INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
RICE UNIVERSITY

SEMANTIC AND PHONOLOGICAL FACTORS IN SPEECH PRODUCTION:
EVIDENCE FROM PICTURE-WORD INTERFERENCE EXPERIMENTS

by

MARKUS FRIEDRICH DAMIAN

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

_______________________________
Randi C. Martin, Professor of Psychology, Chair

_______________________________
Michael J. Watkins, Professor of Psychology

_______________________________
Jeffrey S. Bowers, Assistant Professor of Psychology

_______________________________
Suzanne E. Kemmer, Associate Professor of Linguistics

Houston, Texas

May, 1998
ABSTRACT

SEMANTIC AND PHONOLOGICAL FACTORS IN SPEECH PRODUCTION:
EVIDENCE FROM PICTURE-WORD INTERFERENCE EXPERIMENTS

by

Markus Friedrich Damian

Discrete models of speaking maintain that semantic-syntactic and phonological representations are largely independent, whereas interactive accounts allow for mutual influence between them. The studies presented here investigated this issue by employing a task in which participants named pictures while instructed to ignore visually or auditorily presented distractor words. Previous results using this paradigm with auditory distractors have been used to support the discrete view (e.g., Schriefers, Meyer, & Levelt, 1991) whereas results with visual distractors have been used to argue for an interactive account (e.g., Starreveld & La Heij, 1996a). The first two experiments served to clarify the discrepancy across distractor modalities. Experiment 1 demonstrated that with auditorily presented distractors semantic effects preceded phonological effects whereas with visual distractors phonological effects had an earlier onset than semantic effects. Experiment 2 provided a means for accounting for this discrepancy by demonstrating that the results for visual distractors followed the auditory pattern when presentation time was limited. The following two experiments addressed the issue of interactivity vs. modularity in speaking by employing auditory distractors and investigating the effects of complex
types of distractors. Experiment 3 factorially crossed the factors of semantic and phonological relatedness by employing both semantically and phonologically related distractors (FLY-FLEA). An interaction between the two factors was obtained which was interpreted as supporting an interactive account of speaking. Experiment 4 investigated the effects of mediated distractors which are related to the picture name via an intervening word (TIDE-(TIGER)-LION). A potential effect of such distractors would require an interplay between semantic and phonological levels and thus further support an interactive view. The results showed no effects on naming latencies, a finding that probably is not diagnostic with regard to the question of interactive vs. modular accounts.

The second part of this thesis introduced an interactive computational model closely related to Dell's (1986) model of speech production. This model yielded a sequence of semantic and phonological effects and showed an interaction between the two factors as well as the absence of mediated effects. In summary, the experiments and the model favor an interactive view of speaking in which semantic and phonological levels are closely interconnected.
Acknowledgments

I would like to express my sincere appreciation to all who have helped me during the completion of this dissertation project. I would especially like to thank my thesis advisor, Dr. Randi Martin, for providing continuous feedback and expertise during the various stages involved in this project. I would also like to thank Dr. Jeffrey Bowers, Dr. Suzanne Kemmer, and Dr. Michael Watkins who served as members on the dissertation committee. Further thanks go to Andrew Shirey for assisting me in the collection of the data. Finally, I am especially grateful to Raquel Guirardello, who has provided me with much emotional support, as well as my parents, who have endured my absence for such a long time.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>OVERVIEW</td>
<td>1</td>
</tr>
<tr>
<td>SPEECH PRODUCTION - REPRESENTATIONS AND PROCESSES</td>
<td>2</td>
</tr>
<tr>
<td>Representational Levels in Speech Production</td>
<td>2</td>
</tr>
<tr>
<td>Discrete vs. Interactive Accounts</td>
<td>4</td>
</tr>
<tr>
<td>Complex Patterns of Lexical Relationship</td>
<td>9</td>
</tr>
<tr>
<td>The Picture-Word Interference Procedure</td>
<td>12</td>
</tr>
<tr>
<td>EXPERIMENT 1 - Visual vs. Auditory Distractors</td>
<td>17</td>
</tr>
<tr>
<td>Method</td>
<td>18</td>
</tr>
<tr>
<td>Participants</td>
<td>18</td>
</tr>
<tr>
<td>Materials</td>
<td>18</td>
</tr>
<tr>
<td>Design</td>
<td>18</td>
</tr>
<tr>
<td>Apparatus</td>
<td>19</td>
</tr>
<tr>
<td>Procedure</td>
<td>19</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Results</td>
<td>20</td>
</tr>
<tr>
<td>Discussion</td>
<td>25</td>
</tr>
<tr>
<td>EXPERIMENT 2 - Limited Presentation Duration of Distractors</td>
<td>35</td>
</tr>
<tr>
<td>Method</td>
<td>35</td>
</tr>
<tr>
<td>Participants</td>
<td>35</td>
</tr>
<tr>
<td>Materials</td>
<td>35</td>
</tr>
<tr>
<td>Design, Apparatus and Procedure</td>
<td>35</td>
</tr>
<tr>
<td>Results</td>
<td>36</td>
</tr>
<tr>
<td>Discussion</td>
<td>38</td>
</tr>
<tr>
<td>EXPERIMENT 3 - Effects of Combined Semantic and Phonological Relationship</td>
<td>40</td>
</tr>
<tr>
<td>Method</td>
<td>43</td>
</tr>
<tr>
<td>Participants</td>
<td>43</td>
</tr>
<tr>
<td>Materials</td>
<td>43</td>
</tr>
<tr>
<td>Design</td>
<td>45</td>
</tr>
<tr>
<td>Apparatus</td>
<td>45</td>
</tr>
<tr>
<td>Procedure</td>
<td>45</td>
</tr>
<tr>
<td>Results</td>
<td>46</td>
</tr>
<tr>
<td>Discussion</td>
<td>50</td>
</tr>
<tr>
<td>EXPERIMENT 4 - Mediated Relationships</td>
<td>61</td>
</tr>
<tr>
<td>Method</td>
<td>70</td>
</tr>
<tr>
<td>Participants</td>
<td>70</td>
</tr>
</tbody>
</table>
Appendix A ................................................................. 121
Appendix B ................................................................. 122
Appendix C ................................................................. 123
Appendix D ................................................................. 124
List of Tables

Table 1  Experiment 1: Mean Response Latencies (in ms), varied by Distractor Type (Visual vs. Auditory), Relatedness (Unrelated vs. Semantic vs. Phonological) and Picture-Word Onset Asynchrony (SOA)

Table 2  Experiment 2: Mean Response Latencies (in ms), varied by Relatedness (Unrelated vs. Semantic vs. Phonological) and Picture-Word Onset Asynchrony (SOA)

Table 3  Experiment 3: Mean Response Latencies (in ms), varied by Relatedness (Unrelated vs. Semantic vs. Phonological vs. Semantic & Phonological) and Picture-Word Onset Asynchrony (SOA)

Table 4  Experiment 4: Mean Response Latencies (in ms), varied by Relatedness (Control vs. Unrelated vs. Semantic vs. Phonological vs. Mediated I vs. Mediated II) and Picture-Word Onset Asynchrony (SOA). Mediated I: Distractor is form-related to a category coordinate of the picture; Mediated II: Distractor is a category coordinate of a word form-related to the picture

Table 5  Parameter Settings in the Computational Model
List of Figures

Figure 1 Three varieties of complex priming involving phonological and/or semantic links. Upward arcs indicate semantic relatedness, downward arcs denote phonological links. Adapted from O’Seaghdha and Marin (1997).

Figure 2 Target picture paired with A) unrelated, B) semantically related, C) form-related Distractor Word.

Figure 3 Distractor word tapping into various substages of picture naming preparation depending on picture-distractor onset asynchrony (SOA).

Figure 4 Experiment 1: Effects of semantically and phonologically related distractors (unrelated minus related condition) varied by distractor modality (visual vs. auditory) and SOA (200 ms, -100 ms, 0 ms, +100 ms, +200 ms).

Figure 5 Processing of target pictures and distractors with fixed (A) or unlimited (B) presentation duration.

Figure 6 Experiment 2: Effects of semantically and phonologically related distractors (unrelated minus related condition) varied by SOA (200 ms, -100 ms, 0 ms, +100 ms, +200 ms).

Figure 7 Experiment 3: Effects of semantically related, phonologically related, and semantically and phonologically related distractors (unrelated minus related condition) varied by SOA (-150 ms, 0 ms, +150 ms).

Figure 8 Discrete Two-step model of speaking. From Roelofs, Meyer, and Levelt (1996), p. 247.

Figure 9 Experiment 3: Effects of semantically related, phonologically related, and mediated distractors (unrelated minus related condition) varied by SOA (-150 ms, 0 ms, +150 ms). Mediated I: Distractor is form-related to a category coordinate of the picture; Mediated II: Distractor is a category coordinate of a word form-related to the picture.

Figure 10 Simple network implementing a lexical layer (T = target, R = node related to target, U = node unrelated to target) as well as a conceptual layer in which features representing T and R overlap.
Figure 11 Absolute selection criterion: Target activation in units as a function of time step and relation to simultaneously activated distractor node (unrelated vs. related).

Figure 12 Choice rule: Target selection probability as a function of time steps and relation to simultaneously activated distractor node (unrelated vs. semantically related).

Figure 13 Lateral inhibition: Target Activation in units as a function of time steps and relation to simultaneously activated distractor node (unrelated vs. semantically related).

Figure 14 Simple network implementing a lexical layer ($T = \text{target, } R = \text{node related to target, } U = \text{node unrelated to target}$) as well as a conceptual layer in which the concepts corresponding to $T$ and $R$ are linked via a similarity connection.

Figure 15 Model in Figure 14 Target selection probability as a function of time steps and relation to simultaneously activated distractor node (unrelated vs. semantically related).

Figure 16 Interactive model of speech production. $P = \text{phonologically related to target, } S = \text{semantically related, } U = \text{unrelated, } S&P = \text{semantically and phonologically related, } M1 = \text{mediated; phonologically related to a category coordinate, } M2 = \text{mediated; category coordinate of a phonologically related item.}$

Figure 17 Activation of lexical and phonological target nodes as a function of time step. Onset occurs at time step $= 0$.

Figure 18 Activation of distractor lexical node in the unrelated condition.

Figure 19 Time course of semantic and phonological effects (Unrelated minus Related condition).

Figure 20 Effect (unrelated minus related) of semantically and phonologically related distractors in comparison to the effects of semantic distractors and phonological distractors.
Figure 21  Effect (unrelated minus related) of mediated distractors in comparison to the effects of semantic distractors and phonological distractors. Mediated I: Distractor is form-related to a category coordinate of the picture; Mediated II: Distractor is a category coordinate of a word form-related to the picture.

Figure 22  Activation of lexical distractor nodes under various corresponding conditions with an SOA of 0 time steps.
Semantic and Phonological Factors in Speech Production:
Evidence from Picture-Word Interference Experiments

by

Markus Friedrich Damian

OVERVIEW

This research investigated the relationship between semantic and phonological representations involved in the production of single words. Models of speech production generally assume that speaking proceeds from the retrieval of conceptual codes to the access of corresponding word forms. "Discrete" models of speaking (e.g., Garrett, 1975; Levelt, 1989; Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991) propose an independence among the involved levels, whereas "interactive" accounts of speech production (e.g., Dell, 1986; Stemberger, 1985) allow bidirectional flow of activation between semantic and phonological levels. To date, the bulk of evidence with regard to this issue stems from patterns of naturally occurring speech errors, while experimental studies are rare. In the work reported here, this topic was investigated with a timed picture naming task.

The literature review begins with a description of the shared characteristics of all models of speaking, and then outlines in detail the discrete and the interactive account of speech production. Evidence from speech error corpora taken to argue for or against each position is summarized. In the following section, "complex" relationships between lexical items are described which can serve as a test to investigate predictions derived from either account. Several experiments are reported which made use of a picture-word
interference procedure in which participants named picture while presented with
distractor words that were phonologically related, semantically related, or unrelated to the
picture name. These experiments examined critical assumptions of the discrete account
of speaking, and evidence compatible with an interactive account was obtained. In
addition to these experimental studies, a computational model was employed which tests
whether the findings obtained in the experiments can be accounted for within an
interactive framework. The general discussion summarizes the findings and discusses
implications for the architecture of speech production.

SPEECH PRODUCTION - REPRESENTATIONS AND PROCESSES

Representational Levels in Speech Production

Researchers interested in speech production widely agree on the assumption that
lexical access in language production requires a distinction of at least two levels of
representation: a first level that is concerned with semantic and syntactic properties, and
a second level specifying phonological characteristics. In Levelt's (1989) model of
speaking, for instance, preparation of an utterance consists of a sequence of stages which
is based on a prelinguistic conceptualization, or message, of the planned utterance. First,
appropriate lexical entries are retrieved that specify language-specific semantic (e.g.,
conceptual arguments) and syntactic (e.g., word class and grammatical gender) properties.
Following Kempen and Huijbers (1983), the term lemmas is often used for these
representations. Subsequently, corresponding word forms are encoded. These
representations contain sound representations, including segmental and metric
information, and are commonly termed *lexemes*. Finally, a phonetic plan is formed and executed in the articulation process. Because the lexicon is separated into two representational levels (lemma and lexeme access), such a model is commonly named a “two-step” account of speaking. Although a great deal of disagreement exists as to the exact characteristics of these stages, the necessity to hypothesize two levels of lexical access in speaking is relatively undisputed.

Evidence for the separation of lexical retrieval into a semantic-syntactic and a phonological component comes from various sources. For example, when speakers are in a "tip-of-the-tongue" state, they appear to have access to semantic (Brown & McNeill, 1966) properties of the target word, but are unable to retrieve the appropriate word form or gain only partial phonological access. Similar, but much more dramatic, problems can be shown in amnestic patients (e.g., Goodglass, Kaplan, Weintraub, & Ackerman, 1976). These patients often produce a multitude of semantic circumlocutions: they generate descriptions of the semantic properties of the word to be retrieved, for example saying "that's a good eating bird" when trying to retrieve "turkey" (Martin, Lesch, & Bartha, 1998). These circumlocutions suggest normal or near-normal retrieval of semantic information with some difficulties in retrieving phonological information. Likewise, a number of studies have demonstrated access to syntactic information in either amnestic patients or the tip-of-the-tongue state (Badecker, Miozzo, & Zanuttini, 1995; Caramazza & Miozzo, 1997; Vigliocco, Antonini, & Garrett, 1997; Vigliocco, Vinson, Martin, & Garrett, 1998). All these cases suggest a blockage between the semantic-syntactic and the phonological level and thus support the notion of two distinct stages in lexical retrieval.
A further source of evidence comes from speech errors from normal speakers. For instance, Garrett (1975) suggested a distinction between word and sound exchanges, with each type possessing distinct characteristics: word exchanges normally take place between syntactic phrases or even clauses, preserve their mutual word class, and fulfill similar grammatical functions. In contrast, sound exchange errors mostly affect adjacent words which are often from dissimilar syntactic categories, but typically share form-related characteristics. These findings are taken to indicate that word exchanges result from the selection of the wrong lemma, whereas phonological errors reveal the malfunctioning of the phonological encoding system. A final source of evidence for the distinction between semantic-syntactic and phonological stages in speaking comes from psycholinguistic studies on the time course of speech production in normal speakers (Levelt et al., 1991; Schriefers, Meyer, & Levelt, 1990) which have suggested an early stage of lexical-semantic activation, followed by a later stage of phonological encoding.

Discrete vs. Interactive Accounts

On a very general level, representational stages of a cognitive model can be conceptualized as “discrete”, “cascaded”, or “interactive”. In a discrete information processing model, subprocesses are identified as successive temporal stages, each of which occupies a particular time interval. A cognitive process is then portrayed as the succession of these discrete components. Such a characterization is largely congruent with Fodor’s (1983) notion of cognitive modularity, according to which cognitive modules are automatic, fast-acting, and informationally encapsulated. In contrast,
“cascaded” information processing models (McClelland, 1979) do not require that processing within one component must be completed before a second one can start. In contrast, a continuous operation is hypothesized in which processing can take place on more than one representational level simultaneously. In its original formulation, the cascaded model type conceptualized activation as flowing through the system in an exclusively feedforward fashion. In contrast, the interactive activation framework (McClelland & Rumelhart, 1981) assumes cascaded processing, but additionally hypothesizes that later processing stages can affect earlier ones. Such models therefore implement feedforward as well as feedback links and are therefore termed “interactive”.

The general distinction between discrete, cascaded, and interactive models of information processing is mirrored in issues pertaining to speech production. "Discrete two-step models" (Garrett, 1975; Levelt, 1989; Levelt et al., 1991) assume that speaking proceeds in serial stages from lemma to lexeme retrieval. For instance, Levelt, Roelofs, and Meyer (in press) argue on theoretical grounds that semantic and phonological representational systems are so fundamentally different from each other that they should be conceptualized as distinct and discrete. They assume that before infants show the first signs of real speech, they develop cognitively in two separate ways: on the one hand, infants acquire basic conceptual notions of agency, interactancy, object permanence, etc., and the developing conceptual structures are eventually labeled with words acquired from the natural environment. On the other hand, they also acquire a set of syllabic articulatory gestures, or “babbles”. The acquisition of speech production skills commences at the point when the infant connects lexical concepts to articulatory gestures. Speaking,
therefore, is seen as the coupling of two cognitive domains - lexical concepts and articulatory patterns - that were initially separate from each other, and the distinction between these two systems is perpetuated in the mature speech production system.

Within the discrete two-stage framework of speaking, specific claims have been forwarded as to how lexical access proceeds from semantic-syntactic retrieval to phonological encoding. The first claim asserts that phonological encoding is restricted to one single lexical item (Levett et al., 1991, p.124). The occurrence of semantically conditioned speech errors is usually interpreted as suggesting that a conceptual message activates a cohort of semantically related lexical items (Bock & Levett, 1994). However, only the single lemma that is eventually selected from among the activated cohort propagates its activation to the *phonological* level. Processing in this way shows discrete characteristics since phonological encoding cannot commence until semantic-syntactic retrieval has been entirely accomplished. The second assumption forwarded by proponents of the discrete account of speaking holds that there is no feedback from the lexeme onto the lemma level (Levett et al., 1991, p.125). Since processing is conceptualized as a succession of discrete stages, the - structurally and probably temporally - later stage of phonological encoding is not permitted to influence the earlier stage of semantic-syntactic lexical retrieval. In combination, discrete models of speaking conceive of semantic and phonological codes as largely independent and modular, and apart from the fact that the first stage provides the input for the second, there is no interactivity in between them.
There are a number of findings in the literature that appear to be incompatible with the discrete two-step view, however. Most prominently, corpora of naturally occurring speech errors appear to contain an above-chance ratio of speech errors that are both semantically and phonologically related to the target word, or so-called "mixed errors" (e.g., Dell & Reich, 1981). Also, recent neuropsychological evidence suggests that some aphasics with a high occurrence of word substitutions produce mixed errors at above-chance levels (Best, 1996; Blanken, 1990). The discrete two-step theory of speaking proposes strict independence between the involved representational levels. Accordingly, both semantic-syntactic and phonological components should contribute independently to the production of errors, and mixed errors should be found at no more than chance rate. The above-chance occurrence of this type of error appears to be incompatible with discrete models unless additional assumptions are introduced. For instance, it has been argued that the speaker internally monitors his/her own speech by means of a "lexical monitoring" system (Baars, Motley, & MacKay, 1975; Levelt, 1989) which filters out all but occasional speech errors before they are overtly produced. Such a system can account for the occurrence of mixed errors by claiming that these error words resemble the target word twofold, namely in both semantic and phonological respects. Consequently, such errors are more likely to go undetected than words that are either semantically or phonologically related to the target, resulting in a higher occurrence than would be predicted by pure chance. Within this hypothesis, the mechanism that accounts for mixed errors is not a component of the speech production system per se, but rather an external process.
Alternatively, however, mixed errors might arise as a natural consequence of the structure of the production system itself. Interactive models (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1996; Stemberger, 1985) abandon the notion of discrete, modular stages and the assumptions made within this framework and instead propose semantic, lexical, and phonological levels that are interconnected in a network fashion. As outlined above, the assumptions are derived from the discrete view that a) phonological encoding is restricted to the selected semantic-syntactic item, and b) phonological encoding has no influence on semantic-syntactic retrieval. In contrast, interactive models dispose with these assumptions and, by virtue of their interactive nature, propose the parallel phonological encoding of multiple items as well as feedback from the phonological back to the semantic-syntactic level. That is, an entire set of lemma nodes that has been activated on the basis of the preverbal message propagates some degree of activation to the word form level. Furthermore, lexeme activation feeds back to the lemma level (but see Caramazza and Miozzo, 1997, and Humphreys, Riddoch, and Quinlan, 1988, for models that are cascaded, but not fully interactive and thus discard the first, but not the second claim of the discrete view). Models that implement feedback of activation account for the occurrence of mixed errors by assuming that words semantically as well as phonologically related to the target word receive additional activation from their semantic overlap with the target word as well as from their phonological resemblance to the correct word. As a result, “mixed” lexical units possess a higher selection probability than would be predicted by either type of relationship alone. Therefore, interactive models account for mixed speech errors within
the production system itself and without proposing external mechanisms like a lexical monitoring system.

