuzulassen. Ein Modell mit geteilten Repräsentationen kann auch diese Ergebnisse erklären. Weiterhin macht nicht einmal die Annahme getrennter Repräsentationen die angenommenen Verbindungen zwichen den Modalitäten notwendig. Die Ergebnisse dieser Experimentserie können auch durch eine direkte Verbindung von Phonemrepräsentationen im Sprachproduktionssystem zur Phonemüberwachungseinheit (z.B. den Decision Nodes, siehe Figure 1.1) erklärt werden. Die Voraktivierung würde dann direkt dort stattfinden, anstelle der Voraktivierung im Spracherkennungssystem. Dies würde ebenfalls zu einer Erleichterung der Erkennung führen, indem die Decision Nodes weniger Input vom Spracherkennungssystem bräuchten, um die Antwortschwelle zu erreichen. Auch wenn diese Möglichkeit zusätzliche Verarbeitungsstrukturen nötig macht und dadurch weniger ökonomisch ist, ist es eine mögliche Alternative.

Die dritte Experimentserie (Kapitel 4) sollte zwischen den zwei möglichen Quellen des Effektes der zweiten Experimentserie (eine Extraverbindung vom Sprachproduktionssystem zu den Decision Nodes versus einer Überwachung durch Verbindungen zum Spracherkennungssystem) entscheiden. Diese Fragestellung wurde anhand der Vorhersagen von zwei unterschiedlichen Ansätzen zur Selbstüberwachung von interner Sprache erörtert. Sogenannte produktionsbasierte Ansätze nehmen an, dass die Überwachung eines internen Sprachplans auf der Basis sprachproduktionsinterner Repräsentationen geschieht (was eine direkte Verbindung von Sprachproduktionsrepräsentationen zur Überwachungseinheit wiederspiegelt). Die Perceptual-Loop Theorie auf der anderen Seite nimmt an, dass die Selbstüberwachung mithilfe des Spracherkennungssystems geschieht und sich somit der Spracherkennungsstrukturen bedient. Der entscheidende Punkt ist, dass die Perceptual-Loop Theorie Effekte während der Selbstüberwachung vorhersagt, die spezifisch für Spracherkennungsprozesse sind, was produktionsbasierte Ansätze nicht tun. Der Uniqueness Point Effekt ist ein solch spezifischer Effekt und wurde als Testfall gewählt, um zwischen den beiden Selbstüberwachungsansätzen zu unterscheiden. Es sollte geklärt werden, ob die in Kapitel 3 gefundenen Effekte aufgrund von Verbindungen zwischen dem Sprachproduktions- und dem Spracherkennungssystem entstanden sind, oder aber aufgrund einer direkten Verbindung von Sprachproduktionseinheiten zu den Decision Nodes.

Zwei Aufgaben wurden zu einem Bilderset gestellt, Bildbenennung und interne Phonemüberwachung. In der Bildbenennungsaufgabe wurde kein Uniqueness-Point Effekt erwartet, da dies ein spezifischer Effekt der seriellen Verarbeitung von Sprache während deren Erkennung ist. Für die interne Phonemüberwachung jedoch sagt die Perceputal-Loop Theorie einen Uniqueness Point Effekt vorher. Der innere Sprachplan wird laut dieser Theorie in das Spracherkennungssystem übertragen und dort wie externe Sprache verarbeitet, was einen Effekt der Einzigartigkeit und Erkennung zu einem bestimmten Zeitpunkt möglich macht. Produktionsbasierte Ansätze andererseits sagen einen solchen Effekt nicht voraus, da sie auf sprachproduktionsinternen Repräsentationen und Prozessen basieren, in denen Phoneme parallel aktiviert werden und Einzigartigkeit bei Erreichen der phonologischen

Verarbeitungsebenen immer gegeben ist. Es ist wichtig zu erwähnen, dass beide Aufgaben, Bildbenennung und interne Phonemüberwachung, auf dem gleichen Aktivierungsfluss basieren. Es gab zu keiner Zeit während des Experimentes ein externes Sprachsignal. Aus diesem Grund können unterschiedliche Effekte in den beiden Aufgaben nur aufgrund einer Involvierung verschiedener Repräsentationen (Sprachproduktion versus Spracherkennung) und Prozesse (parallele versus serielle Aktivierung) erklärt werden.

