Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 5, 900–913

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.5.900

Influences of Spoken Word Planning on Speech Recognition

Ardi Roelofs
Radboud University Nijmegen

Rebecca Özdemir and Willem J. M. Levelt
Max Planck Institute for Psycholinguistics

In 4 chronometric experiments, influences of spoken word planning on speech recognition were examined. Participants were shown pictures while hearing a tone or a spoken word presented shortly after picture onset. When a spoken word was presented, participants indicated whether it contained a prespecified phoneme. When the tone was presented, they indicated whether the picture name contained the phoneme (Experiment 1) or they named the picture (Experiment 2). Phoneme monitoring latencies for the spoken words were shorter when the picture name contained the prespecified phoneme compared with when it did not. Priming of phoneme monitoring was also obtained when the phoneme was part of spoken nonwords (Experiment 3). However, no priming of phoneme monitoring was obtained when the pictures required no response in the experiment, regardless of monitoring latency (Experiment 4). These results provide evidence that an internal phonological pathway runs from spoken word planning to speech recognition and that active phonological encoding is a precondition for engaging the pathway.

Keywords: speech perception, speech production, self-monitoring

In psycholinguistics, speech production and speech comprehension are often studied as if they were separate and independent processes. However, in everyday communication, speech production and recognition often seem to happen simultaneously. Speakers regularly seem to plan their turn in a conversation while simultaneously listening to an interlocutor. Moreover, speakers seem to monitor their own speech for errors while simultaneously planning an upcoming utterance (e.g., Levelt, 1989). This raises the question of how ongoing speech recognition processes influence the planning of speech and vice versa. In the present article, we examine the issue of how production affects recognition for the phonological level of planning and processing. Influences of speech production on speech recognition and vice versa should occur if phonological representations and processes are shared between production and recognition, but also when the production and perception systems are separate but tightly linked. We start by discussing some of the scarce evidence about the influences of production on perception and about whether the phonological systems are shared or separate. The evidence comes from chronometric, dual-task accuracy, neuropsychological, and functional brain imaging studies. Next, we discuss some of the functional reasons for the existence of links running from production to perception in separate but closely linked systems, and we describe our working model WEAVER++, which assumes separate but connected phonological systems. Then, we provide an overview of the new experiments reported in the present article, which are reported next. Finally, we discuss the theoretical implications of our findings in the General Discussion section.

Evidence for Cross-Talk Between Planning and Recognition

Chronometric evidence from object naming studies suggests that hearing spoken words influences spoken word planning (e.g., A. S. Meyer & Schriefers, 1991; Roelofs, 1997, 2002, 2005; Schriefers, Meyer, & Levelt, 1990; Starreveld, 2000). For example, Schriefers et al. (1990) observed that when participants named pictured objects (e.g., a pictured cat), the time it took to name the objects was less when they simultaneously heard a phonologically related spoken word (e.g., "cap") compared with an unrelated word (e.g., "tree"). Moreover, experiments have shown that both spoken words and their initial fragments speed up object naming (Özdemir, 2006; Roelofs, 1997, 2002; Starreveld, 2000). The priming effect from hearing words and word fragments on object naming suggests the existence of phonological links running from speech perception to speech planning, or, alternatively, a sharing of phonological representations between production and perception. D. E. Meyer and Gordon (1983) and Gordon and Meyer (1984) asked participants to produce spoken syllables in response to other spoken or written syllables. For example, they said "buh" to "duh" or "duh" to "tuh." Compared with unrelated syllables, responses were faster when the onset consonants of the heard and produced syllables shared the voicing feature (e.g., voiced "buh"—"duh"), but not when the place of articulation was shared (e.g., alveolar "duh"—"tuh"). For the responses to printed syllables, sharing of phonological features had no effect at all. These findings also indicate that speech perception may influence production and suggest a close relationship between the perception and production of voicing features in speech.

It is far less clear, however, whether there exist phonological influences from speech planning on speech recognition. Chrono-

Ardi Roelofs, Nijmegen Institute for Cognition and Information and F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Nijmegen, the Netherlands; Rebecca Özdemir and Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.

We are indebted to Anne Cutler and James McQueen for helpful comments and discussions.