**Complex Patterns of Lexical Relationship**

Although the issue of modularity vs. interactivity in speaking was first addressed with regard to speech errors, the relationship between semantic and phonological codes can also be investigated by means of response time studies. Traditionally, such studies compare conditions in which two or more stimuli are either semantically or form-related to each other with a condition in which the two stimuli are in no obvious relationship to each other. For instance, in priming studies, a probe word or picture is preceded by either an unrelated or a related stimulus. Likewise, in the picture-word interference procedure, subjects name pictures while simultaneously being presented with related or unrelated distractor words. In both cases, the effects of a relationship between prime/distractor and probe item on response latencies are measured based on the assumption that such effects reflect processing at the corresponding cognitive stages. Apart from obvious relationships such as semantic or phonological relatedness, there are other, less direct and more complex, types of relationship that can be subject to empirical investigation. These are schematized in Figure 1. Such complex patterns of priming have been predominantly addressed in research on speech comprehension. However, an obvious parallel can be drawn to speech production, as will be outlined below.

The majority of studies employing such complex effects has addressed the issue of indirect types of relationships exclusively at the semantic level (see A) in Figure 1).
Figure 1. Three varieties of complex priming involving phonological and/or semantic links. Upward arcs indicate semantic relatedness, downward arcs denote phonological links. Adapted from O’Seaghdha and Marin (1997).

Effects of semantic-semantic mediation have been demonstrated in a number of studies in the literature (Balota & Lorch, 1986; McNamara, 1992a, 1992b; McNamara & Altarriba, 1988) and are usually attributed to automatic spreading activation taking place in the conceptual system, although it has recently been argued (McKoon & Ratcliff, 1992;
Ratcliff & McKoon, 1994) that such effects might better be explained in terms of a weak direct link between prime and target than by means of a relationship mediated by an intervening item. Furthermore, although effects of semantic-semantic mediation are interesting because they test the limits of the spreading activation metaphor, they are not relevant concerning whether lexical access can be characterized as interactive or discrete: the mediated relationship resides exclusively in the conceptual domain and is therefore not diagnostic with regard to the relation of semantic and phonological codes. For these reasons, semantically mediated effects will not be further regarded in the present discussion.

The other two types of relationship outlined in Figure 1 are more relevant for the issue of how semantic and phonological codes relate to each other. Type B) in Figure 1 represents a case in which prime/distractor and target word are both semantically and phonologically related to each other, and thus represents the direct analog to the occurrence of "mixed" speech errors described in the preceding section. Just as in the case of speech errors, cognitive theories that propose the strict independence of semantic and phonological levels predict an effect of these items that derives from the independent contribution of either type of relationship alone. In contrast, an effect for these items that differs from the one that would be predicted from either type of relationship alone would appear to favor non-modular, interactive accounts of speech production. This logic has been applied to speech production in a number of studies (Rayner & Springer, 1986; Starreveld & La Heij, 1995,1996a) and will be further explored in Experiment 3.
The third type of a complex relationship between a prime/distractor and a target word, outlined in Panel C, consists of a word that is semantically related to a lexical item form-related to the target word, or vice versa. For these items to have an effect on each other, an interplay between semantic and phonological codes is required which is permitted in interactive, but not in modular, accounts. This argument has also been applied to speech production, albeit not very extensively (Leveld et al., 1991; Jescheniak & Schriefers, 1997), and will be investigated in Experiment 4.

The Picture-Word Interference Procedure

One of the tools by means of which the nature of the two stages of lexical access in speaking can be experimentally investigated is the picture-word interference procedure. This task will be extensively employed in the following experiments and is a variation of Stroop’s (1935) seminal study. In the Stroop task, participants are instructed to name the color of a word which itself is a different color name (e.g., the color of the word “red” printed in blue ink has to be named). The fact that the word itself carries - incompatible - lexical information about color massively interferes with naming the color of the word, compared to a condition in which the word itself is not a color name. The enormous amount of research inspired by this finding is comprehensively reviewed in MacLeod (1991). Although the Stroop effect still awaits a satisfactory explanation, most theorists would agree that it is in part due to the fact that the word information undergoes automatic lexical processing which in turn interferes with the retrieval of the correct color name.
Figure 2. Target picture paired with A) unrelated, B) semantically related, C) form-related distractor word.

The picture-word interference procedure is schematized in Figure 2. Color naming has been replaced with a task in which participants perform timed naming responses to pictures depicting basic objects, henceforth referred to as *targets*. At the same time, they are being confronted with words that are visually embedded in the object. The effect of these so-called *distractor* words, which participants are instructed to ignore, on naming latencies for the pictures is the measure of interest. In parallel to the Stroop task, distractor words interfere with the picture naming response and slow down latencies (hence the term "interference procedure"). Furthermore, distractors can bear various relationships to the picture naming. Words that are semantically related to the picture have been shown to slow naming responses compared to unrelated words (e.g., Picture: APPLE, Distractor: ORANGE), a finding first reported by Rosinski, Golinkoff, and Kukish (1975). On the other hand, form-related words speed up naming latencies relative to unrelated words (Picture: BELL, Distractor: BELT), as shown by, e.g., Posnansky and Rayner (1977, 1978) and Rayner and Posnansky (1978).
Studies that involve the picture-word interference procedure arose within the domain of research on the Stroop task, and consequently, the theoretical conclusions drawn from these studies cannot easily be transferred to issues relevant to speech production. Traditionally, a speech response has been thought of as an "output variable" which by itself was not considered deserving further investigation (Bock, 1996). For instance, within the Stroop domain, the interference effect is commonly attributed to "response competition" (e.g., Morton & Chambers, 1973), but it is not immediately clear how such a concept relates to the phases of semantic-syntactic and phonological retrieval in picture naming. More recently, the picture-word interference task has been adopted to more explicitly address issues pertaining to speech production. Here, there exists evidence that the semantic interference effect in this task is not merely a conceptual, preverbal conflict, but crucially involves the stage of semantic-syntactic lexical retrieval (Schriefers et al., 1990; Damian, Bowers, & Katz, 1997). In contrast, the phonological facilitation effect appears to reside at the stage of phonological encoding (Schriefers et al., 1990; Starreveld & La Heij, 1996a). Consequently, semantic and phonological effects can be used to investigate issues pertaining to their corresponding representational levels involved in speaking.

There are at least two advantages that the picture-word interference procedure has over the original Stroop task. First, whereas the response set in the Stroop task is restricted to a small number of easily identifiable colors, the use of pictures allows the investigation of a large number of semantic domains. Second, in the Stroop task, the color to be named and the identity of the word are properties of one and the same
stimulus, whereas picture and distractor in the picture-word interference task are separate stimuli. Thus, the onset of picture and distractor can be systematically varied. The underlying idea is schematized in Figure 3: if it is assumed that distractor words have a constant processing time, then the manipulation of stimulus-onset asynchrony (SOA) between picture and distractor will tap into one of various stages involved in the preparation of the naming response.

In recent years, a number of studies have manipulated SOAs in the picture-word interference procedure. In these studies, the convention has been adopted to denote SOAs as relative to the picture onset. That is, a condition in which distractor onset precedes picture onset is labeled as a negative SOA, conditions in which distractor onset follows picture onset are labeled as a positive SOA, and a condition in which distractor and picture onset are simultaneous are labeled as a SOA of zero. For instance, Glaser and Düngelhoff (1984) demonstrated a semantic interference effect that centered around an SOA of 0 ms, and Starreveld and La Heij (1996a) additionally showed that phonological

![Diagram](image)

**Figure 3.** Distractor word tapping into various substages of picture naming preparation depending on picture-distractor onset asynchrony (SOA).
facilitation effects can be obtained with SOAs ranging from -200 ms to +100 ms.

A further variation of the picture-word interference task has been introduced by Schriefers et al. (1990). Because orthographic and phonological characteristics are highly confounded in most languages, there is some doubt as to the exact locus and origin of the facilitation caused by form-related visual distractor words in the studies described above. In order to ensure that phonologically related distractor words immediately tapped into phonological output representations involved in speaking, Schriefers et al. presented distractor words auditorily instead of visually. A first experiment demonstrated that, similar to a task in which distractor words are presented visually, auditory distractors significantly interfered with the naming response. In a second and critical experiment, semantic interference, but no phonological effect, was obtained at an SOA of -150 ms, whereas phonological facilitation, but no semantic effect, was obtained at SOA = 0 ms and SOA = +150 ms. This pattern was interpreted as reflecting the separate and serial contributions of lemma retrieval and lexeme access during the preparation of the naming response.

If the picture-word interference procedure is to be used as a vehicle in the investigation of speech production, it must be demonstrated that all relevant experimental variables are sufficiently understood. The following two experiments provide a clarification of a crucial aspect of the picture-word interference procedure that has not been sufficiently addressed in previous research, namely, the time course of semantic and phonological effects with visually or auditorily presented distractors.
EXPERIMENT 1 - VISUAL VS. AUDITORY DISTRACTORS

A close survey of the findings from the picture-word interference paradigm demonstrates a striking discrepancy between visual and auditory distractors with regard to the pattern of semantic and phonological effects. Starreveld and La Heij's (1996a) study with visually presented distractors yielded phonological facilitation effects that preceded and followed semantic interference effects in time. In contrast, Schriefers et al.'s study with auditory distractors demonstrated a sequence of semantic interference followed by phonological facilitation. However, both versions of the procedure have been taken as supporting particular claims or constraining hypotheses regarding speech production. For instance, Schriefers et al. interpreted the sequence of semantic and phonological effects obtained with auditory distractors as in line with discrete two-step accounts of speech production. In contrast, Starreveld and La Heij obtained an interaction between semantic and phonological factors with visual distractors and concluded, based on Sternberg's (1969) additive-factors method, that a discrete model cannot be adequate (the specific details of this experiment will be outlined in Experiment 3). For obvious reasons, either conclusion is premature as long as the discrepancy between the effects of visual and auditory distractors across modalities has not been satisfactorily explained. However, to date, there has been little attempt to uncover the source of this disparity.

A further difficulty in interpreting these findings arises from the fact that the effects of visual and auditory distractors have been investigated by different groups of researchers and in different experiments. That is, it is not entirely clear whether the discrepant findings reflect some yet unknown stimulus or procedural property, or a
genuinely interesting fact that a model of speech production must account for. In order to clarify this issue, the following experiment directly compared the effects of visually and auditorily presented distractors.

Method

Participants. Thirty Rice University undergraduate students served as participants in order to meet a class requirement for experimental participation.

Materials. Twenty-eight line drawings of common objects were selected from the Snodgrass and Vanderwart (1980) set. For each picture, a categorically related word was selected that served as the semantically related distractor. An orthographically/phonologically related word was chosen that shared a minimum of the initial two phonemes as well as the initial two letters with the picture label. Finally, an unrelated word was selected that stood in no obvious relationship to the target picture. This yielded a total of 84 picture-distractor combinations (28 pictures x 3 conditions). All three distractor conditions were matched in terms of the average number of letters and phonemes.

Design. The experimental design included Distractor Modality (Visual vs. Auditory) as a between-subjects factor, as well as two within-subjects factors, namely SOA with five levels (-200 ms, -100 ms, 0 ms, +100 ms, and +200 ms), and Target-Distractor Relation with three levels (Unrelated, Semantically Related, Phonologically Related). Fifteen participants were randomly assigned to each Distractor Modality condition. Each participant saw the set of 84 picture-word pairs at all five SOAs (-200 ms, -100 ms, 0 ms, +100 ms, and +200 ms). The order in which participants received the
SOA blocks was balanced according to a Latin Square design. Items were presented in a pseudorandom fashion such that the same picture never appeared twice on subsequent trials.

**Apparatus.** The pictures and distractors were presented from an Apple PowerPC computer using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). The pictures were digitized at a size of approximately 3 x 3 in and presented as black line drawings on white background. The auditory distractors were recorded by a female speaker, digitized with a sampling frequency of 20 kHz, and presented via Labtec LT-100 headphones. The visual distractors were shown in uppercase Helvetica 18 Bold font in the center of each picture. The computer scheduled presentation of the stimuli and recorded a participant's response times to the nearest millisecond by means of a voice-activated relay.

**Procedure.** Participants were tested individually. At the beginning of the experiment, they were familiarized with the set of experimental pictures by viewing each picture on the computer screen for 2000 ms with the appropriate name printed below it. A practice block was then administered in which participants performed a naming response to each of the pictures presented in random order. Responses other than the expected ones were corrected by the experimenter. Next, a second practice block was administered in which the pictures were presented accompanied by unrelated distractor words. Again, unexpected responses were corrected. Finally, the five experimental blocks were carried out, consisting of 84 naming responses each. Breaks were provided between the experimental blocks. Each testing session thus consisted of 420 experimental trials and lasted approximately one hour.
On each individual trial, participants first viewed a fixation cross presented at the center of the screen for 1000 ms. After a blank interval of 500 ms, the picture appeared. At varying intervals before or after the picture onset (depending on the SOA block), the distractor was played over the headphones (in the auditory distractor group) or presented at the center of the picture (in the visual distractor group). Participants performed the naming response by speaking into a microphone; the picture disappeared from the screen as soon as the voice key triggered. Visually presented distractor words were terminated simultaneously with the picture offset. Following each naming response, the experimenter judged the response to be either correct or incorrect (which included picture names other than the expected ones, repairs, stuttering or mouth clicks, or malfunctioning of the voice key) by typing a code into the computer. Each trial was followed by a 1500 ms intertrial interval.

Results

All responses judged to be incorrect by the experimenter were excluded from the analysis. Likewise, response times longer than 2000 ms or shorter than 250 ms were eliminated. Finally, response times deviating more than three standard deviations from a participant's conditional mean were deleted. These procedures resulted in the exclusion of 2.0%, 0.0%, and 1.4% of the data, respectively. Thus, a total of 3.4% of data points were deleted in this way.

Table 1 lists the mean response latencies, varied by Distractor Modality, Target-Distractor Relation, and SOA. Figure 4 displays the semantic and phonological effects (Unrelated minus Related condition). With visual distractors, semantically related
Table 1

Experiment 1: Mean Response Latencies (in ms), varied by Distractor Type (Visual vs. Auditory), Relatedness (Unrelated vs. Semantic vs. Phonological) and Picture-Word Onset Asynchrony (SOA)

<table>
<thead>
<tr>
<th>SOA (in ms)</th>
<th>-200</th>
<th>-100</th>
<th>0</th>
<th>+100</th>
<th>+200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Distractors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrelated</td>
<td>700</td>
<td>714</td>
<td>744</td>
<td>693</td>
<td>620</td>
</tr>
<tr>
<td>Semantic</td>
<td>698</td>
<td>725</td>
<td>786</td>
<td>716</td>
<td>640</td>
</tr>
<tr>
<td>Phonological</td>
<td>666</td>
<td>687</td>
<td>693</td>
<td>643</td>
<td>608</td>
</tr>
<tr>
<td>Auditory Distractors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unrelated</td>
<td>695</td>
<td>706</td>
<td>707</td>
<td>637</td>
<td>605</td>
</tr>
<tr>
<td>Semantic</td>
<td>723</td>
<td>730</td>
<td>734</td>
<td>641</td>
<td>606</td>
</tr>
<tr>
<td>Phonological</td>
<td>690</td>
<td>683</td>
<td>671</td>
<td>591</td>
<td>581</td>
</tr>
</tbody>
</table>

distractors cause an interference effect that peaks at an SOA of 0 ms. Critically, phonological/orthographic effects precede the semantic effect, as indicated by a substantial amount of facilitation at SOA = -200 ms and -100 ms. In contrast, auditory
Figure 4. Experiment 1: Effects of semantically and phonologically related distractors (unrelated minus related condition) varied by distractor modality (visual vs. auditory) and SOA (200 ms, -100 ms, 0 ms, +100 ms, +200 ms).
distractors yield semantic interference effects ranging from SOA = -200 ms to SOA = 0 ms, as well as phonological facilitation effects ranging from SOA = -100 ms to SOA = +200 ms.

A 2 (Distractor Modality) x 5 (SOA) x 3 (Condition) ANOVA was conducted on the data with Distractor Modality as a between-subjects factor, and SOA and Condition as within-subjects factors. The effect of Distractor Modality was insignificant in the participants analysis, $F_1 = 1.32$, and marginally significant in the items analysis, $F_2(1,54) = 3.74, p = .059, MSE = 109,309$, indicating that visual and auditory distractors showed roughly comparable overall amounts of interference. A main effect of SOA, $F_1(4,112) = 63.55, p = .000, MSE = 188,318, F_2(4,216) = 435.30, p = .000, MSE = 348,568$, as well as main effect of Condition, $F_1(2,56) = 80.84, p = .000, MSE = 91,225, F_2(2,108) = 82.40, p = .000, MSE = 179,574$, were obtained. Furthermore, a significant SOA x Condition interaction was obtained, $F_1(8,224) = 7.93, p = .000, MSE = 3,590, F_2(8,432) = 9.15, p = .000, MSE = 6,590$, as well as a significant SOA x Distractor Modality interaction, $F_1(4,112) = 6.68, p = .000, MSE = 19,785, F_2(4,216) = 45.33, p = .000, MSE = 36,300$. Critically, a significant three-way SOA x Condition x Distractor Modality interaction was found, $F_1(8,224) = 2.80, p = .006, MSE = 1,267, F_2(8,432) = 3.55, p = .001, MSE = 2,559$, indicating that the way in which SOA and Condition modified each other varied depending on whether distractors were presented visually or auditorily.
Two further ANOVAs were conducted in which the data were analyzed for each Distractor Modality condition separately. For visual distractors, a main effect of SOA, $F_1(4,56) = 24.17, p = .000, \text{MSE} = 84,358$, $F_2(4,108) = 203.43, p = .000, \text{MSE} =$ 155,239, a main effect of Condition, $F_1(2,28) = 55.92, p = .000, \text{MSE} = 55,381$, $F_2(2,54) = 50.72, p = .000, \text{MSE} = 112,585$, and a SOA x Condition interaction, $F_1(8,112) = 7.14, p = .000, \text{MSE} = 3,252$, $F_2(8,216) = 7.89, p = .000, \text{MSE} = 6,237$, were obtained.

Newman-Keuls pairwise comparisons that compared Unrelated to either the Semantic or the Phonological condition within each SOA group indicated that at both SOA = -200 ms and SOA = -100 ms, the Phonological condition differed significantly from both the Unrelated and the Semantic condition, which did not differ significantly from each other. At SOA = 0 ms and SOA = +100 ms, all three conditions differed significantly from each other. Finally, at SOA = +200 ms, the Semantic condition differed significantly from both the Unrelated and the Phonological condition, which did not differ from each other.