Das Targetphonem war immer in finaler Position des Wortes und variierte bezüglich seiner Entfernung zum Uniqueness Point des Wortes. Es gab drei Bedingungen, *no-distance* (Targetphonem ist der Uniqueness Point des Wortes), *small-distance* (Targetphonem ist ein Phonem hinter dem Uniqueness Point des Wortes), und *large-distance* (Targetphonem ist zwei Phoneme hinter dem Uniqueness Point des Wortes). Wenn das Spracherkennungssystem in die interne Phonemüberwachung involviert ist, wie von der Perceptual-Loop Theorie angenommen, dann sollten die Erkennungszeiten des Targetphonems kürzer sein, um so weiter dieses vom Uniqueness Point entfernt ist (wie es in Spracherkennungsstudien der Fall ist). Wenn die interne Selbstüberwachung allein auf sprachproduktionsinternen Repräsentationen beruht, sollte dieser Effekt nicht auftreten.

In der Ergebnissen fand sich kein Uniqueness Point Effekt in der Bildbenennungsaufgabe, wie es von beiden Selbstüberwachungsansätzen vorhergesagt wird. Es gab jedoch einen Effekt der Entfernung vom Uniqueness Point in der internen Phonemüberwachungsaufgabe. Dieses Ergebnismuster ist kompatibel mit der Perceptual-Loop Theorie und fordert sprachproduktionsbasierte Ansätze heraus. Die Ergebnisse suggerieren, dass der interne Sprachplan zur Überwachung und Kontrolle ins Spracherkennungssystem übertragen wird. Dies führt zu Effekten in internen Selbstüberwachungsaufgaben, die spezifisch für Spracherkennungsprozesse sind (wie der Uniqueness Point Effekt).

Interaktive Modelle haben, obwohl sie in der Lage sind die Ergebnismuster der ersten beiden Experimentserien (Kapitel 2 und 3) zu erklären, Schwierigkeiten mit den Ergebnissen dieses letzten Experimentes (Kapitel 4).

Die in dieser Arbeit präsentierten Ergebnisse haben zur Beantwortung der Ausgangsfragen in folgender Weise beigetragen:

Beeinflussen sich Spracherkennung und Sprachproduktion gegenseitig?

Diese Frage kann mit Gewissheit mit *Ja* beantwortet werden. In jedem Experiment hatte der eingefügte Einfluss (Spracherkennungseinfluss in Kapitel 2 und Sprachproduktionseinfluss in Kapitel 3) einen starken Effekt auf die Aufgabenbewältigung in der jeweils anderen Modalität.

Wenn ja, auf welchen Ebenen der Verarbeitung beeinflussen sie sich?

Bezüglich des Einflusses von Spracherkennung zu Sprachproduktion (Kapitel 2), wurden Verbindungen auf sublexikalischen sowie lexikalischen Formverarbeitungsebenen identifiziert. In der anderen Richtung des Einflusses von Sprachproduktion zu Spracherkennung (Kapitel 3), gab es Evidenz für Verbindungen auf sublexikalischer Ebene. Mit dem Experimentdesign der zweiten Experimentserie konnte die Existenz von Verbindungen auf lexikalischer Formverarbeitungsebene weder bestätigt noch ausgeschlossen werden.

Gibt es getrennte Repräsentationen für Spracherkennung und Sprachproduktion auf der Ebene der Formverarbeitung?

Das letzte Experiment (Kapitel 4) zeigte einen spezifischen Spracherkennungseffekt, den Uniqueness Point Effekt, in einer internen Selbstüberwachungsaufgabe. Das legt nahe, dass Repräsentationen ausserhalb des Sprachproduktionssystems bei der Bewältigung dieser

Aufgabe beteiligt sind, was zu der Annahme führt, dass die sublexikalischen Repräsentationen für die Spracherkennung und die Sprachproduktion getrennt sind. Zusammen gesehen sprechen die drei in dieser Arbeit präsentierten Experimentserien für ein Sprachverarbeitungssystem mit getrennten, jedoch eng verbundenen Formverarbeitungsebenen für Sprachproduktion und Spracherkennung, wie es im Arbeitsmodell (Figure 1.1) illustriert ist.

CURRICULUM VITAE

Rebecca Özdemir was born in 1979 in Duisburg, Germany. She studied Linguistics, Information Science and German at Heinrich-Heine-University, Düsseldorf, Germany. She graduated in 2002 with a masters thesis on the processing of relative clauses in first and second language speakers of English. From 2003 until 2005 she was affiliated with the Utterance Encoding group of the Max-Planck-Insitute for Psycholinguistics, working on her PhD project, reported in this thesis.