For auditory distractors, a main effect of SOA, $F_1(4,108) = 50.80, p = .000, \text{MSE} =$ 229,629, $F_2(4,108) = 273.90, p = .000, \text{MSE} = 229,629$, a main effect of Condition, $F_1(2,54) = 29.06, p = .000, \text{MSE} = 69,684$, $F_2(2,54) = 32.58, p = .000, \text{MSE} = 69,684$, and a SOA x Condition interaction, $F_1(8,216) = 3.57, p = .000, \text{MSE} = 2,913$, $F_2(8,216) = 4.48, p = .000, \text{MSE} = 2,913$, were obtained. Newman-Keuls tests indicated that at SOA = -200 ms, the Semantic condition differed significantly from both the Unrelated and the Phonological condition, which did not differ from each other. For SOA = -100 ms and SOA = 0 ms, all three conditions differed significantly from each other. Finally,
at SOA = +100 ms and SOA = +200 ms, the Phonological condition differed significantly
from both the Unrelated and the Semantic condition, which did not differ significantly
from each other.

Discussion

The results replicated the discrepancy between the effects of auditory and visual
distractors previously reported in the literature. With visual distractors, semantic
interference appeared in a time window ranging from SOA = 0 ms to SOA = +200 ms,
whereas the phonological/graphemic facilitation effect spanned a range of SOAs from -
200 ms to +100 ms. These findings are quite similar to those from Starreveld and La
Heij’s (1996a) study. In contrast, auditorily presented distractors caused a semantic
interference effect at the earlier SOAs (-200 ms, -100 ms, and 0 ms) and a phonological
facilitation effect at later SOAs (-100 ms, 0 ms, +100 ms, and +200 ms). Such a pattern
roughly resembles the sequence of semantic and phonological effects obtained in
Schriefers et al.’s (1991) original study. Note, however, that their study showed a sharp
distinction between SOAs revealing semantic and those revealing phonological effects,
whereas the present experiment demonstrated a considerable degree of overlap at
intermediate SOAs, and exclusively semantic or phonological effects only at more
peripheral SOAs.¹

Of course, SOAs are not directly comparable across distractor modalities.
Visually presented words are probably processed in parallel and can be assumed to gain
quite rapid access to their semantic codes. In contrast, when a word is auditorily
presented, a certain time span will be required until its recognition point is reached and
semantic effects can arise. In the picture-word interference procedure, a longer processing duration of the distractor word translates into a larger negative SOA at which an effect can be obtained, and accordingly auditory distractors should show semantic effects at earlier SOAs than visual ones. Roughly, this is the pattern that was found: auditorily presented distractors showed semantic interference at SOA = -200 ms, SOA = -100 ms, and SOA = 0 ms, whereas visually presented distractors showed similar effects at SOA = 0 ms, SOA = +100 ms, and SOA = +200 ms. When phonological facilitation effects are analyzed not in terms of their absolute SOAs, but relative to the onset and offset of the corresponding semantic effects, the discrepancy regarding distractor modality becomes even more apparent: with auditory distractors, the onset of semantic interference preceded the onset of phonological facilitation by two SOA steps, whereas with visual distractors, the onset of semantic interference followed the onset of phonological facilitation by two SOA steps.

The above findings suggest that distractor modality is crucial in regard to the pattern of effects that will be obtained in the picture-word interference procedure. Two possible explanations for such a modality effect have been forwarded in earlier discussions in the literature. The first hypothesis discounts the sequence of semantic and phonological effects obtained within the auditory version of the picture-word interference procedure as an artifact that bears no consequence on the question of how semantic and phonological representational stages in speaking relate to each other. The second hypothesis rejects the visual version of the picture-word interference procedure as inconclusive regarding how speech production works and implies that visual and auditory
distractors differ in their activation of form-related competitors. These two arguments will now separately be described, and it is concluded that neither one provides a satisfactory explanation of the different patterns across modality.

According to the first hypothesis suggested by Dell and O'Seaghdha (1991) and Starreveld and La Heij (1996b), the apparent sequence of semantic and phonological effects obtained with auditorily presented distractors is an artifact resulting from the extension of auditory stimuli over time. Phonologically related distractor words yield a facilitation effect at the later SOAs because at these SOAs only the initial segments of the words have been processed when the phonological form of the target is accessed. These segments are congruent with the target, whereas the incongruent segments of the distractor word have not yet been reached. As a result, a facilitation effect arises that is functionally like an “identity priming” condition, i.e., one in which picture label and distractor are the same word. In contrast, at the earlier SOAs, the whole distractor word is available to the participant at the time of phonological retrieval of the target, and its incongruence with the target name is clearly detected. As a result, phonological facilitation arising from partial form overlap is counteracted by the recognition of the incongruence between distractor and target, and response times in the form-related condition do not differ from those in the unrelated one. For semantically related distractors, the opposite relation holds between the distractor effect and partial and full processing of the distractor: in order to interfere with lemma retrieval of the target, the identity of the distractor has to be fully recognized, which is the case only at earlier, but not at later, SOAs. The prediction derived from these assumptions is a pattern of early
semantic interference, followed by later phonological facilitation, just as was obtained by Schriefers et al. and the above experiment. However, the “identity priming” hypothesis attributes this pattern not to a processing sequence of lemma and lexeme retrieval, but merely to the fact that auditorily presented distractor words possess an extension over time. Consequently, the obtained pattern of effects should not be taken as evidence for the discrete two-step notion of lexical retrieval in speech production.

The “identity priming” hypothesis has been shown to be implausible in a study by Meyer and Schriefers (1991). This study utilized phonologically related distractor words that were either word-initially related (“klos” -> chock) or end-related (“mok” -> mug) to the target name (“klok” -> clock). For end-related targets, the whole word would be perceived by the time the overlapping phonological segments were processed. Consequently, there should be no phonological facilitation for these items if the identity priming hypothesis holds. However, a similar facilitation effect was obtained for both conditions when the timing of the stimuli was manipulated such that the overlapping phonological segments occurred at the same point in time in relation to picture onset. This finding implies that the sequential character of auditory word recognition is irrelevant in regard to the facilitation effect, and consequently, the claim that the pattern of semantic and phonological effects in this task is just an artifact resulting from the temporal extension of auditory stimuli is not valid. In contrast, Meyer and Schriefers' findings lend further support to the assumption that such a sequence reflects the respective contributions of lemma and lexeme processing in speech production.
The second hypothesis which attempts to account for the difference in the effects of visual and auditory distractors by discounting the visual effects will be called the "input priming" hypothesis. As briefly described in the introduction, Starreveld and La Heij (1995, 1996a) experimentally crossed semantic and phonological relatedness in a picture naming task with visually presented distractors. The obtained interaction was interpreted as contradicting the predictions derived from "discrete" models of speech production. In a reply, Roelofs, Meyer, and Levelt (1996) suggested that form-related effects obtained with visual distractors might have nothing to do with the phonological stage - or lexeme retrieval - of picture naming. Instead, they might originate in the process of input processing of the distractor word. That is, visually presented distractor words might activate form-related lexical neighbors in the visual word form lexicon (e.g., "cap" would activate "cat" to some extent). If the target lemma that is to be named ("cat") is among the activated set, retrieval time might be reduced. However, such a form-related facilitation effect from visually presented distractors should not be interpreted as reflecting processes residing at the lexeme activation stage of speaking. Instead, it would be hypothesized to arise from form-related lemma activation due to purely visual processes, or due to input phonological processes. In either case, form-related facilitation effects would not be diagnostic in regard to phonological encoding in speech production. Furthermore, both phonological priming effects and semantic interference effects would affect the lemma retrieval stage, and as a consequence, an interaction between them, as demonstrated by Starreveld and La Heij (1996a), would bear
no consequence as to whether discrete or interactive models of speaking are more appropriate.

Indeed, the hypothesis that lexical co-activation effects might account for form-related priming obtained with visual distractors is in accordance with certain models of visual word recognition (e.g., McClelland & Rumelhart, 1981). Furthermore, Lupker (1982) attempted to disentangle the respective contributions of orthography and phonology to priming effects in picture naming. He demonstrated substantial interference from distractor words that were orthographically identical to the picture name except their first letter, but their single vowel sounds differed (Picture: FOOT, Distractor: BOOT). Reduced, but still significant, interference was obtained from distractor words that were phonologically similar, but visually dissimilar to the target word (Picture: PLANE, Distractor: BRAIN). The findings were taken to suggest that both visual and phonological similarity significantly contribute to the net facilitation effect obtained from visually presented distractors. Consequently, Roelofs et al.'s argument regarding the possibility of a purely visual effect receives some support from this study. Further evidence supporting their claim comes from outside the scope of the picture-word interference task. A number of experiments have found form-related priming effects with visually presented prime and target words (Grainger & Segui, 1992; Lukatela & Turvey, 1990; Meyer, Schvaneveldt, & Ruddy, 1974; Segui & Grainger, 1990), suggesting that prime words indeed co-activate form-related lexical items. In addition, a number of studies have investigated the processing of single target words while manipulating competitor frequency or set size. Such competitor effects, first investigated by
Coltheart, Davelaar, Jonassen, and Besner (1977), have been investigated in a variety of tasks, such as perceptual identification (Grainger & Jacobs, 1996; Grainger & Segui, 1990; Snodgrass & Mintzer, 1993), lexical decision (Andrews, 1989, 1992; Forster & Shen, 1996; Grainger & Jacobs, 1996; Johnson & Pugh, 1994), word naming (Andrews, 1989, 1992; Grainger, 1990; Sears, Hino, & Lupker, 1995), and semantic categorization (Forster & Shen, 1996). Although these studies show a complex pattern of findings, they nevertheless strongly suggest the existence of neighborhood effects in visual word recognition. Consequently, this body of research appears to support Roelofs et al.'s claim that at least part of the phonological facilitation effect obtained with visually presented distractors in picture naming might be attributable to visual effects during input processing.

A problematic aspect of this hypothesis is that in the picture-word interference procedure, participants process the distractor word only incidentally. It is an open question whether neighborhood effects arising during distractor processing are powerful enough to influence target retrieval, as claimed by Roelofs et al. Furthermore, interactive models of word recognition, such as McClelland and Rumelhart's (1981) model, do indeed suggest co-activation effects, but these effects are so transient that it is debatable whether they could have a detectable impact in the picture-word interference paradigm. A further difficulty with the "visual priming" hypothesis is that it would seem to imply that neighborhood effects should be observed with auditory presentation of the distractor as well. That is, the phonological input should activate phonologically similar words and prime target retrieval if target and distractor are form-related. Indeed, lexical co-
activation effects have been demonstrated by a number of studies investigating form-related priming in the auditory domain (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Radeau, Morais, & Dewier, 1989; Slowiaczek & Pisoni, 1986; Slowiaczek, Nusbaum, & Pisoni, 1987). Although investigations of lexical co-activation effects in the auditory domain are quite scarce (but see Marslen-Wilson, 1990), such effects are also predicted by current models of auditory speech processing (e.g., Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986). As a result, it can be assumed that neighborhood effects should arise during processing of auditory distractors as well. However, Schriefers et al.’s experiment and our Experiment 1 demonstrated that auditorily presented distractors yielded no phonological priming at negative SOAs. Phonological facilitation only emerged at a point in time subsequent to the onset of semantic effects. Unless one is willing to assume that visual and auditory word recognition differ fundamentally from each other (i.e., visually presented words activate lexical neighbors, whereas auditorily presented words do not), this discrepancy suggests that neighborhood effects - disregarding the modality in which they reside - are irrelevant in regard to how form-related priming in the picture-word interference task comes about.

In summary, the “visual input priming” hypothesis receives support from some experimental studies suggesting lexical co-activation effects in visual word recognition. On the other hand, this hypothesis can account for the discrepancy between visual and auditory distractors only if it assumes a fundamental difference in the way words are processed in the visual and auditory domain. Evidence for such an assumption is weak.
There is yet another, much simpler explanation for the discrepancy between visually and auditorily presented distractors, however. In their use of the picture-word interference procedure, Schriefers et al. (1990) and others have manipulated the SOA between picture and word in order to tap into the various representational stages involved in speech preparation. In order for this manipulation to achieve this aim, it must be assumed that distractor words have a roughly fixed access time (see Panel A of Figure 5) and, depending on their onset relative to the picture, either interfere with semantic retrieval of the target word (at earlier SOAs) or facilitate phonological encoding (at later SOAs). The assumption of a fixed and limited access time seems appropriate for auditorily presented distractors which have a fixed extension over time and thus terminate "naturally" with their offset. Although one might assume that the phonological and semantic representations remain activated for some time following the end of the word, the duration of this activation might be expected to be relatively brief for unattended words.

In contrast, visually presented distractors do not have a fixed duration, and as a consequence, the interval for which they are presented is arbitrary. In Experiment 1, visually presented distractors - together with the picture - remained on the screen until the participant triggered the voice key. In particular, visual distractors in the SOA = -200 ms and SOA = -100 ms conditions appeared on the screen before the picture was presented, but then remained on throughout presentation of the picture. In Starreveld and La Heij's (1996a) study, both pictures and distractor words remained on the screen for 375 ms after picture onset. According to Levelt et al.'s (in press) estimates, both lemma and lexeme
Figure 5. Processing of target pictures and distractors with fixed (A) or unlimited (B) presentation duration.

activation roughly take place within the first 400 ms of naming preparation. This implies that in Experiment 1 as well as Starreveld and La Heij’s study, distractor words were present on the screen while both target lemmas and lexemes were accessed, independent of the SOA of their onset (see Panel B of Figure 5). Both experiments revealed phonological facilitation effects at early SOAs, which crucially distinguished the pattern from the one obtained with auditory distractor words. I hypothesize that such “early” form-related effects arise because, due to their unlimited presentation duration, these distractors nevertheless tap into a “late” stage of speech production (i.e., lexeme activation). That is, it is assumed that the visual distractors unintentionally activate their lexical representations while they remain on the screen.
An experimental test of this hypothesis is relatively straightforward. The only procedural detail that needs to be altered is the presentation duration of the distractor words. A complete match between auditory and visual distractors in this regard is of course not possible: Semantic access in auditory word recognition critically depends on when "lexical uniqueness" is achieved, whereas visually presented words probably gain relatively faster access to semantic codes. However, the issue at question is whether visually presented distractors, if their presentation duration is limited, yield a pattern of effects comparable to the one obtained with auditory distractors. The following experiment used the stimuli from Experiment 1 and cleared distractors from the screen after a preset interval, even though the target pictures remained on the screen until a response was made.

EXPERIMENT 2 - LIMITED PRESENTATION DURATION OF DISTRACTORS

Method

Participants. Twenty Rice University undergraduate students served as participants in order to meet a class requirement for experimental participation.

Materials. The employed stimuli were identical to those used in Experiment 1.

Design, Apparatus and Procedure. Only distractors in the visual modality were used. The conditions were the same as in Experiment 1. Thus, the experimental design included SOA with five levels (-200 ms, -100 ms, 0 ms, +100 ms, and +200 ms) and
Target-Distractor Relation with three levels (Unrelated, Semantically Related, and Phonologically Related) as within-subjects factors.

The procedure was identical to the one employed in Experiment 1, with one exception. Within each trial, the distractor word was displayed for only 200 ms and then replaced by a string of "X" letters that served as a mask and stayed on the screen for 500 ms. All other temporal parameters remained identical to Experiment 1. In particular, the picture remained on the screen until the voice key triggered.

The same apparatus as the one used in Experiment 1 was used.

Results

As in Experiment 1, all responses judged to be incorrect by the experimenter, response times shorter than 250 ms or longer than 2000 ms, and response times deviating more than three standard deviations from a participant's conditional mean were excluded. These procedures resulted in the exclusion of 3.0%, 0.6%, and 1.4% of the data, respectively. A total of 5.0% of data points were deleted in this way.

Table 2 and Figure 6 display the results. Substantial semantic interference effects can be found for SOAs ranging from -100 ms to +100 ms. In contrast, facilitation effects from form-related distractors appear to set in at SOA = 0 ms and stretch into the positive SOAs. Critically, no form-related effect is found at SOA = -200 ms, and the effect at SOA = -100 ms is very small.

An ANOVA conducted on the data showed a significant main effect of SOA, $F_1(4,76) = 7.40, p = .000, MSE = 36,233$, $F_2(4,108) = 46.36, p = .000, MSE = 50,479$, a main effect of Condition, $F_1(2,38) = 39.81, p = .000, MSE = 64,240$, $F_2(2,54) = 48.48$,
Table 2
Experiment 2: Mean Response Latencies (in ms), varied by Relatedness (Unrelated vs. Semantic vs. Phonological) and Picture-Word Onset Asynchrony (SOA)

<table>
<thead>
<tr>
<th>SOA (in ms)</th>
<th>-200</th>
<th>-100</th>
<th>0</th>
<th>+100</th>
<th>+200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated</td>
<td>714</td>
<td>710</td>
<td>738</td>
<td>741</td>
<td>689</td>
</tr>
<tr>
<td>Semantic</td>
<td>731</td>
<td>735</td>
<td>782</td>
<td>774</td>
<td>673</td>
</tr>
<tr>
<td>Phonological</td>
<td>715</td>
<td>704</td>
<td>695</td>
<td>671</td>
<td>660</td>
</tr>
</tbody>
</table>

Figure 6. Experiment 2: Effects of semantically and phonologically related distractors (unrelated minus related condition) varied by SOA (200 ms, -100 ms, 0 ms, +100 ms, +200 ms).
\[ p = .000, \text{MSE} = 94,386, \text{and a significant SOA} \times \text{Condition interaction, } F_1(8,152) = 10.57, p = .000, \text{MSE} = 10,277, F_2(8,216) = 20.13, p = .000, \text{MSE} = 14,872. \]

Newman-Keuls pairwise comparisons were performed within each SOA group. At SOA = -200 ms, none of the conditions differed significantly from each other. At SOA = -100 ms, the Semantic condition differed significantly from both the Unrelated and the Phonological condition, which did not differ from each other. At SOA = 0 ms and SOA = +100 ms, all three conditions differed significantly from each other. Finally, at SOA = +200 ms, the Phonological condition differed significantly from both the Unrelated and the Semantic condition, which did not differ from each other.

**Discussion**

Similar to Experiment 1, the semantic interference effect is most pronounced at SOA = 0 ms and present to a lesser extent at the neighboring SOAs. In contrast, phonological facilitation is now found exclusively at the later SOAs ranging from 0 ms to +200 ms, whereas, in Experiment 1, such an effect was found at SOAs ranging from -200 ms to +100 ms. In other words, Experiment 2 yielded a sequence of semantic interference, followed by phonological facilitation. Such a pattern bears a high resemblance to the pattern obtained with auditory distractors, but differs substantially from the one obtained with the unlimited visual distractors in Experiment 1. I conclude that it is neither the "identity priming" nor the "input priming" hypothesis detailed above that accounts for the discrepancy in the pattern obtained with visual and auditory distractors. Rather, an overlooked procedural detail - the presentation duration of the distractor - caused the difference in how semantic and phonological effects temporally
relate to each other. When the presentation duration of the distractor was limited, a similar pattern - semantic interference followed by phonological facilitation - was obtained with visual and auditory distractors.

This finding implies that reasonable trust can now invested in Schriefers et al.'s suggestion that the sequence of semantic and phonological effects obtained in their experiment reflects the underlying stages of lexical access in speaking. The fact that even visually presented distractors displayed such a pattern demonstrates that the temporal extension of auditorily presented words cannot be the cause for the sequence of effects found with auditory distractors. Thus, potential differences in how visually and auditorily presented distractor words are processed are irrelevant with regard to the picture-word interference procedure. In this way, our findings are in agreement with Meyer and Schriefers' (1991) results demonstrating that not only word-initially related, but also end-related phonological distractors can facilitate responses. Both theirs and our findings underscore the assumption that the sequence of semantic and phonological effects constitutes a basic fact about lexical access in speaking that models of language production have to account for.

The implications in regard to the "input priming" hypothesis forwarded by Roelofs et al. are less clear. Note that these authors tailored their argument as a rebuttal to Starreveld and La Heij's conclusions, but did not explicitly outline their claims in regard to the difference between visually and auditorily presented distractors. At the very least, our findings suggest that lexical co-activation during visual word processing cannot be the only difference between the visual and auditory patterns. Such co-activation
effects should be most pronounced at early stages of word processing, which correspond to a late SOA in the picture-word interference procedure. However, the difference between visual and auditory distractors consisted in the presence or absence of phonological effects preceding the onset of semantic effects, and this difference was eliminated when presentation duration of visual distractors was limited. This finding would appear to imply that such cohort effects are irrelevant for the phonological facilitation to arise. I conclude that with auditory distractors, phonological facilitation effects most likely reflect the exclusive effects of the distractor on word form encoding for production.

EXPERIMENT 3 - EFFECTS OF COMBINED SEMANTIC AND PHONOLOGICAL RELATIONSHIP

The following experiment constitutes a test of the predictions derived from the discrete model of speaking by investigating the effects of distractor words that are both semantically and phonologically related to the picture name. This issue has been recently addressed in two reports by Starreveld and La Heij (1995, 1996a) whose study was based on Rayner and Springer's (1986) observation that semantic and form-related effects appear to have a non-additive effect in the picture-word interference procedure. However, Rayner and Springer presented their conditions in a blocked fashion, which might have given rise to potential anticipation effects. Starreveld and La Heij (1995) randomly mixed the conditions and, in accordance with Rayner and Springer, replicated the interaction between the two factors at an SOA of 0 ms. Furthermore, Starreveld and
La Heij (1996a) demonstrated this interaction between semantic and orthographic similarity for a wide range of SOAs. Based on Sternberg’s (1969) additive-factors method, the authors interpreted the obtained statistical interaction between semantic and orthographic factors as indicating one of two possibilities: either the process of lemma retrieval does not exist or is irrelevant in speech production, or if it exists, lemma retrieval must receive feedback from the word form level. The former account, favored by the authors, was termed the “name retrieval” account. Starreveld and La Heij concluded that, in order to explain their findings, a separate lexical-semantic stage is not necessary; all that is required is a word form level separate from the conceptual system. Retrieval of the correct lexical item, in this model, is a one-step process residing at the word form level that is subject to semantic interference (for instance via spread of activation in the semantic system, see, e.g., Roelofs, 1992a, 1992b, for a possible mechanism) as well as orthographic/phonological facilitation (resulting from the priming of word form information). As both the experimental factors of semantic and phonological relatedness presumably affect a single representational level, such an account would indeed predict the obtained interaction between the two factors.

However, Schriefers et al.’s results, as well as the findings from Experiment 2, establish a sequential pattern of semantic and phonological processing as a basic characteristic of speech production. The suggested two separate stages of lexical access are difficult to reconcile with Starreveld and La Heij’s “name retrieval” account, which assumes that both semantic and form-related factors affect a single representational level. In contrast, a two-step account of speaking is compatible with such a sequence. Of
course, this preliminary conclusion does not speak to the question of whether such a two-step account must necessarily be conceived of as modular or "discrete". Indeed, Harley (1993) demonstrated that an interactive model is able to produce a rough sequence of early semantic and late phonological activation. Although the computational details of this model are open for debate, his report indicates that the sequential pattern should not be interpreted as necessarily ruling out non-modular accounts. Consequently, the present findings should be interpreted as corroborating a two-step account of speech production, irrespective of whether modular or interactive versions are favored.

This opens the possibility that the logic of Starreveld and La Heij's experiment could be employed in order to test the assumptions made by the "discrete" two-step model of speaking. Auditorily presented distractors or visual distractors with limited presentation duration yield a sequence of semantic and phonological effects. Assuming that this pattern reflects the respective contributions of lemma and lexeme representations, the additive-factors method could test whether these two stages are additive or interact. In this regard, the discrete two-step model of speech production allows straightforward predictions: the presence of semantic interference in the absence of a phonological effect at a negative SOA would be taken as indicating that the speech production process is at the lemma retrieval stage, and phonological properties of the distractor word are entirely irrelevant at this point. Consequently, semantically related distractors that are also phonologically related should yield the same semantic interference effect as semantically related, phonologically unrelated distractors. The reverse prediction would be made for positive SOAs: the presence of phonological
facilitation in the absence of semantic interference presumably reflects the completion of semantic-syntactic retrieval and the access of word form information. As a result, distractors both semantically and phonologically related should yield facilitation effects comparable to those obtained from phonologically related, semantically unrelated words. In other words, the discrete two-step model clearly predicts additivity of the respective effects of semantic and phonological similarity.

If instead an interaction between the two factors is obtained, such a finding would clearly pose problems for the basic assumptions of the two-step model. If additive-factor logic holds, then either a) the two factors affect the same representational level, or b) there is some interactivity between the two levels. Given that Experiments 1 and 2 suggest the seriality of representational stages, possibility a) can be discounted. Thus, an interactive relationship would argue in favor of non-modular accounts of speech production.

In order to investigate this issue, the following experiment employed an analog of Starreveld and La Heij's (1996a) experiment by factorially crossing semantic and phonological relatedness while using auditory distractors.

**Method**

**Participants.** Forty-two Rice University undergraduate students served as participants in order to meet a class requirement for experimental participation.

**Materials.** Eighteen line drawings of common objects were selected from the Snodgrass and Vanderwart (1980) set which had a semantic coordinate that also shared a minimum of the initial two phonemes with the picture name (henceforth S&P distractors).
In the following, semantically (S) and phonologically (P) related distractors were selected that were matched in terms of their semantic and phonological overlap to the S&P distractors. As suggested, but not implemented by Starreveld and La Heij (1996a), it is important to match distractor words in the S&P condition to those in the S and the P condition on an individual item basis rather than just by averaging scores for each condition. This is important in regard to an items analysis, which will otherwise show deviant results. Again, phonological similarity was defined as an overlap of the two initial phonemes as a minimal requirement. In addition, a phonological overlap score was calculated that matched P and S&P distractor words in regard to their phonological similarity to the picture label.² For each individual picture label, the phonological overlap scores of the corresponding P and S&P distractors deviated no more than 10% from each other. Likewise, it was ensured that the unrelated (U) distractors were as phonologically unrelated to the picture label as the S distractors. In regard to conceptual overlap, semantic relatedness ratings were collected for the stimuli. Thirty participants were presented with pairs of words and instructed to rate conceptual similarity on a scale from 1 (very unrelated) to 5 (very related). The rated pairs consisted of the eighteen picture labels paired with the S&P distractor words as well as each picture label paired with three different potential S distractors, plus the picture label pairs with the P and the U distractors, presented in random order. From these ratings, the one S distractor word was chosen for each individual picture whose rating score was closest to the corresponding S&P score. Again, none of the individual scores for the S&P and the S condition deviated more than 10% from each other. Also, it was ensured that P and U
distractors were semantically unrelated to comparable degrees. This selection procedure ensured that both phonological and semantic similarity were matched on an individual item basis. This experiment also employed a Control condition in which no distractor word was displayed.

**Design.** The experimental design included SOA (-150 ms, 0 ms, and +150 ms), Semantic Relatedness (Related vs. Unrelated), and Phonological Relatedness (Related vs. Unrelated) as within-subjects factors. Each set of 90 picture-word pairs was displayed three times at different SOAs. The order in which participants received the SOA blocks was balanced according to a Latin Square design. Items were presented in a pseudorandom fashion such that the same picture never appeared twice on subsequent trials.

**Apparatus.** The same apparatus as the one in the above experiments was used.

**Procedure.** The general procedure was similar to the one employed in Experiment 1 and 2. However, in accordance with Schriefers et al.'s (1990) Experiment 2, only three SOA blocks (-150 ms, 0 ms, +150 ms) were administered. Again, each participant was familiarized with the pictures before the experiment began. Subsequently, two practice blocks of 18 trials each, one without and one with auditorily presented unrelated distractors, were conducted. Finally, the three experimental blocks of 90 trials each were administered, resulting in a total of 270 experimental trials for each participant. Each testing session lasted approximately half an hour.
Results

Responses judged to be incorrect by the experimenter, response times shorter than 250 ms or longer than 2000 ms, and response times deviating more than three standard deviations from a participant's conditional mean were eliminated. These procedures resulted in the exclusion of 2.3%, 0.2%, and 1.2% of the data, respectively. A total of 3.7% of data points were deleted in this way.

Table 3 displays the mean response times, varied by SOA and picture-distractor relation. Figure 7 displays the semantic, phonological and semantic as well as phonological effects (Unrelated minus Related condition). Parallel to the half of Experiment 1 that employed auditory distractors, an early stage of semantic interference in the absence of phonological effects is followed by a late stage of exclusively phonological facilitation, with an intervening stage in which both effects are obtained. Semantically as well as phonologically related distractors have an effect that appears roughly additive only at the latest SOA, but clearly not so at the earlier two SOAs.

A 3 (SOA) x 5 (Type of Relatedness) ANOVA was conducted on the data with both factors as within-participants factors. A main effect of SOA, \( F_1(2,82) = 6.61, p = .002, \text{MSE} = 30,305, \; F_2(2,34) = 26.46, p = .000, \text{MSE} = 13,584 \), as well as main effect of Type of Relatedness, \( F_1(4,164) = 54.75, p = .000, \text{MSE} = 95,787, \; F_2(4,68) = 20.22, p = .000, \text{MSE} = 42,301 \), was obtained. Furthermore, a significant SOA x Type of Relatedness interaction was obtained, \( F_1(8,328) = 5.67, p = .000, \text{MSE} = 5,120, \; F_2(8,136) = 4.85, p = .000, \text{MSE} = 2,145 \).
Table 3
Experiment 3: Mean Response Latencies (in ms), varied by Relatedness (Unrelated vs. Semantic vs. Phonological vs. Semantic & Phonological) and Picture-Word Onset Asynchrony (SOA)

<table>
<thead>
<tr>
<th>SOA (in ms)</th>
<th>0</th>
<th>+150</th>
</tr>
</thead>
<tbody>
<tr>
<td>-150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>630</td>
<td>643</td>
</tr>
<tr>
<td>Unrelated</td>
<td>670</td>
<td>702</td>
</tr>
<tr>
<td>Semantic</td>
<td>702</td>
<td>733</td>
</tr>
<tr>
<td>Phonological</td>
<td>665</td>
<td>683</td>
</tr>
<tr>
<td>Semantic &amp; Phonological</td>
<td>679</td>
<td>676</td>
</tr>
</tbody>
</table>

Due to the fact that the factors Semantic and Phonological Relatedness were factorially crossed, an additional 3 (SOA) x 2 (Semantic Relatedness) x 2 (Phonological Relatedness) ANOVA was performed in which the response data from the Control condition were omitted and all three factors were treated as within-subjects variables. A main effect of SOA was obtained, $F_1(2,82) = 7.90, p = .001, \text{MSE} = 31,752, F_2(2,34) = 23.86, p = .000, \text{MSE} = 14,131$. Furthermore, a main effect of Semantic Relatedness was obtained, $F_1(1,41) = 39.10, p = .000, \text{MSE} = 23,772, F_2(1,17) = 12.53, p = .003, \text{MSE} = 10,417$, as well as a main effect of Phonological Relatedness, $F_1(1,41) = 79.10, p = .000,$
Figure 7. Experiment 3: Effects of semantically related, phonologically related, and semantically and phonologically related distractors (unrelated minus related condition) varied by SOA (-150 ms, 0 ms, +150 ms).

$\text{MSE} = 128,853$, $F_2(1,17) = 13.62, p = .002$, $\text{MSE} = 59,888$. The SOA x Semantic Relatedness interaction approached significance, $F_1 = 2.93, p = .059$, $F_2 = 2.48, p = .099$. The SOA x Phonological Relatedness interaction was significant, $F_1(2,82) = 12.83, p = .000$, $\text{MSE} = 10,581$, $F_2(2,34) = 9.64, p = .001$, $\text{MSE} = 4,252$. The Semantic Relatedness x Phonological Relatedness interaction was significant in the subjects analysis, $F_1(1,41) = 10.58, p = .002$, $\text{MSE} = 11,756$, and approached significance in the items analysis, $F_2 = 3.19, p = .092$. Critically, a significant SOA x Semantic Relatedness x Phonological Relatedness interaction was obtained, $F_1(2,82) = 5.10, p = .008$, $\text{MSE} = 4,308$, $F_2(2,34) = 3.87, p = .031$, $\text{MSE} = 1,728$. 
The significant three-way interaction described above indicates that the extent to which semantic and phonological relatedness modify each others' effects varies according to SOA. Consequently, the effects of Semantic and Phonological Relatedness were investigated at each level of the factor SOA separately. For SOA = -150 ms, this analysis showed a main effect of Semantic Relatedness, $F_1(1,41) = 27.92$, $p = .000$, $MSE = 20,783$, $F_2(1,17) = 19.25$, $p = .000$, $MSE = 10,012$, and a main effect of Phonological Relatedness, $F_1(1,41) = 11.95$, $p = .001$, $MSE = 8,208$, that approached significance in the items analysis, $F_2(1,17) = 3.47$, $p = .080$, $MSE = 4,534$. Furthermore, the Semantic x Phonological Relatedness interaction was significant in the participants analysis, $F_1(1,41) = 6.60$, $p = .014$, $MSE = 4,059$, but not significant in the items analysis, $F_2 = 2.41$.

Newman-Keuls pairwise comparisons indicated that the Control condition as well as the Semantically Related condition differed significantly from all other conditions. In contrast, the Phonologically Related, the Unrelated, and the Semantically & Phonologically Related Condition did not differ from each other, but differed significantly from either the Control or the Semantically Related Condition.

For SOA = 0 ms, the analysis showed a main effect of Semantic Relatedness in the subjects analysis, $F_1(1,41) = 5.03$, $p = .030$, $MSE = 5,406$, that was not significant in the items analysis, $F_2 = 1.90$. Furthermore, a main effect of Phonological Relatedness was obtained, $F_1(1,41) = 35.13$, $p = .000$, $MSE = 60,611$, $F_2(1,17) = 11.22$, $p = .004$, $MSE = 27,549$, as well as a significant Semantic x Phonological Relatedness interaction, $F_1(1,41) = 10.63$, $p = .002$, $MSE = 16,301$, $F_2(1,17) = 8.29$, $p = .014$, $MSE = 7,565$. 
Newman-Keuls tests showed that the Control condition as well as the Unrelated and the Semantically Related condition differed from all other conditions. In contrast, the Phonologically Related and the Semantically & Phonologically Related condition did not differ from each other, but differed from all other conditions.

Finally, for SOA = +150 ms, a main effect of Semantic Relatedness was found in the participants analysis, $F_1(1,41) = 5.46, p = .024, \text{MSE} = 2,437$, that was marginally significant in the items analysis, $F_2(1,17) = 3.33, p = .086, \text{MSE} = 996$. Furthermore, a main effect of Phonological Relatedness was obtained, $F_1(1,41) = 93.66, p = .000, \text{MSE} = 81,196, F_2(1,17) = 23.95, p = .000, \text{MSE} = 36,309$. In contrast, the Semantic x Phonological Relatedness interaction was not significant in either subjects or items analysis, $F_1 = 0.02, F_2 = 0.02$. Newman-Keuls tests indicated that the Control condition and the Phonologically Related condition did not differ from each other. Likewise, the Phonological and the Semantically & Phonologically Related condition did not differ. Finally, the Unrelated and the Semantically Related condition did not differ from each other.

Discussion

If one ignores the results from the Semantically & Phonologically Related condition, the findings from this experiment once again replicated Schriefers et al.'s (1990) findings in that an early stage of exclusively semantic effects contrasted with a later stage of exclusively phonological effects. In accordance with the findings from Experiment 1, the intermediate SOA yielded both semantic interference and phonological facilitation. Distractor words that were semantically as well as phonologically related to
the picture name showed an attenuated semantic interference effect at the earliest SOA (-150 ms), in which the prediction from additivity for these items was underestimated 20 ms at SOA = -150 ms. At the intermediate SOA (0 ms), the semantic interference effect for these items was entirely eliminated. Only at the latest SOA did semantic and phonological effects roughly add up for the semantically as well as phonologically related distractor words.

These findings appear to contradict the predictions made from the discrete two-step model of speaking. Such a model hypothesizes complete independence of lemma and lexeme stages and as a result would predict an additive relationship among the two experimental factors of semantic and phonological relatedness. Before this conclusion is accepted, the validity of the additive-factors method should be assessed in more detail. According to Pachella (1974), the additive-factors method is based on the premise that the reaction time interval is filled with a sequence of stages that provide various informational transformations. These cognitive stages are assumed to be entirely modular such that the manipulation of experimental factors solely affects the duration of processing taking place on a particular cognitive stage, but not its content. Data from an experiment employing the additive-factors method can be informative in two ways. First, under a "functionally based approach to theorizing" (p.57), the presence of a statistical additivity among a pair of experimental factors can serve as an operational device to infer the existence of corresponding processing stages. Second, there are psychological issues in which there already exists some independent and external justification for assuming the existence of two or more independent stages. In such a case, the additive-factors
framework can test specific assumptions about the character of these stages: additivity between the relevant experimental factors indicates that the corresponding stages indeed adhere to a modular definition. In contrast, an interaction shows that the processing stages are not modular in respect to each other. Note that this use of the additive-factors method does not exclude the possibility that two factors might affect a single stage additively or that they affect different stages while interacting. However, in this case the definition of what characterizes a “stage” would be undermined. In this sense, the additive-factors method provides a method of testing specific claims about the definition and character of particular processing stages.

It is this latter sense of interpretation that is relevant to the question of modularity vs. interactivity in speaking. The existence of separate semantic and phonological representational stages in speech production is an undisputed issue in the literature; indeed, Caramazza and Miozzo (1997, p.329) recently called it "... as close to a universally shared position as anything is in cognitive science". The temporal sequence of semantic interference and phonological facilitation, once again replicated in the present experiment, further supports the assertion that lexical access in speech production involves two separable representational stages. Accordingly, the additive-factors method can be used to test the modularity assumption of the discrete two-step model.

To reiterate, the two-step model assumes that lexical-semantic selection converges on a single candidate which subsequently undergoes phonological encoding at the lexeme level. Specifically, only one lexical-semantic item undergoes phonological encoding, and lexeme retrieval has no influence on the preceding stage of lexical-semantic selection. It
is clear that this characterization of lexical access in speaking adheres very closely to the modular nature of a cognitive stage suggested by the additive-factors method. However, if these assumptions of the two-step model hold, it is hard to see how such an account could explain the interaction obtained in Experiment 3. As Roelofs et al. (1996, p.247) correctly note, the finding of an interaction among two experimental factors does not exclude the possibility that these two factors affect separate representational stages. However, this case would undermine the definition of a cognitive stage as provided by the additive-factors method, namely its modular sense.

In the following, I will discuss whether models of speech production that abandon the assumption of strict independence between the involved representational stages might be able to provide a plausible account of the obtained pattern. To this aim, I will look more closely at Roelofs' (1992a, 1992b) discrete model of speech production and then consider a modification of this model that provides it with interactive characteristics.

Figure 8 displays a schema of Roelofs' model. Within this framework, it is assumed that distractor words gain simultaneous access to their corresponding lemma entries (route a) as well as their lexemes (route b). An unrelated distractor word will compete with the target lemma, and since lexical access at the lemma level involves a selection probability function, this competition will result in a prolonged simulated selection time compared to a condition in which no distractor words are presented. In order to simulate a semantic relationship between a distractor and a target, it is assumed that conceptual similarity will cause spreading activation between the two corresponding concepts. As a result, a trade-off of activation at the lemma level will arise. For reasons
detailed in Roelofs (1992a, 1992b), this mutual trade-off will benefit the distractor more than it benefits the target lemma, and consequently, lemma retrieval of the target will be slowed even more than in the unrelated condition. In contrast, *phonologically related* distractors will activate form-related lexeme representations such as the target lexeme, and consequently, lexeme encoding will be accelerated. In this way, the model provides a plausible account of the general interference caused by the presence of a distractor word, as well as the more specific effects of semantic interference and phonological facilitation obtained in the picture-word interference effect. In order to account for the sequential pattern of these effect, the model would have to assume that distractor activation
transmitted to the lexeme level decays quickly. As a consequence, distractor words that are presented before the onset of the picture (i.e., at earlier SOAs) would show no effect on the stage of phonological encoding of the target because their activation had already died out by the time this stage was reached. Note that this model is a discrete two-step model in that activation is assumed to proceed unidirectionally from the lemma to the lexeme level. The mechanisms outlined above that account for semantic interference and phonological facilitation would predict an additive effect of these two factors, as the contributing representational levels are entirely independent of each other.

Interactive models of speech production (e.g., Dell, 1986, Stemberger, 1985) differ from such an account in that they permit feedback from the lexeme to the lemma level. In this way, phonological activation can to some degree exert an influence on the earlier semantic-syntactic level. As described in the introduction, such a feedback link might provide a plausible explanation of why so-called "mixed" speech errors (i.e., errors that are semantically and phonologically related to the target word) occur more frequently than would be expected by chance. In order to modify Roelofs' framework, it could be assumed that the link between the stages of lemma and lexeme is bidirectional. In this way, distractor words can transmit activation to the lemma level (via route a in Figure 4), to the lexeme level (via route b) and back from the lexeme to the lemma level (via feedback between the levels).

Such a feedback connection has no obvious consequences for semantically related distractor words: the mechanism accounting for semantic interference outlined above does not involve the lexeme level, and as a result, the same mutual trade-off between
target and distractor lemma activation should take place. However, the feedback link has an effect on form-related distractor words. A phonologically related distractor would be assumed to gain access to its corresponding lexeme level representations and consequently prime form-related units, such as the target lexeme. If activation is permitted to flow backwards through the system, then it can be assumed that not only the target lexeme, but also its corresponding lemma unit receives some priming. Note that the hypothesized feedback connection cannot be very strong as otherwise form-related priming effect would be predicted to appear at earlier SOAs, which is only the case to a small extent.

Now consider the special case of a semantically as well as phonologically related distractor word. The mechanism described above predicts that such words - via their phonological relationship with the target - activate the target lexeme as well as to some extent the target lemma via feedback. At the same time, conceptual overlap with the target word prolongs selection time for the target lemma. However, because the correct lemma node has already received some priming via feedback from the lexeme level, semantic overlap now has less time to exert its influence on lemma retrieval. In other words, the semantic interference caused by semantically and phonologically related distractors will be attenuated through the priming via the feedback link. In this way, both a semantic and a phonological relationship could affect lemma retrieval time. A non-additive relationship between the two factors would be predicted if the model itself showed non-linear characteristics. For instance, if units at the lemma level possess a nonlinear activation function, the resulting pattern for the semantically and
phonologically related distractors would be non-additive; this type of nonlinear activation dynamics is implemented in a large number of existing psycholinguistic models (e.g., Harley, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981). Other possible mechanisms of non-linearity are the marking of lexical selection by a jolt of activation added to the target item, as in Dell's (1986) model, or the implementation of lateral inhibition at the lexical level, as in Harley's (1991) and Cutting and Ferreira's (1997) models. Each of these mechanisms would introduce a nonlinear character to the model which could potentially account for the pattern of results observed in Experiment 3. As a consequence, the effects of the two factors of semantic and phonological relatedness would not be expected to add up, and an interaction would arise, just as was demonstrated in Experiment 3. More specifically, semantic interference would be attenuated by a simultaneously present phonological relationship between distractor and target.

Note that the proposed mechanism is contingent on the presence of at least some degree of phonological facilitation per se at early SOAs, as the hypothesized transmission of activation back from the lexeme onto the lemma level predicts a detectable influence of form relatedness by itself. Indeed, at an SOA of -150 ms, facilitation obtained from phonologically related distractors was numerically weak and statistically not significant, and correspondingly, the interaction between the two experimental factors was not very powerful (significant in the participants, but not in the items analysis). In contrast, an SOA of 0 ms yielded a robust phonological facilitation effect of 19 ms. As a result, the feedback mechanism was able to effectively eliminate the interference effect created by
the semantic overlap between distractor and target when a simultaneous phonological relationship was present. At an SOA of +150 ms, semantic interference per se had all but disappeared, and as a result, the feedback of activation ceased to show an influence on response times. Here, semantically as well as phonologically related distractors yielded effects comparable to those obtained from form-related items.

A problematic aspect of such an account might derive from the fact that semantically and phonologically related distractors always yielded an effect comparable to the one obtained from phonologically related distractors. In other words, the semantic interference effect for theses distractors was statistically (but not numerically) eliminated at SOA = -150 ms, and entirely disappeared at SOA = 0 ms. From an intuitive viewpoint, it is less than obvious whether a feedback mechanism as the one hypothesized above can account for the finding that a phonological relationship between distractor and target not only reduces, but "overrides" a simultaneously present semantic relationship. Only an actual computational implementation within an interactive activation framework of speech production would show whether this type of architecture could yield such an effect. In the second section of this dissertation, I will attempt to provide such a computational account.

A very similar proposal has been forwarded in a recent study by Cutting and Ferreira (1997). This study made use of the fact that homophones by definition have dissimilar conceptual characteristics and lemmas, but identical word forms. Participants named pictures of homophones (e.g., a "ball" as a toy) while confronted with distractor words that were semantically related to the alternative meaning of the picture name (e.g.,
"dance", which is related to "ball" as a festive event). The authors demonstrated that such distractors significantly sped up naming latencies at an SOA of -150 ms. Because in this condition distractor and target are conceptually unrelated, the obtained interference effect must be mediated via the fact that both meanings of "ball" have identical phonological representations. Similar to our account of the interaction between semantic and phonological relatedness outlined above, the authors hypothesized a feedback link from the lexeme level to the lemma level. In this way, the distractor "dance" could activate a semantically related lemma like "ball" as a festive event which in turn would undergo phonological encoding. If feedback from the lexeme to the lemma level is assumed, then lemma retrieval for "ball" as a toy would be accelerated. Hence, distractors semantically related to the opposite meaning of a homophonic picture could reduce the time required to access the correct lexical entry of the homophonic picture.

This account faces a similar problem as our hypothesis outlined above: the critical effect appears to involve phonological codes, yet phonologically related distractors usually do not show a substantial effect on naming latencies at this SOA. Cutting and Ferreira argued that this was the case as phonological distractors merely share a few segments with the target name and therefore do not provide the target lemma with much activation through the feedback connection. In contrast, homophones possess identical word forms, and as a result, the activation transmitted via the feedback link is strong enough to reduce lemma retrieval time. Whether this is a plausible hypothesis remains to be seen, but it is important to note that their hypothesis is essentially identical to the one forwarded to account for the results from Experiment 3.
Finally, it is worth considering a claim that proponents of the discrete view of speaking have forwarded in order to account for "mixed" speech errors without abandoning the central assumptions of their view. It has been argued (e.g., Baars et al., 1975; Levelt, 1989) that speech output is supervised by a production-external "monitoring system" which filters out most of potential errors before they overtly occur. "Mixed" speech errors are assumed to occur at an overproportional rate because here the error item resembles the intended word in a twofold respect and is thus less likely to be filtered out by the monitoring system. Whereas the "monitoring" hypothesis constitutes a plausible possibility in regard to speech errors, it is hard to see how it would account for my findings from the picture-word interference procedure. After all, here the speaker produces the correct word in the overwhelming majority of cases, and the monitor would be assumed to accept all these responses. Under these circumstances, I believe that my finding provides evidence for a production-internal mechanism through which the factors of semantic and phonological relatedness modify each other. These findings should not be understood as contesting the existence of monitoring systems in general. There is an intuitive plausibility to the notion that speakers do indeed monitor their own output, as well as some independent experimental evidence (e.g., Baars et al., 1975; Levelt, 1983; Motley, Camden, & Baars, 1982). However, it is unclear how such a monitoring system could account for the findings obtained in Experiment 3.

In summary, Starreveld and La Heij (1995, 1996a) raised relevant issues, but used a procedure that does not provide clear answers. In contrast, my own study used a procedure that, by employing auditory distractors, reflected the respective contributions of
lemma and lexeme stages of lexical access. In accordance with Starreveld and La Heij's findings, an interaction between semantic and phonological relatedness was obtained. Such a finding might be best explained by abandoning the notion of discrete, modular stages in speech production and allowing for some extent of interactive feedback between the representational levels. The following experiment constitutes a further attempt to distinguish the two models.

EXPERIMENT 4 - MEDIATED RELATIONSHIPS

As outlined in the introduction, there are several types of complex priming patterns that are suitable to test whether lexical access is modular or interactive in nature. The previous experiment employed distractors that were both semantically and phonologically related to the target picture and obtained results that appeared to argue against a modular and in favor of an interactive view of speech production. "Mediated" lexical relationships constitute a further way to investigate this issue. In such pairs, primes/distractors and targets are only indirectly linked via an intervening lexical item (CAT-(dog)-FOG or FOG-(dog)-CAT). Such mediated chains of items incorporate both a semantic and a phonological component, and thus, effects caused by these items appear to require a complex interplay between the involved representational levels which is permissible in interactive accounts of speaking. In contrast, the absence of mediated effects might be interpreted as favoring a modular account of lexical access that denies the possibility of interacting semantic and phonological representations.
Relatively few studies have investigated the effects of mediated items. In a study on speech comprehension, McNamara and Healy (1988) employed a lexical decision task in which participants responded to pairs of words that were either semantically related (LIGHT-LAMP), rhymed (LAMP-DAMP), or were related by means of an intervening item (LIGHT-DAMP). In accordance with earlier studies, the presence of either a semantic or a form relationship between the two items yielded a response time facilitation. The authors argued that mediated word pairs should show a likewise facilitation effect if semantic and phonological levels interact with each other. Contradicting this prediction, an inhibitory effect was obtained for these items. By means of a further experiment that used a self-paced reading task, as well as a manipulation of nonwords, McNamara and Healy were able to attribute this effect to strategic processes operating at a postlexical point in time rather than automatic processes. The authors concluded that their experiments provided no evidence for the hypothesis that spreading activation can occur between semantic and phonological levels.

Levelt et al. (1991) pursued a similar logic in a study that more specifically targeted speech production issues. A series of experiments employed a procedure that combined picture naming and auditory lexical decision: Participants named pictures of objects presented on a screen, while on a certain percentage of trials the preparation of the naming response was interrupted by the presentation of an auditory probe word on which participants performed a lexical decision. Picture and probe word bore various relationships to each other, and the picture-probe onset interval was systematically varied. In accordance with the findings from the picture-word interference studies, semantic
interference was found at the shortest SOA interval, which vanished at longer SOAs.

Somewhat deviating from the findings with the picture-word interference procedure, a
phonological relationship yielded an interference - instead of facilitation - effect which
was pronounced at both short and long SOAs.

Importantly, two further experiments tested one of the central assumptions of the
discrete view of speaking which holds that only one lexical item, namely the selected
target word, receives phonological encoding. According to this claim, conceptually
related words which are co-activated during lexical selection should never transmit their
activation to the phonological stage. Levelt et al. tested this assumption by employing
mediated probe words that were indirectly related to the picture name. Experiment 5
tested the effects of naming preparation of a picture (SHEEP) whose close associate
(WOOL) was form-related to an auditory probe word (WOOD). Similarly, Experiment 6
examined naming preparation of a picture (SHEEP) whose category coordinate (GOAT)
was form-related to an auditory probe (GOAL). In both cases, it was argued that if
phonological encoding is restricted to the target item, then response times for the
mediated probes should not differ from those for unrelated probes. These predictions
were confirmed in both experiments: mediated probe words did not significantly differ in
their response times from the unrelated condition. Levelt et al. interpreted this failure as
being in accordance with the predictions derived from the discrete account of speech
production, but incompatible with interactive accounts.

A problematic aspect about this conclusion was pointed out in a reply by Dell and
O'Seaghdha (1991; see also Dell & O'Seaghdha, 1992): the absence of mediated effects
might not be diagnostic with regard to the issue of interactivity vs. modularity. Whereas the presence of mediated effects would appear to require an interplay between semantic and phonological codes that is not permitted in modular models, the failure to find such effects - as in Levelt et al.’s experiments - does not argue against the notion of interactivity. This is because an interactive framework does not necessarily entail a detectable effect of a mediated relationship: For one lexical item to have an effect on a second, mediated item, activation of the first item must first spread to semantic neighbors, which will receive a fraction of the original activation. Then, one of these neighbors must in turn spread a fraction of its own activation to the critical, form-related item in order to have an effect on its processing. As there are two steps involved, each one of which has an attenuating influence, Dell and O’Seaghdha argued that it is far from clear whether mediated effects are detectable in response time studies. Accordingly, the absence of such effects should not be used as definitive evidence against interactive characteristics of speech production.

Despite the fact that a mediated effect presumably is rather small in size even if lexical access is indeed interactive, O’Seaghdha and Marin (1997) made a further attempt to track down such an effect. In their study, participants named visually presented target words that were preceded by either unrelated or related prime words. Compared to the lexical decision task employed in McNamara and Healy’s study described above, naming has been shown to be less susceptible to strategic confounds (e.g., Seidenberg, Waters, Sanders, & Langer, 1984). Pooled across four experiments and 176 participants, a facilitation of five milliseconds was obtained when primes were semantically related to a
word form-related to the target word, an effect that was significant when analyzed by subjects, but not by items. Two further experiments in which the primes were masked yielded even smaller effects. Nevertheless, the authors took the results as evidence for the interactive characteristics of lexical access, conceded, however, that detection of such an effect might reach the limits of sensitivity of a naming task. In support of their conclusion that the mediated effect revealed the interactive architecture of lexical access, O’Seaghdha and Marin also demonstrated that when mediated items were categorized according to the strength of their semantic relationship (as measured by the size of the semantic effect obtained from the related pair within the mediated chain), strongly related item pairs showed a mediated effect of 7 ms, whereas weakly related item pairs merely yielded an effect of 1 ms. The fact that the effect size appeared to depend on the “strength” of the semantic relationship was taken to suggest that the overall mediated effect, although very small, was nevertheless “real” and diagnostic of the interactive nature of lexical access.

If this conclusion holds and the effects of mediated relationships are so small as to be virtually undetectable, one possible way of magnifying them would be to reinforce the strength of either the semantic or the phonological relationship within the mediated chain. That is, one could use form-related items that are almost identical, or one could employ near-synonymous pairs that are virtually indistinguishable in terms of their semantic characteristics, but are nevertheless lexically distinct. The latter strategy was pursued in the only study to date that has investigated mediated relationships within the picture-word interference paradigm. In a recent short report by Jescheniak and Schriefers (1997),
participants named pictures with two near-synonymous names (rooster: HAHN and GOCKEL in German), one of which (HAHN) was clearly dominant. Distractor words that were form-related to the non-dominant (and thus unselected) name of the picture (GONDEL; gondola) were shown to inhibit the production of the dominant picture name. This finding was interpreted as showing that under certain circumstances, as in the production of near-synonymous names, more than one lexical-semantic item undergoes phonological encoding, contradicting one of the central assumptions of the discrete two-step view of speaking. However, the authors cautioned against generalizing these findings beyond the special case of synonyms. Possibly, the production of concepts with more than one lexical label differs qualitatively from the one for non-synonymous items, and thus phonological encoding of more than one lexical item might be restricted to this uncommon case. Jescheniak and Schriefers pointed out the necessity of further research to investigate whether their effect generalizes to items that are closely related, but not synonymous.

In summary, studies that have investigated mediated effects in lexical access have yielded somewhat ambiguous results. The initial positive finding in McNamara and Healy's (1988) study was probably caused by strategic processes, whereas studies that are less subject to this confound have either failed to show mediated effects (Levelt et al., 1991), obtained findings so small that a meaningful interpretation is questionable (O'Seaghdha & Marin, 1997), or demonstrated mediated effects only for the case of near-synonymous lexical items that might not generalize to lexical access for non-synonymous words (Jescheniak & Schriefers, 1997).
To investigate whether mediated effects could be demonstrated in the picture-word interference procedure, I conducted a pilot study in which 28 pictures were paired with either unrelated or semantically related words. In addition, I employed distractor words (LIME) that were form-related to a category coordinate (LION) of the picture (TIGER). At the earliest investigated SOA (-150 ms), a highly significant semantic interference effect of 83 ms was obtained. Critically, the mediated condition yielded a significant interference effect of 32 ms. The mediated effect was reduced at SOA = 0 ms and disappeared at SOA = +150 ms. These findings appear to suggest that Jescheniak and Schriefers’ findings are not restricted to the special case of near-synonyms, but might indeed generalize to conditions in which the primed intervening item is merely a category coordinate of the target picture.

An account of such an effect in terms of an interactive model of speaking would assume that a mediated distractor (LIME) spreads part of its activation to form-related words at the phonological level (LION). In an interactive framework, activated nodes at the phonological level can, via feedback, exert an influence onto the semantic level. Consequently, such a mechanism could allow LION to interfere with lexical selection of the target TIGER. Furthermore, since the distractor interferes with semantic-syntactic target selection, such an effect would most likely be located at the earlier SOAs which show interference from semantically related distractors. Indeed, this is where the effect was actually found in the pilot experiment.

There are a number of problems related with such an account, the most serious of which is that the same mechanism which causes the mediated effect should also lead to
facilitation in the phonologically related condition. However, this is clearly not the case: at the SOA for which the mediated effect was obtained (SOA = -150 ms), no such phonological facilitation effects are normally obtained. In contrast, phonological facilitation effects are exclusively found at later SOAs and are attributable to processes occurring at the lexeme level. Furthermore, it might be argued that the obtained mediated effect has nothing to do with feedback from the lexeme to the lemma level, but might instead arise as a result of phonological priming within the input phonological system. In other words, hearing the distractor word might activate a cohort of form-related words in the input lexicon, which might then interfere with target selection if they are conceptually related. Although this alternative explanation is possible, it should be noted that it, just like the account proposed above, would also predict phonological facilitation effects at the earliest SOA, which is not what was found.

In summary, the findings from my pilot experiment suggested a mediated effect that is possibly diagnostic with regard to the nature of lexical access in speaking. However, the exact dynamics of how such an effect arises are far from clear. Furthermore, stimuli in this experiment were not tightly controlled in terms of their word length and degree of relatedness, and the finding reported above was based on only 12 participants. Thus, the alleged effect deserves more systematic investigation and is further explored in the following experiment.

The complementary condition, namely one in which a distractor is semantically related to an intervening item which is related in form to the picture name, has not been investigated in the picture-word interference procedure so far. Similar to the scenario
outlined above, an interactive framework might predict an effect of this condition: it
could be hypothesized that such a distractor (ORANGE) gains access to its corresponding
conceptual representation and co-activates a semantic-syntactic node that shares part of
its semantic specification (LIME). One of the assumptions of the interactive account is
that multiple semantic-syntactic items undergo phonological encoding in speaking. Thus,
both the co-activated LIME and the target item LION might activate their corresponding
word forms. Since these two items share phonological characteristics, the target LION
might receive some additional activation, resulting in facilitation from the mediated
condition. Note that, since this effect concerns the stage of phonological encoding, it
would be predicted to reside at a later stage of naming preparation, namely at an SOA that
reveals facilitation from phonologically related distractors.

The following experiment constitutes an attempt to investigate both types of
mediated relationship more systematically. The general design of this experiment is very
similar to the one of Experiment 3. However, instead of employing distractors that are
both semantically and phonologically related, mediated distractors of the two types
outlined above were incorporated. The potential finding of an effect of one or both of
these indirectly related conditions might be taken as evidence for the interactive nature of
lexical access in speaking. Specifically, distractors form-related to a category coordinate
of the picture name were predicted to yield interference at SOAs indicating lexical-
semantic retrieval (-150 ms and 0 ms). In contrast, distractors semantically related to an
item form-related to the picture name were predicted to yield facilitation at SOAs
indicating phonological encoding (0 ms and +150 ms).
Method

Participants. Eighteen Rice University undergraduate students served as
participants in order to meet a class requirement for experimental participation.

Materials. Twenty-four line drawings of common objects were selected from the
Snodgrass and Vanderwart (1980) set. For each of these pictures, a category coordinate
was selected that served as the semantically related distractor word. A further word was
chosen that shared a minimum of the initial two phonemes with the picture name and
served as the phonologically related distractor. Also, a word was chosen that stood in no
such relationship to the picture name and served as the unrelated distractor.

For the first mediated condition, a word was chosen that shared a minimum of the
initial two phonemes with the semantically related distractor of a picture. This condition
constitutes the analog to the one employed in Jescheniak and Schriefers's study and my
own pilot experiment and will henceforth be referred to as Mediated I. The second
mediated condition was formed by selecting a word that was a category coordinate of the
phonologically related distractor of a picture. This condition has not been previously
investigated and will be referred to as Mediated II. Distractor words in all five conditions
were matched with regard to the average number of phonemes. A further condition was
also employed in which no distractor was presented (Control).

Design. The experimental design included SOA (-150 ms, 0 ms, and +150 ms)
and Type of Relatedness (Control, Unrelated, Semantic, Phonological, Mediated I, and
Mediated II) as within-subjects factors. Each set of 144 picture-word pairs was displayed
three times at different SOAs. The order in which participants received the SOA blocks
was balanced according to a Latin Square design. Items were presented in a pseudorandom fashion such that the same picture never appeared twice on subsequent trials.

**Apparatus.** The same apparatus as in the above experiments was used.

**Procedure.** The general procedure was again similar to the one employed in the previous experiments. Three SOA blocks (-150 ms, 0 ms, +150 ms) were administered. Again, each participant was familiarized with the pictures before the experiment began. Subsequently, two practice blocks of 24 trials each, one without and one with auditorily presented unrelated distractors, were conducted. Finally, the three experimental blocks of 144 trials each were administered, resulting in a total of 432 experimental trials for each participant. Each testing session lasted approximately one hour.

**Results**

Responses judged to be incorrect by the experimenter, response times shorter than 250 ms or longer than 2000 ms, and response times deviating more than three standard deviations from a participant’s conditional mean were eliminated. These procedures resulted in the exclusion of 2.7%, 0.3%, and 1.4% of the data, respectively. A total of 4.4% of data points were deleted in this way.

Table 4 displays the mean response times, varied by SOA and picture-distractor relation. Figure 9 displays the semantic, phonological and the two mediated effects (Unrelated minus Related condition). The results show a pattern of results that are familiar from Experiment 1 with auditory distractors and Experiment 3: early semantic interference vanishes at the later SOAs, and phonological facilitation builds up to a
Table 4
Experiment 4: Mean Response Latencies (in ms), varied by Relatedness (Control vs. Unrelated vs. Semantic vs. Phonological vs. Mediated I vs. Mediated II) and Picture-Word Onset Asynchrony (SOA). Mediated I: Distractor is form-related to a Category Coordinate of the Picture; Mediated II: Distractor is a Category Coordinate of a Word form-related to the Picture

<table>
<thead>
<tr>
<th>SOA (in ms)</th>
<th>-150</th>
<th>0</th>
<th>+150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>631</td>
<td>612</td>
<td>618</td>
</tr>
<tr>
<td>Unrelated</td>
<td>683</td>
<td>673</td>
<td>657</td>
</tr>
<tr>
<td>Semantic</td>
<td>713</td>
<td>696</td>
<td>662</td>
</tr>
<tr>
<td>Phonological</td>
<td>677</td>
<td>650</td>
<td>619</td>
</tr>
<tr>
<td>Mediated I</td>
<td>682</td>
<td>672</td>
<td>654</td>
</tr>
<tr>
<td>Mediated II</td>
<td>681</td>
<td>683</td>
<td>660</td>
</tr>
</tbody>
</table>

maximum at the latest SOA. In contrast, both mediated conditions appear to yield only negligible effects on response times.

A 3 (SOA) x 6 (Type of Relatedness) ANOVA was conducted on the data with both factors as within-participants factors. A main effect of SOA was obtained, $F_1(2,34) = 3.49, p = .042, MSE = 29,444, F_2(2,46) = 39.07, p = .001, MSE = 36,381$, as well as a main effect of Condition, $F_1(5,85) = 33.44, p = .001, MSE = 4,252, F_2(5,115) = 41.76,$
Figure 9. Experiment 3: Effects of semantically related, phonologically related, and mediated distractors (unrelated minus related condition) varied by SOA (-150 ms, 0 ms, +150 ms). Mediated I: Distractor is form-related to a category coordinate of the picture; Mediated II: Distractor is a category coordinate of a word form-related to the picture.

\[ p = .001, \text{MSE} = 4.252 \]. Furthermore, a significant SOA x Condition interaction was obtained, \( F_1(10,170) = 3.08, p = .001, \text{MSE} = 1.827, F_2(10,230) = 3.38, p = .000, \text{MSE} = 2.339 \).

Next, the effect of Type of Relatedness was analyzed on each level of the factor SOA separately. At SOA = -150 ms, Type of Relatedness was highly significant, \( F_1(5,85) = 21.62, p = .000, \text{MSE} = 12.686, F_2(5,115) = 20.05, p = .000, \text{MSE} = 17.097 \). Newman-Keuls pairwise comparisons indicated that the Control condition differed significantly from all other conditions. Likewise, the Semantic condition differed
significantly from all other conditions. In contrast, the Unrelated, the Phonological, the Mediated I, and the Mediated II conditions did not significantly differ from each other.

At $\text{SOA} = 0 \text{ ms}$, Type of Relatedness was found to be highly significant, $F_1(5, 85) = 16.43$, $p = .000$, $\text{MSE} = 15,707$, $F_2(5, 115) = 22.07$, $p = .000$, $\text{MSE} = 21,210$. Newman-Keuls tests indicated that the Control condition differed significantly from all other conditions. The Unrelated, the Mediated I, and the Phonological condition did not differ from each other, but differed from all other conditions. Likewise, the Mediated I, the Unrelated, the Mediated II, and the Semantic condition did not differ from each other, but from all other conditions.

Finally, at $\text{SOA} = +150 \text{ ms}$, Type of Relatedness was again highly significant, $F_1(5, 85) = 12.60$, $p = .000$, $\text{MSE} = 7,741$, $F_2(5, 115) = 16.58$, $p = .000$, $\text{MSE} = 10,334$. Newman-Keuls tests indicated that the Control and the Phonological condition did not differ from each other, but differed from all other conditions. Likewise, the Mediated I, the Unrelated, the Mediated II, and the Semantic condition did not differ from each other, but differed significantly from all other conditions.

Discussion

This experiment once again replicated the pattern of semantic interference followed by phonological facilitation in the picture-word interference procedure first shown by Schriefers et al. (1991). However, neither one of the two mediated conditions appeared to have an effect on naming latencies; in particular, the effect of the Mediated I condition which had been obtained in the pilot experiment could not be replicated. The largest deviation from the unrelated condition was obtained at $\text{SOA} = 0 \text{ ms}$ for the
Mediated II condition (an interference effect of 10 ms), but this effect was not even marginally significant. Note furthermore that the interactive framework outlined in the introduction to this experiment would have predicted a facilitatory rather than an inhibitory effect for the Mediated II condition. The fact that no significant mediated effects were obtained, in particular in the Mediated I condition employed in the pilot experiment, raises the question of how this failure can be reconciled with the earlier positive finding. One possible explanation concerns the strength of relatedness within each mediated chain: As described in the introduction of this experiment, O'Seaghdha and Marin (1997) argued that a mediated effect should be of rather small size, and furthermore that its size should depend on the degree of relatedness within the two constituents of the mediated chain (i.e., for the Mediated I condition: semantic relatedness between picture and intervening item, phonological similarity between intervening item and mediated distractor; the reverse for the Mediated II condition). The stimuli for the pilot study were assembled in a somewhat loosely defined manner. Indeed, this raises the possibility that the mediated distractor-picture pairs in Experiment 4 were somehow less related than their counterparts in the pilot experiment.

To evaluate this hypothesis, both semantic and phonological relatedness within the mediated picture-distractor pairs must be investigated separately. Semantic relatedness within a mediated chain can be gauged by the size of the corresponding semantic interference effect since the intervening word itself is used as distractors in the semantic condition (e.g., the intervening item LION in the mediated combination LIME-(LION)-TIGER is used as a semantic distractor paired with TIGER). Experiment 4
yielded a semantic interference effect of 30 ms at SOA = -150 ms which is in line with the findings from Experiment 3 (32 ms) and Experiment 1 (28 ms at SOA = -200 ms). In contrast, the semantic effect in the pilot experiment was much larger (83 ms). On the basis of this difference, it might be argued that the semantic distractors in the pilot experiment were more strongly related to their corresponding pictures than those used in Experiment 4. However, a confound consists of the fact that SOA was a between-subjects factor in the pilot experiment whereas it was a within-subjects factor in Experiment 4 (and all other experiments in this dissertation). Repeated exposure to the same picture-distractor pairings appears to diminish the size of the semantic effect. Hence, the effect sizes are not directly comparable across the two experiments. Only conducting both experiments with the same design would show whether this - seemingly inconsequential - difference in design could entail such a dramatic effect. Alternatively, the stimulus pairings could be subjected to a semantic relatedness rating in which participants judge on a scale how closely related two words are. However, despite the variation in its size, both experiments yielded a semantic interference effect that was significant. This fact makes it a rather remote possibility that the semantic link within the mediated chain was strong enough in the pilot study, but too weak in Experiment 4, to yield a mediated effect.

The phonological overlap between the mediated distractor and the intervening word might be another factor that differentiates the pilot experiment from Experiment 4: in the pilot study, overlap sometimes consisted of only the initial phoneme (JUG-JAM) and sometimes of a much larger congruence (TANGERINE-TAMBOURINE). In
Experiment 4, just as in all other experiments in this thesis, phonological relatedness was defined as an overlap of a minimum of the initial two phonemes. To locate the difference between the two experiments in the strength of phonological similarity, one would have to claim that the phonological link in the pilot experiment was somehow stronger than the corresponding one in Experiment 4, so that mediated effects arose in the former, but not in the latter. However, calculation of a phonological overlap score in a manner similar to Experiment 3 (described in Footnote 2) showed a slightly lower score (0.48) in the pilot experiment than in Experiment 4 (0.56). In summary, neither differences in semantic relatedness nor differences in phonological overlap are likely to account for the differing outcomes in the mediated condition.

A further aspect in which pilot experiment and Experiment 4 might have differed is the unrelated condition. In the picture-word interference task, stimuli are usually selected for the unrelated condition in such a way that they bear no obvious relationship to the picture name. To serve as a baseline for comparison with the other - semantic and phonological - conditions, stimuli in all conditions are usually matched in terms of their average number of phonemes or letters. This was not the case in the pilot experiment. If word length - disregarding the relationship between picture and distractor - is correlated with the interference the word causes, then the stimuli in the unrelated condition might have possibly underestimated the baseline and consequently overestimated the interference created by semantic and mediated distractors. However, a post-hoc analysis of the average number of phonemes in each condition indicated roughly comparable indices for the three critical conditions (an average word length of 4.3 phonemes in the
unrelated condition, one of 4.4 phonemes in the semantic condition, and one of 4.5 phonemes in the mediated condition; none of these differed significantly from each other). In comparison, in Experiment 4, these three conditions were almost exactly matched in their average word lengths (unrelated condition: 4.3 phonemes; semantic condition: 4.3 phonemes; mediated condition: 4.2 phonemes). Of course, since phonemes vary in their duration, average word length provides only a rough estimate of the real duration of the distractor words.

In summary, at present there is no satisfactory account of why the pilot experiment and Experiment 4 yielded such divergent results. However, given that design and stimulus selection of Experiment 4 are in accordance with all other experiments reported in this thesis, the conservative conclusion at this point has to be drawn that there are no mediated effects in the picture-word interference procedure. This negative finding is in accordance with the results from Levelt et al.'s (1991) picture naming/auditory lexical decision procedure, and in a similar manner might be interpreted as favoring a discrete account of speaking. However, for the reasons detailed in Dell and O'Seaghdha (1991) and outlined in the introduction to this experiment, the failure to obtain such indirect effects is probably not diagnostic with regard to whether lexical access in speech production is modular or interactive. To reiterate, Dell and O'Seaghdha argued that the two links involved in a mediated relationship combine in a multiplicative fashion such that effects of this type are unlikely to be detectable in an experimental situation. If their argument is valid, the results obtained in Experiment 4 would not contribute to the question of the relationship between semantic and phonological codes in speaking. In the
light of this argument, the simulations described in the next section now receive special weight: If it can be shown that an interactive model of speaking can a) simulate the interaction between semantic and phonological relatedness observed in Experiment 3, and b) demonstrate the absence of mediated effects, then the interactive view would receive unequivocal support, and the findings from Experiment 3 and 4 would be compatible with each other.

SUMMARY OF EXPERIMENTAL RESULTS

The experiments reported above purport to investigate a central issue in research on speech production, namely the question of how semantic and phonological representations relate to each other. Whereas "discrete" accounts of speaking propose the independence of the two involved levels, "interactive" models conceive of them as largely interdependent. The picture-word interference procedure employed in this dissertation has been taken to argue for (Schriefers et al., 1991) or against (Starreveld & La Heij, 1995, 1996a) the discrete account of speaking, suggesting that its characteristics and parameters are not fully understood. Experiments 1 and 2 intended to clarify issues related to the respective time course of semantic interference and phonological facilitation obtained with this procedure. Experiment 1 directly contrasted visual and auditory distractors. In accordance with existing reports, a sequence of semantic and phonological effects was obtained with auditory distractors, whereas visually presented distractors yielded phonological effects that preceded the onset of the semantic effects. It was suggested that this discrepancy might result from the previously neglected fact that, while
auditory distractors have a fixed duration, visual distractors remained on the screen until a response was made. Therefore, even distractors presented at a relatively early (negative) SOA might provide visual input long enough to affect phonological encoding of the picture name, resulting in “early” phonological facilitation effects. This hypothesis was verified in Experiment 2, which demonstrated that visually presented distractors yielded a sequence of semantic and phonological effects similar to the pattern from auditory distractors if their presentation duration was limited.

With this insight, the initial question of how semantic and phonological representations relate to each other could be newly explored. On a general level, predictions derived from discrete vs. interactive models of speaking can be investigated by means of complex forms of lexical relationships. With regard to the picture-word interference procedure, complex relationships are those that surmount simple semantic or phonological picture-distractor relatedness and consist of either a “mixed” (semantic and phonological) or a mediated (indirect) relationship between distractor and picture name. Experiment 3 employed an additive-factors design in which the two factors of semantic and phonological relatedness were factorially crossed. The obtained interaction was interpreted as violating predictions derived from the discrete view of speaking, but possibly in line with interactive accounts of speech production. Experiment 4 investigated complex distractor-picture relationships in which the two constituents are related via an intervening word. This experiment was based on a pilot experiment in which distractor words form-related to a category coordinate of the picture had been shown to yield significant interference, a finding that appears incompatible with discrete
accounts, but possibly in line with interactive accounts of speaking. Experiment 4 attempted to replicate this mediated effect, as well as investigating its complementary counterpart, namely a condition in which a distractor word is semantically related to a word form-related to the picture name. The results showed that neither of the two mediated conditions had an effect on picture naming times that exceeded the one of unrelated distractors; in particular, the mediated effect obtained in the pilot experiment could not be replicated. This failure to obtain mediated effects in the picture-word interference procedure is not necessarily incompatible with interactive accounts of lexical access, however, since it has been convincingly argued (Dell & O’Seaghdha, 1991) that mediated effects might be so small as to be virtually undetectable even in an interactive framework.

A COMPUTATIONAL MODEL OF SPEECH PRODUCTION

The following section provides an attempt to simulate the empirical evidence described in the above experiments in a computational model. The following key empirical findings to be captured by a model of speaking are:

- Semantically related distractors cause interference.
- Form-related distractors cause facilitation.
- An early semantic interference effect is followed by a later phonological facilitation effect, as evidenced in Experiment 1 (with auditory distractors), Experiment 3, and Experiment 4.
• Semantic and phonological relatedness interact, as observed in Experiment 3.

Specifically, semantic interference appears reduced in the presence of a simultaneous phonological relatedness (SOA = -150 ms and SOA = 0 ms). At the latest SOA (+150 ms), semantically and phonologically related distractors yield an effect identical to phonologically related distractors.

• Mediated distractors fail to show an effect on performance, as observed in Experiment 4.

Of these six findings, the first has probably been most extensively investigated, namely in Roelofs' (1992a, 1992b) computational model of lexical access in speaking. Roelofs (1997) also recently introduced an extension of this model which simulates a variety of phonological priming effects and thus accounts for the second key finding.

However, to my knowledge, this model has not been used yet to simulate the third and fourth type of observation, namely the relative time course of semantic and phonological effects. The interaction between semantic and phonological relatedness has been successfully simulated in a model introduced by Starreveld and La Heij (1996a) which by its very nature, however, fails to account for the sequential time course of semantic and phonological effects. Finally, the effects of mediated distractors have been examined on a very general level by Dell and O'Seaghdha (1991), have not been applied, however, to the picture-word interference procedure.

In the following, I will introduce a simple interactive activation model which borrows heavily from Dell et al.'s (1996) model accounting for speech errors in normal and aphasic speakers, as well as Roelofs' (1992a, 1992b) account of lexical access in
speaking. After I have outlined the general characteristics of the model, I will target each of the five observations described above separately.

**Competitive and Non-Competitive Structures**

The crucial aspect of the picture-word interference procedure consists of the fact that lexical access to the target word occurs while a distracting item is concurrently active in the system. This distracting item can bear various relationships to the target and as a result has an effect on response times. There are some broad principles in which these two nodes, target and distractor, interact with each other. On the most general level, two related nodes exchange some activation via their mutual overlap, whereas unrelated nodes do not. The results of this mutual tradeoff will vary depending on which procedure is chosen to correspond to a response time measure. For the following discussion, an extremely simple network is assumed that only consists of two layers, one of which implements lexical nodes, and the other represents semantic features (note, however, that the same type of relationship could exist between lexical nodes and phonological segments). Such a network is shown in Figure 10.

Input to the network is provided to the conceptual features of the target node, and this activation spreads to the lexical level. Concurrently, a certain amount of activation is added to the activation value of one of the non-target lexical nodes, taken to represent the presentation of a distractor word. Activation at the lexical level can then be taken to represent an analog of response times in one of the following ways:
Figure 10. Simple network implementing a lexical layer ($T = \text{target}$, $R = \text{node related to target}$, $U = \text{node unrelated to target}$) as well as a conceptual layer in which features representing $T$ and $R$ overlap.

(a) Absolute Threshold of Target Activation

It could be assumed that lexical access occurs as soon as one lexical node, normally the target, has exceeded a certain absolute threshold value. Such a measure would essentially ignore the activation state of competing nodes and exclusively focus on the activation of the target.

Under the assumption of an absolute threshold criterion, a relation between target and distractor node will predict facilitation: via their shared semantic features, the distractor node will add a fraction of its activation to the target node, which in turn will reach its criterion faster than before. The fact that the distractor node profits as well from the mutual overlap is irrelevant as the selection criterion ignores the activation of competing nodes.
The following simulation demonstrates this facilitation effect from a related distractor node. Simultaneously with the activation of the conceptual target features, an - either unrelated or related - distractor node receives a jolt of activation. Figure 11 shows that if absolute target activation is taken as a measure of performance, the net effect is facilitatory: in the related condition, the target node reaches an arbitrary selection threshold faster than in the unrelated condition.

The assumption that lexical access occurs without considering the activation state of simultaneously active distractor items is psychologically questionable. More likely, the cognitive system takes into account the target node’s activation as well as those of competing units. It can do so by one of the following mechanisms:

![Graph](image)

**Figure 11.** Absolute selection criterion: Target activation in units as a function of time step and relation to simultaneously activated distractor node (unrelated vs. related).
(b) Choice Rule

One commonly employed mechanism to this aim (McClelland & Rumelhart, 1981; Roelofs, 1992a, 1992b) is Luce's (1959) choice rule:

\[
p(\text{selection}) = \frac{a_{(m,t)}}{\sum_{e} a_{(e,t)}}
\]  

(1)

In this equation, \(a_{(m,t)}\) is the activation of node \(a\) at time step \(t\), and \(\sum_{e} a_{(e,t)}\) represents the summed activation of all competing nodes (including the target node) at that time step. The resulting ratio \(p\) reflects the probability with which target selection occurs. Simultaneously active nodes will lower the denominator and therefore decrease the selection probability of the target node.

The following simulation (Fig. 12) demonstrates that if selection probability is taken as a measure of performance, the constellation described in Figure 10 predicts interference from related distractors. In this case, a selection criterion is reached slower in the related than in the unrelated condition.

(c) Lateral Inhibition

The choice rule is external to the network structure per se since selection probability is calculated on the performance output of the network. A different mechanism by which a competitive structure among nodes can be "built in" is mutual
inhibition (e.g., Cutting & Ferreira, 1997). As in (a), the selection criterion consists of just a simple activation threshold, but the lexical layer incorporates strong inhibitory links between competing nodes. Within such a structure, all competing nodes attempt to inhibit each other, with the most activated node constituting the strongest inhibitory force. Such a process has reasonable support on a neurophysiological (e.g., Feldman & Ballard, 1982; Grossberg, 1978), as well as some plausibility on a psychological level (Dell & O'Seaghdha, 1994). In this case, the mutual overlap between target and distractor will supply the distractor node with some additional activation, which in turn will increase its inhibitory force.

The following simulation (Fig. 13) demonstrates that the implementation of lateral inhibition among lexical nodes will cause interference from related distractors.
Figure 13. Lateral inhibition: Target activation in units as a function of time steps and relation to simultaneously activated distractor node (unrelated vs. semantically related).

Apart from depressing the general level of activation in the network, a selection criterion is reached slower in the related than in the unrelated condition.

With regard to the first step of speech production, or lemma access, it is clear that one or the other form of competitive mechanism must be implemented into the model. If it is assumed that conceptual input temporarily activates a cohort of semantically related items, then a competitive selection mechanism seems required in order to enable the system to settle on the correct lexical candidate.

The situation is a little more complicated in the case of phonological facilitation observed in the picture-word interference procedure. For models that entirely separate lemma from word form retrieval, the seemingly paradoxical situation arises that presentation of an unrelated word prolongs response time relative to a control condition with no distractor (which would appear to imply a competitive mechanism), yet the
presence of a phonological relationship between target and distractor reduces this interference and thus yields a facilitatory effect relative to the unrelated condition (a finding that appears to deny a competitive structure). Roelofs' (1997) model solves this problem by letting articulatory syllabic programs compete with each other, which accounts for the interference from an unrelated distractor. At the same time, by means of the priming of segment nodes and the absence of a competitive structure at this level (cf. Figure 10), a phonologically related distractor word will benefit the target word form and thus create a facilitatory effect that occasionally outweighs the more general interference described above. On a general level, then, this model solves the problem by attributing interference from unrelated distractors and relative facilitation from form-related distractors to different levels within word-form encoding.

The model described below is not intended to provide a detailed account of the full speech production mechanism. Instead, it targets merely the relative time course of semantic interference and phonological facilitation in the picture-word interference procedure. Therefore, I will assume that selection at the lexical level occurs via a competition-sensitive mechanism, whereas selection at the phonological level is absolute and thus allows for phonological priming effects. In other words, it is assumed that there is at least one level within word-form encoding at which no competition takes place. Although this solution might appear to be ad-hoc, the dynamics of interactive activation networks are too complex to take the approach suggested by Roelofs (1997) and implement more than one representational level at the stage of word form encoding.
Representational Format - Features vs. Compositional Structure

A long-standing debate exists in cognitive science as to whether semantic representations are better conceptualized in terms of semantic features (e.g., Smith, Shoben, & Rips, 1974) or in terms of units representing whole concepts (Collins & Loftus, 1975). With regard to speech production, some researchers (e.g., Levelt, 1989; Garrett, 1975) have argued that decompositional conceptual representations face insurmountable problems and that only localized representations can adequately solve these problems (but see Bowers & Vigliocco, 1997). Likewise, it has been argued (i.e., Bock & Levelt, 1994) that the fact that phonological speech errors mostly possess lexical status demonstrates that there must be a unified whole-word representation at the phonological level (but see Dell’s, 1986, and Caramazza and Miozzo’s, 1997, models which specify phonological representations sublexically).

The simple network employed above and shown in Figure 10 specified the conceptual layer as feature-based and is thus an analog to Dell’s (1986) model of speech production. Figure 14 depicts the analog of this model, but with holistic conceptual representations. Figure 15 demonstrates that it is irrelevant for the current matter whether the representational format for the conceptual layer is feature-based or holistic: the choice rule applied to the activation at the lexical layer predicts interference from related items (as in Figure 12 where features instead of holistic units were implemented). Therefore, for the issues at stake in the present discussion, it is irrelevant whether the representational format is based on features or on whole concepts: here, it does not
Figure 14. Simple network implementing a lexical layer (T = target, R = node related to target, U = node unrelated to target) as well as a conceptual layer in which the concepts corresponding to T and R are linked via a similarity connection.

matter whether target and distractor node exchange activation via mutual features, or via a direct link among related holistic concept nodes.

However, the feature-based version has a disadvantage from a modeling perspective: the extent of relatedness can only be varied in discrete steps. That is, although the number of overlapping features can be modified, this can be done only in discrete steps, and such a modification always requires "re-wiring" of the existing connections between concepts and lexical units. In contrast, in a holistic model, overlap can be continuously varied by modifying a single parameter. For this reason, the following simulations will adapt holistic representations at both the conceptual and the phonological layer. Thus, the model's representational format was chosen to allow for the simplest variation of the model's characteristics. This choice should not, however, be regarded as a commitment for or against decompositional structure. At least for the
Figure 15. Model in Figure 14: Target selection probability as a function of time steps and relation to simultaneously activated distractor node (unrelated vs. semantically related).

For present purposes, both the feature and the holistic format render mathematically equivalent results, as demonstrated above.

General Architecture

The model employed for the following simulations is schematized in Figure 16. It is closely modeled after Dell et al.'s (1996) model in that its layers are connected in a fully interactive manner. The model consists of a conceptual, a lexical, and a phonological layer. On each of these three layers, seven units are specified that represent a target picture to be named as well as six conditions that include all the relevant relationships to the target.
Figure 16. Interactive model of speech production. P = Phonologically related to target, S = Semantically related, U = Unrelated, S&P = Semantically and phonologically related, M1 = Mediated; Phonologically related to a category coordinate, M2 = Mediated; category coordinate of a phonologically related item.

Connections

Three types of connections are implemented in the model:

Direct links between corresponding concepts, lexical units, and phonological word forms

These “vertical” links (labeled w_global) are strongly excitatory and allow the mapping of an activated conceptual representation onto its corresponding phonological code. These weights have to be set to a fairly high value, representing the fact that lexical
access usually occurs without errors. The model adheres to the "globality assumption" (see Dell et al., 1996) in that all weights of this type are set to the same value.

**Semantic similarity links**

These weak excitatory links between overlapping concepts (labeled $w_{cc}$) represent the graded nature of conceptual representations and implement spreading activation between related concepts. As shown above, this type of representation, at least for the present purposes, is equivalent to a semantic feature representation in which related concepts share a certain number of conceptual features.

**Phonological similarity links**

These weak excitatory links between overlapping word forms (labeled $w_{ff}$) represent shared phonological segments between form-related lexical items.

**General mutual inhibition at the lexical level**

By means of the semantic similarity links, lexical nodes that are semantically related to the target will receive some activation. Likewise, lexical nodes that are form-related to the target will be co-activated to some extent. Thus, a set of lexical nodes will be activated that share characteristics with the target node. Inhibitory links between nodes at the lexical layer (labeled $w_{ll}$, but omitted in Figure 16) represent a common competitive mechanism whose function it is to suppress this secondary activation and enhance the target activation. Furthermore, they provide a plausible means of accounting for interference from concurrently active distractor words, as outlined above. Note, however, that these inhibitory links are indiscriminatory in that they do not consider semantic of phonological similarity. That is, each word inhibits the activity of each other
word, regardless of whether or not the corresponding nodes are related. This type of inhibition contrasts with the possibility of "smart" or similarity-based type of inhibition examined in Dell and O'Seghdha (1994). In a similarity-based inhibitory structure, only units that are possible competitors tend to inhibit each other. For instance, in the current model, activation of a distractor node "cat" should result in more competition for the selection of the target node "dog" than would activation of an unrelated distractor node such as "hip". Such a similarity-based inhibitory structure could quite plausibly be acquired in an error-correcting learning procedure. However, Dell and O'Seaghdha reviewed the existing findings on inhibitory effects in comprehension and production and concluded that there is little evidence supporting such a structure. In contrast, general inhibition is a common feature of computational models (such as the Interactive Activation model of word recognition introduced by McClelland & Rumelhart, 1981) and receives at least some empirical support.

**Updating function**

Each node in the network carries a real-valued activation with a minimum activation of zero. At each time step, each node is updated in a linear fashion according to the function:

\[
a_{t,j} = a_{t-1,j}(1-q) + \sum_i w_{i,j} a_{t-1,i}
\]

(2)
where $a_{(i,t)}$ is the activation of node $i$ at time step $t$, $q$ is a decay parameter, $w_{(i,j)}$ is the connection weight from a sending node $h$ to the receiving node $i$, and $a_{(i,t-1)}$ is the activation of node $h$ at the preceding time step. In this way, a node's activation from the preceding iteration decays, and a weighted summation of the activation of all the connected nodes is added.

**Lexical selection**

Dell's (1986; Dell et al., 1996) model of speaking, which serves as a blueprint for the model outlined here, predominantly targets speech errors which come about due to random noise in the network, combined with a manipulation of global weight or decay settings. In contrast, the temporal order in which events occur in the network is fixed: after input has been provided to the model, it is allowed to cycle for a number of iterations, after which the lemma node with the highest activation is selected and provided with some further activation. After a further fixed number of iterations, the simulation is stopped, and the phonemes with the highest activation are selected. This structure allows the simulation of correct responses as well as a large variety of naturally occurring speech errors (unrelated, semantic, formal, and mixed errors as well as neologisms).

However, in normal speaking, which the present model intends to simulate, errors in picture naming are negligible: in Dell et al.'s study, non-brain-damaged controls named 96.9% of a large set of pictures correctly. For this reason, the noise-adding procedures of Dell's network are omitted in the present model. Some procedure is needed to convert performance of the model into reaction times. In the present model, this occurs by
making lexical selection time-dependent: the simulation is terminated when the phonological target node reaches a critical activation threshold. In this way, the model provides a measure that can be considered an analog of response times in an experimental situation.

**Processing Assumptions**

Target processing is simulated by setting the conceptual target node to the fixed value of 1.0 at a particular time step, after which its activation is held constant at that level. This activation is then permitted to cycle through the network and spreads to linked nodes. The simulation stops when the activation value of the target phonological node has reached a value of 1.0; the number of iterations required to reach that threshold provides an analog of the response time means obtained in picture naming procedures.

Distractor processing is simulated by setting the activation of the corresponding lexical and phonological units to 1.0 for only one time step. This “pulse” of distractor activation is chosen such that it by itself would never cause the criterion at either the lexical or the phonological level to be exceeded, and in the absence of other sources of activation it will fade away rapidly.

**Parameter Settings**

The particular parameter setting for the following simulations was constrained by a number of observations: first, the speed at which activation spreads across a network is generally determined by a combination of the weight and the decay parameter \( w_{\text{global}} \) and \( decay \), respectively). These two values had to be chosen such that reasonably high activation was obtained at the phonological level. Second, distractor activation is
provided to both the lexical and the phonological level. Since both corresponding nodes are always interconnected, a "reverbratory" pattern develops in which both nodes co-activate each other and activation easily grows out of bounds. To prevent the model from doing so, the decay parameter must be set to a fairly high value such that the distractor "pulse" provided to the model quickly fades away. Third, the two parameters that determine semantic and phonological overlap (w_cc and w_ff, respectively), are constrained in two ways: on the one hand, related nodes must be activated to such an extent that presentation of a related distractor shows a detectable increase or decrease in the number of iterations required to finish the simulations. In other words, if these parameters are set to very small values, unrelated and related conditions will yield the same results. On the other hand, if these parameters are set to values that are too high, the co-activation combined with the distractor activation will keep the target node from ever reaching its selection criterion. Thus, selection of these parameters required a balance between maximizing the effect size and minimizing the likelihood that the model fails to reach its target criterion. In the following simulations, the semantic overlap parameter was first optimized, and the phonological overlap parameter was then chosen such as to yield an effect size roughly comparable to the semantic effect. Table 5 displays the parameter settings for the model obtained in this way.

Results

Figure 17 shows the spread of activation across the network in the absence of distracting activation. Given that weight and decay parameters are the same across the
Table 5  
Parameter Settings in the Computational Model

\[
\begin{align*}
  w_{\text{global}} &= 0.26 \\
  w_{\text{cc}} &= 0.185 \\
  w_{\text{ll}} &= -0.45 \\
  w_{\text{ff}} &= 0.12 \\
  \text{decay} &= 0.45
\end{align*}
\]

whole network, the resulting level of activation at the phonological level will be lower than at the lexical level. The simulation is terminated when phonological activation of the target node reaches the criterion of 1.0, which is the case at time step 17.

Figure 18 displays the activation of the distractor node. It is shown that the activation reaches a brief peak, but decays to zero within six iterations. This pattern results from the relatively large setting for the global decay parameter (\(\text{decay} = 0.45\))

![Activation Graph](image_url)

**Figure 17.** Activation of lexical and phonological target nodes as a function of time step. Onset occurs at time step = 0.
which lowers existing activation quickly in the absence of external input. In this way, a “pulse” of distracting activation can be simulated that models the presumed effect of distractors with a limited presentation duration.

Figure 19 displays the relative time course of semantic and phonological effects. It is shown that a semantic interference effect builds up quickly and peaks at an SOA of 0 time steps, then slowly drifts toward zero which is reached at a time step of 25. In contrast, the phonological facilitation effect builds up more slowly and reaches a peak at time step 15, then relatively quickly goes back to zero. In combination, the model displays a rough sequence of semantic and phonological effects that to some degree resembles the one obtained in the above experiments (cf. Figures 4, 6, 7, and 9).

Next, the effects of distractors that are semantically and phonologically related to

![Graph showing distractor activation over time steps.](image)

**Figure 18.** Activation of distractor lexical node in the unrelated condition.
Figure 19. Time course of semantic and phonological effects (unrelated minus related condition).

the target are examined. Figure 20 shows the results for this condition. The curve shows a slight interference effect that peaks at the same time step as the semantic interference effect (time step = 0), then quickly turns into a facilitatory effect, and at time step 15 equals the curve for the phonologically related condition. This pattern shares at least some resemblance to the results from Experiment 3: at the earlier SOAs, the amount of interference obtained in the critical condition underestimates the amount that would be predicted from an additive relation between semantic and phonological factors. In contrast, at the latest investigated SOA (+150 ms), the effects of the phonological and the semantic and phonological condition were virtually identical, as is the case in the model at the later SOAs (e.g., time step 15 to 25).
Figure 20. Effect (unrelated minus related) of semantically and phonologically related distractors in comparison to the effects of semantic distractors and phonological distractors.

The fit between the model's performance and the empirical data is largely compromised by the fact that the model, although displaying some effect of its interactive characteristics, still does not fully capture the fact that semantically and phonologically related distractors in Experiment 3 essentially acted like phonologically related distractors. In particular, at SOA = 0 ms, semantic interference in this condition was not only reduced, but entirely eliminated. There is good reason to assume that no interactive model could exactly simulate this finding. However, the critical empirical finding - the fact that an additive account overestimates the extent of semantic interference in the critical condition, particularly at earlier SOAs - is captured to some extent by the model,
which puts it at a distinct advantage relative to a modular account of speaking which predicts entirely additive results.

Finally, the effects of the two types of mediated distractors are examined in the model. Figure 21 shows the results for this simulation. Neither of the two mediated conditions have any detectable impact on the number of time steps that it takes the model to reach criterion. Such a finding is in line with the failure to obtain mediated effects in the picture-word interference procedure (cf. Figure 9).

In this model, a competitive structure is set up by lateral inhibition at the lexical level. Consequently, special attention should be paid to the activation dynamics at this

Figure 21. Effect (unrelated minus related) of mediated distractors in comparison to the effects of semantic distractors and phonological distractors. Mediated I: Distractor is form-related to a category coordinate of the picture; Mediated II: distractor is a category coordinate of a word form-related to the picture.
level. Figure 22 shows distractor activation for all relevant conditions at SOA = 0 time steps. First, the phonologically related distractor node has an identical activation curve to the unrelated one, suggesting that the facilitatory effect of 3 time steps obtained with the model resides exclusively at the phonological level and is not mediated via the lexical units monitored here. In contrast, the semantically related distractor node receives input both from its conceptual overlap with the target and the external input provided to a distractor, and as a result, it remains active much longer and thus is able to interfere with target selection. Of most interest is the curve for the semantically and phonologically related condition. Activation for this type of distractor node, although much above the level of an unrelated word, is lower than the corresponding semantic distractor. This finding suggests that, although phonological overlap between distractor and picture alone has no impact on lexical target processing (as evidenced by the phonological curve which is identical to the unrelated one), such overlap dampens the amount of semantic interference created by a simultaneously present conceptual overlap with the target (demonstrated by the fact that this curve is substantially flatter than the curve for the semantic condition). In this way, the model offers a possible mechanism for why semantic and phonological relatedness do not act in an additive fashion. Finally, the curves for the two mediated conditions suggest that such indirect relationships are too weak to have a detectable influence on lexical distractor activation: the level of activation is identical to an unrelated distractor node.
Figure 22. Activation of lexical distractor nodes under various corresponding conditions with an SOA of 0 time steps.

Summary

This model contributes a number of insights to the debate over whether lexical access in speaking is modular or interactive in nature. Proponents of the modular view have claimed (e.g., Levelt et al., 1991) that a) an interactive model cannot yield a sequential pattern of semantic and phonological effects, and b) that an interactive model renders mediated effects, which - at least to date - have not been demonstrated in empirical findings. These assertions have been questioned by Dell and O'Seaghdha (1991) on a theoretical basis. The model outlined here is fully interactive and yet a) shows a clear sequential pattern of semantic and phonological results, and b) does not reveal any influence of mediated distractors. Consequently, it would be premature to
interpret the existing empirical findings on the time course of semantic and phonological
effects and the effects of mediated conditions as definitive evidence against the
interactive account of speaking. The model demonstrates that these findings do not
constitute an insurmountable obstacle for an interactive account.

Furthermore, the model offers some suggestions as to how distractors that are
both semantically and phonologically related to the picture can have an effect that violates
the additive predictions derived from the modular account. A competitive structure at the
lexical level might not be substantially influenced by feedback from the phonological
level in the case of phonologically related distractors, but such feedback might be strong
enough to modify the extent of semantic interference created by a simultaneously present
conceptual relationship between distractor and target.

In a number of ways, the model presented here is of very preliminary nature.
Most importantly, the model merely describes the patterns of results, but so far no attempt
has been made to mathematically fit the performance of the model to the response time
means. The units of the model's performance are somewhat arbitrary and need to be fit to
chronometric units in a mathematical way. A particular problem would be encountered in
fitting the curves of the model to the means derived from the empirically investigated
SOAs. The simulations described in Figures 19, 20, and 21 are each based on the
modeling of twelve SOAs. In contrast, Experiments 3 and 4 each employed only three
SOAs. How these SOAs should be matched to each other is not clear at present. One
possible strategy would be to relate the peak of semantic interference in the model (time
step = 0) to the earliest empirical SOA (-150 ms), the peak of the phonological facilitation
effect (time step = 15) to the latest SOA (+150 ms), and to examine the performance of
the model in one further time step that is in between the two others (time step = 7 or 8;
SOA = 0 ms). Consequently, the data points derived from the model could be compared
to the empirically investigated response time means; for instance, a “goodness of fit”
statistic could be computed between the simulated and the real data. Of course, even
such a procedure is somewhat arbitrary in that we cannot be sure that the empirically
investigated SOAs really captured the maximum of semantic and phonological effects, as
must be assumed here. In summary, there are some complex problems involved in
comparing the model’s performance to the results from the experiments. The goal of the
present work was to provide some suggestions as to whether certain empirical
constellations can be accounted for in an interactive model. At this point, an extensive
investigation of whether the model can exactly fit the data will be relegated to future
research.

A further limitation of the model consists of the fact that is has been set up with
the minimum number of nodes required to model all relevant conditions. To generalize
its performance, it would appear necessary to set up a larger network that has a more
realistic set of lexical items. If it could be shown that such a large network yields
equivalent results, then the conclusions derived from the obtained patterns would receive
much stronger support. By itself, the strategy of starting with a minimal constellation and
then later expanding its structure is legitimate and, for instance, used in Roelofs’ (1992a,
1992b, 1997) and Dell’s (1986; Dell et al., 1996) modeling efforts. However, such an
expansion requires a considerable amount of computational effort which is beyond the
scope of the present investigation. Again, this endeavor will be addressed in future research.

GENERAL DISCUSSION

The question of how semantic and phonological representations relate to each other has been a long-standing issue in research on speech production. The research reported here made extensive use of the picture-word interference procedure to investigate whether lexical access in word production is modular or interactive in nature. Experiments 1 and 2 served to clarify this procedure by providing an explanation for a striking discrepancy between the pattern of semantic and phonological effects caused by either visually or auditorily presented distractor words. When visual distractors were presented with a fixed display duration, both versions yielded a pattern of early semantic interference, followed by later phonological facilitation. These findings point out the importance of such a sequence as a central fact about lexical access in speech production.

The subsequent two experiments employed complex forms of relatedness between picture and distractor word to test predictions derived from either account of speech production. Experiment 3 investigated the effects of distractor words that were both semantically and phonologically related, and demonstrated a clear non-additive influence of the experimental factors of semantic and phonological relatedness. This finding is in accordance with findings from speech error research indicating an occurrence of "mixed" speech errors at a rate above chance, and suggests that one of the critical assumptions of the discrete two-step model of speaking, namely the claim that activation proceeds
unidirectionally from semantic to phonological representations, might have to be abandoned. Within an interactive framework, the results from Experiment 3 might be explained by assuming that both factors modify each other via feedback from the phonological onto the semantic-syntactic level.

Experiment 4 investigated potential effects of "mediated" distractors that are related to the picture via an intervening item. In order to have an effect, such indirect relationships would require an interplay between semantic and phonological levels, which is permitted in an interactive, but not in a modular framework. Consequently, the potential demonstration of a mediated effect might be used as further support for the interactive view. The results from this experiment indicated that these distractors failed to show an effect on response times, a finding which might be interpreted as favoring a modular account. However, it was argued that the absence of mediated effects might also be compatible with an interactive view of speaking as such effects might so small as to be virtually undetectable.

In the final section, a computational model was introduced that is closely related to Dell's (1986; Dell et al., 1996) interactive account of speaking. In line with the results from the experiments, the model demonstrated a somewhat sequential pattern of semantic and phonological effects, reduced semantic interference for distractors that were both semantically and phonologically related to the target, and no effects of mediated relationships. These findings show that, contrary to claims forwarded by proponents of the modular position, an interactive account is not generally incompatible with (i) a sequential pattern of semantic and phonological effects, and (ii) the failure to obtain
mediated effects in experimental studies. Furthermore, the model provides some
suggestions as to how an interactive model might account for the interaction between
semantic and phonological relatedness obtained in Experiment 3, a finding that appears
principally incompatible with a modular account of speaking.

In combination, the experiments as well as the computational model favor a view
of speaking in which semantic and phonological representations appear closely
interrelated. In contrast, modular accounts of speech production seem to face serious
problems in accounting for the empirical findings, particularly for the interaction between
semantic and phonological relatedness obtained in Experiment 3. Of course, in order to
settle the debate between modular and interactive accounts of speech production, the
present findings would have to be confirmed in other experimental procedures. To this
aim, it is encouraging that a number of tasks other than the picture-word interference
procedure have recently been forwarded that might have the potential to further clarify
these issues. For instance, the picture naming/auditory lexical decision task introduced by
Levelt et al. (1991), described in the introduction to Experiment 4, could be employed
with a factorial crossing between semantic and phonological relatedness. If, similar to
Experiment 3 reported above, a statistical interaction was obtained, the conclusions
drawn from Experiment 3 which favored an interactive account of speaking would be
corroborated.
References


Footnotes

1 Other experiments conducted in our laboratory have shown comparable results in that with auditory distractors, semantic and phonological effects always appear to overlap at SOAs near 0 ms. It is not clear at present whether higher items and/or participants variability in our experiment could account for the difference to Schriefers et al.’s original findings. However, we do not believe that such an overlap should be interpreted as evidence against the discrete two-step view favored by these authors.

2 Adapted from Lesch and Pollatsek’s (1993) visual similarity estimate, this phonological similarity measure consists of an index ranging from 0 to 1, calculated as the average of a) the fraction of shared phonemes between two words (e.g., “rabbit” and “rat”) in and out of position (e.g., / r æ b î t / and / r æ t / = 3/4 = 0.75), and b) the fraction of shared phonemes that occur in the same position within the words (e.g., / r æ b î t / and / r æ t / = 2/4 = 0.5). Thus, the phonological similarity measure for “rabbit” and “rat” is (0.75 + 0.5) / 2 = 0.625.

3 For this and the following three simulations, the particular parameter settings are of no relevance and influence the size, but not the general character, of the model performance. Consequently, units of activation or selection probability have been omitted from Figure 11, 12, 13, and 15.

4 Following McClelland and Rumelhart’s (1981) suggestions, selection probability is commonly implemented as a “running average” that takes into account not only a
node's present activation, but also its scaled activation in preceding time steps, and is thus less affected by sudden changes in activation.
### Appendix A

**Stimuli used in Experiments 1 and 2**

<table>
<thead>
<tr>
<th>Picture Name</th>
<th>Semantically Related Distractor</th>
<th>Phonologically Related Distractor</th>
<th>Unrelated Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>apple</td>
<td>peach</td>
<td>apparel</td>
<td>nickel</td>
</tr>
<tr>
<td>banana</td>
<td>plum</td>
<td>band</td>
<td>deer</td>
</tr>
<tr>
<td>bed</td>
<td>couch</td>
<td>bend</td>
<td>duke</td>
</tr>
<tr>
<td>bell</td>
<td>gong</td>
<td>belt</td>
<td>fate</td>
</tr>
<tr>
<td>book</td>
<td>journal</td>
<td>boom</td>
<td>horde</td>
</tr>
<tr>
<td>bowl</td>
<td>cup</td>
<td>bone</td>
<td>sand</td>
</tr>
<tr>
<td>camel</td>
<td>giraffe</td>
<td>camera</td>
<td>tide</td>
</tr>
<tr>
<td>dog</td>
<td>fox</td>
<td>doll</td>
<td>rank</td>
</tr>
<tr>
<td>dress</td>
<td>shirt</td>
<td>dread</td>
<td>tool</td>
</tr>
<tr>
<td>duck</td>
<td>swan</td>
<td>duct</td>
<td>tower</td>
</tr>
<tr>
<td>foot</td>
<td>knee</td>
<td>fool</td>
<td>uncle</td>
</tr>
<tr>
<td>hand</td>
<td>elbow</td>
<td>hash</td>
<td>louse</td>
</tr>
<tr>
<td>horse</td>
<td>zebra</td>
<td>horn</td>
<td>germ</td>
</tr>
<tr>
<td>mouse</td>
<td>skunk</td>
<td>mount</td>
<td>hunch</td>
</tr>
<tr>
<td>mushroom</td>
<td>carrot</td>
<td>muscle</td>
<td>jewel</td>
</tr>
<tr>
<td>onion</td>
<td>radish</td>
<td>onset</td>
<td>mist</td>
</tr>
<tr>
<td>pear</td>
<td>orange</td>
<td>pedal</td>
<td>organ</td>
</tr>
<tr>
<td>pig</td>
<td>cow</td>
<td>pin</td>
<td>bulk</td>
</tr>
<tr>
<td>ring</td>
<td>brooch</td>
<td>ridge</td>
<td>lock</td>
</tr>
<tr>
<td>saw</td>
<td>axe</td>
<td>salt</td>
<td>mask</td>
</tr>
<tr>
<td>seal</td>
<td>whale</td>
<td>seed</td>
<td>threat</td>
</tr>
<tr>
<td>sheep</td>
<td>goat</td>
<td>shield</td>
<td>acid</td>
</tr>
<tr>
<td>sock</td>
<td>glove</td>
<td>sorrow</td>
<td>mail</td>
</tr>
<tr>
<td>sun</td>
<td>mars</td>
<td>suck</td>
<td>puppy</td>
</tr>
<tr>
<td>sweater</td>
<td>blouse</td>
<td>sweep</td>
<td>needle</td>
</tr>
<tr>
<td>table</td>
<td>desk</td>
<td>tail</td>
<td>gutter</td>
</tr>
<tr>
<td>tomato</td>
<td>pepper</td>
<td>tornado</td>
<td>baron</td>
</tr>
<tr>
<td>truck</td>
<td>bus</td>
<td>trunk</td>
<td>scout</td>
</tr>
<tr>
<td>Picture Name</td>
<td>Semantically Related Distractor</td>
<td>Phonologically Related Distractor</td>
<td>Semantically and Phonologically Related Distractor</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>apple</td>
<td>peach</td>
<td>apathy</td>
<td>apricot</td>
</tr>
<tr>
<td>basket</td>
<td>crib</td>
<td>ban</td>
<td>bag</td>
</tr>
<tr>
<td>bee</td>
<td>spider</td>
<td>beacon</td>
<td>beetle</td>
</tr>
<tr>
<td>bread</td>
<td>donut</td>
<td>brick</td>
<td>bran</td>
</tr>
<tr>
<td>camel</td>
<td>pig</td>
<td>cash</td>
<td>calf</td>
</tr>
<tr>
<td>carrot</td>
<td>spinach</td>
<td>cast</td>
<td>cabbage</td>
</tr>
<tr>
<td>duck</td>
<td>raven</td>
<td>dub</td>
<td>dove</td>
</tr>
<tr>
<td>elephant</td>
<td>moose</td>
<td>elm</td>
<td>elk</td>
</tr>
<tr>
<td>fly</td>
<td>moth</td>
<td>flu</td>
<td>flea</td>
</tr>
<tr>
<td>lamp</td>
<td>candle</td>
<td>landing</td>
<td>lantern</td>
</tr>
<tr>
<td>peanut</td>
<td>almond</td>
<td>piano</td>
<td>pecan</td>
</tr>
<tr>
<td>rabbit</td>
<td>beaver</td>
<td>raft</td>
<td>rat</td>
</tr>
<tr>
<td>snake</td>
<td>eel</td>
<td>snack</td>
<td>snail</td>
</tr>
<tr>
<td>spoon</td>
<td>ladle</td>
<td>sparkle</td>
<td>spatula</td>
</tr>
<tr>
<td>squirrel</td>
<td>mole</td>
<td>skate</td>
<td>skunk</td>
</tr>
<tr>
<td>train</td>
<td>bus</td>
<td>trophy</td>
<td>trolley</td>
</tr>
<tr>
<td>truck</td>
<td>jeep</td>
<td>trap</td>
<td>tractor</td>
</tr>
<tr>
<td>trumpet</td>
<td>horn</td>
<td>traffic</td>
<td>trombone</td>
</tr>
</tbody>
</table>
### Appendix C

**Stimuli used in the Pilot Experiment**

<table>
<thead>
<tr>
<th>Picture Name</th>
<th>Semantically Related Distractor</th>
<th>Mediated Distractor</th>
<th>Unrelated Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>apple</td>
<td>tangerine</td>
<td>tambourine</td>
<td>sign</td>
</tr>
<tr>
<td>barrel</td>
<td>jug</td>
<td>jam</td>
<td>thyme</td>
</tr>
<tr>
<td>basket</td>
<td>case</td>
<td>cave</td>
<td>scene</td>
</tr>
<tr>
<td>book</td>
<td>magazine</td>
<td>mandarin</td>
<td>plant</td>
</tr>
<tr>
<td>bread</td>
<td>muffin</td>
<td>mutton</td>
<td>race</td>
</tr>
<tr>
<td>broom</td>
<td>mop</td>
<td>map</td>
<td>rope</td>
</tr>
<tr>
<td>butterfly</td>
<td>moth</td>
<td>mouth</td>
<td>angle</td>
</tr>
<tr>
<td>cake</td>
<td>pie</td>
<td>pin</td>
<td>frame</td>
</tr>
<tr>
<td>carrot</td>
<td>squash</td>
<td>square</td>
<td>dance</td>
</tr>
<tr>
<td>church</td>
<td>temple</td>
<td>tenant</td>
<td>spot</td>
</tr>
<tr>
<td>clown</td>
<td>jester</td>
<td>justice</td>
<td>device</td>
</tr>
<tr>
<td>foot</td>
<td>claw</td>
<td>clash</td>
<td>squirrel</td>
</tr>
<tr>
<td>hammer</td>
<td>chisel</td>
<td>cheddar</td>
<td>dog</td>
</tr>
<tr>
<td>harp</td>
<td>piano</td>
<td>piranha</td>
<td>wolf</td>
</tr>
<tr>
<td>house</td>
<td>barn</td>
<td>beer</td>
<td>wound</td>
</tr>
<tr>
<td>lemon</td>
<td>grapefruit</td>
<td>graphite</td>
<td>seal</td>
</tr>
<tr>
<td>lion</td>
<td>panther</td>
<td>pantry</td>
<td>dawn</td>
</tr>
<tr>
<td>lock</td>
<td>bolt</td>
<td>build</td>
<td>dolphin</td>
</tr>
<tr>
<td>peanut</td>
<td>cashew</td>
<td>canvas</td>
<td>armadillo</td>
</tr>
<tr>
<td>pear</td>
<td>orange</td>
<td>order</td>
<td>coat</td>
</tr>
<tr>
<td>potato</td>
<td>rhubarb</td>
<td>rubbish</td>
<td>club</td>
</tr>
<tr>
<td>rabbit</td>
<td>hare</td>
<td>hail</td>
<td>student</td>
</tr>
<tr>
<td>snail</td>
<td>worm</td>
<td>war</td>
<td>needle</td>
</tr>
<tr>
<td>sun</td>
<td>moon</td>
<td>mold</td>
<td>brain</td>
</tr>
<tr>
<td>tomato</td>
<td>beet</td>
<td>boot</td>
<td>railroad</td>
</tr>
<tr>
<td>train</td>
<td>ship</td>
<td>shop</td>
<td>forest</td>
</tr>
<tr>
<td>elephant</td>
<td>rhinoceros</td>
<td>regatta</td>
<td>wine</td>
</tr>
<tr>
<td>swan</td>
<td>gull</td>
<td>game</td>
<td>note</td>
</tr>
</tbody>
</table>
## Appendix D

**Stimuli used in Experiment 4**

<table>
<thead>
<tr>
<th>Picture Name</th>
<th>Semantically Related Distractor</th>
<th>Mediated I Distractor</th>
<th>Phonologically Related Distractor</th>
<th>Mediated II Distractor</th>
<th>Unrelated Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>ant</td>
<td>wasp</td>
<td>wand</td>
<td>anchor</td>
<td>hook</td>
<td>bulk</td>
</tr>
<tr>
<td>bell</td>
<td>gong</td>
<td>golf</td>
<td>belt</td>
<td>tie</td>
<td>party</td>
</tr>
<tr>
<td>bowl</td>
<td>pot</td>
<td>posture</td>
<td>boat</td>
<td>ship</td>
<td>letter</td>
</tr>
<tr>
<td>brush</td>
<td>comb</td>
<td>cone</td>
<td>bread</td>
<td>muffin</td>
<td>mole</td>
</tr>
<tr>
<td>cannon</td>
<td>rifle</td>
<td>rider</td>
<td>camel</td>
<td>giraffe</td>
<td>chop</td>
</tr>
<tr>
<td>carrot</td>
<td>spinach</td>
<td>spindle</td>
<td>cashew</td>
<td>almond</td>
<td>brass</td>
</tr>
<tr>
<td>clock</td>
<td>watch</td>
<td>wash</td>
<td>clown</td>
<td>jester</td>
<td>batter</td>
</tr>
<tr>
<td>cow</td>
<td>donkey</td>
<td>dollar</td>
<td>couch</td>
<td>sofa</td>
<td>tank</td>
</tr>
<tr>
<td>flower</td>
<td>bush</td>
<td>bull</td>
<td>flag</td>
<td>banner</td>
<td>swim</td>
</tr>
<tr>
<td>fly</td>
<td>beetle</td>
<td>beer</td>
<td>flute</td>
<td>horn</td>
<td>stamp</td>
</tr>
<tr>
<td>glass</td>
<td>mug</td>
<td>mustard</td>
<td>glove</td>
<td>mitten</td>
<td>van</td>
</tr>
<tr>
<td>goat</td>
<td>sheep</td>
<td>shield</td>
<td>gold</td>
<td>silver</td>
<td>supper</td>
</tr>
<tr>
<td>hammer</td>
<td>chisel</td>
<td>chicken</td>
<td>hand</td>
<td>arm</td>
<td>cabin</td>
</tr>
<tr>
<td>house</td>
<td>barn</td>
<td>bark</td>
<td>hound</td>
<td>wolf</td>
<td>blood</td>
</tr>
<tr>
<td>leg</td>
<td>elbow</td>
<td>elk</td>
<td>leopard</td>
<td>panther</td>
<td>trigger</td>
</tr>
<tr>
<td>lion</td>
<td>tiger</td>
<td>tide</td>
<td>lime</td>
<td>orange</td>
<td>rocket</td>
</tr>
<tr>
<td>mouse</td>
<td>hamster</td>
<td>hanger</td>
<td>mountain</td>
<td>hill</td>
<td>cent</td>
</tr>
<tr>
<td>pants</td>
<td>vest</td>
<td>vent</td>
<td>parrot</td>
<td>canary</td>
<td>tail</td>
</tr>
<tr>
<td>pig</td>
<td>mule</td>
<td>mural</td>
<td>pitcher</td>
<td>bottle</td>
<td>snow</td>
</tr>
<tr>
<td>rabbit</td>
<td>beaver</td>
<td>beacon</td>
<td>radish</td>
<td>celery</td>
<td>club</td>
</tr>
<tr>
<td>skirt</td>
<td>dress</td>
<td>drum</td>
<td>skunk</td>
<td>raccoon</td>
<td>bumper</td>
</tr>
<tr>
<td>spider</td>
<td>hornet</td>
<td>horde</td>
<td>spoon</td>
<td>fork</td>
<td>asphalt</td>
</tr>
<tr>
<td>stool</td>
<td>chair</td>
<td>chain</td>
<td>stork</td>
<td>eagle</td>
<td>thumb</td>
</tr>
<tr>
<td>trumpet</td>
<td>clarinet</td>
<td>clamp</td>
<td>train</td>
<td>bus</td>
<td>author</td>
</tr>
</tbody>
</table>