Correspondence concerning this article should be addressed to Ardi Roelofs, Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, Montessorilaan 3, Nijmegen 6525 HR, the Netherlands. E-mail: A.Roelofs@nici.ru.nl

metric experiments that examined the influence of speech production planning on concurrent speech recognition are rare. Gordon and Meyer (1984) described an unpublished experiment from their laboratory (i.e., D. E. Meyer & Gordon, 1984) in which participants had to fully prepare a specified syllable for production (e.g., "buh"). On half of the trials, a cue presented after preparation indicated to actually produce the prepared syllable. On the other half of the trials, an auditory syllable was presented over headphones (e.g., "duh" or "tuh") and participants had to manually classify the heard syllable as one of four alternatives. It was observed that the manual classification was slowed if the onset consonants of the auditory syllable and the syllable prepared for production shared the voicing feature ("buh"—"duh"). However, no effect was observed when the onset consonants of the heard and prepared syllables shared the place of articulation ("duh"—"tuh"). It is important to note that the observed influence of production on perception (i.e., interference from sharing the voicing feature) is opposite to the observed influence of perception on production (i.e., facilitation). According to Gordon and Meyer (1984) and D. E. Meyer and Gordon (1984), the interference from sharing the voicing feature between the prepared and heard syllables suggests that this phonological feature is common between speech production and speech perception or, alternatively, that production and perception compete for a common clock mechanism for the timing of voice onset in consonants.

Levelt et al. (1991) combined picture naming with auditory lexical decision. Participants were asked to name pictured objects, and on some critical trials, they had to make a lexical decision by means of a keypress to an auditory probe presented shortly after picture onset. Thus, the speakers had to monitor for the lexical status of spoken probes while preparing to speak the name of the object. Compared with unrelated spoken words, monitoring responses were slower for word probes that were semantically related, phonologically related, or even identical to the picture name. These findings suggest influences of speech planning on spoken word recognition.

Cross-talk between production and perception is expected if representations are shared, as proposed by several theorists, including Wernicke (1874), Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967), and MacKay (1987). Theorists have argued for shared representations on the basis of evidence from, inter alia, neuropsychological and functional brain imaging studies. For example, it has been observed that brain-damaged patients with speech comprehension deficits may have fluent but phonemically disordered speech production (e.g., Shallice, 1988). Also, brain imaging studies have shown that the left posterior superior temporal cortex (Wernicke's area) participates in the phonological level of processing in both speech perception and production (for reviews, see Buchsbaum, Hickok, & Humphries 2001; Indefrey & Levelt, 2004).

However, the view of shared representations also meets with a number of difficulties. First, examining performance accuracy (i.e., percentage correct), Shallice, McLeod, and Lewis (1985) observed that there was little dual-task interference in terms of loss of accuracy when participants performed auditory name detection and oral reading tasks simultaneously, suggesting that production and perception tasks may engage separate pathways. Second, under the assumption of shared representations, one expects a strong correlation between production and comprehension accuracy of aphasic speakers. However, such strong correlations are not observed empirically (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Jacquemot, Dupoux, & Bachoud-Lévi, 2007; Nickels & Howard, 1995), although some associations between production and recognition abilities may occur (e.g., N. Martin & Saffran, 2002). Third, on the basis of the available imaging findings, it is difficult to distinguish between overlapping systems and closely linked ones. Actually, some imaging evidence suggests that different parts of Wernicke's area are involved in production and perception (Wise et al., 2001). If phonological input and output networks are separate but tightly connected, then activation of one phonological network would lead to the activation of the other, explaining the co-activation in imaging studies (Roelofs, 2003b).

To summarize, the available evidence for cross-talk between speech production and recognition from chronometric, neuropsychological, and functional brain imaging studies does not necessarily imply shared representations or mechanism, but it is also compatible with the assumption that the phonological production and perception systems are separate but closely linked. In fact, on the basis of reviews of the literature, Monsell (1987) and Roelofs (2003b) concluded that the position of two linked systems best explains the available data. Given that some studies have suggested some influences from planning on recognition, the question arises about why links from production to perception would exist in separate but closely linked systems. Monsell gave three functional reasons for the existence of these links.

Function of Planning-to-Recognition Links

According to Monsell (1987), phonological links from production to recognition may serve a number of functions. In particular, such links would be important for subvocal rehearsal, language learning, and self-monitoring of speech production.

First, an internal pathway from speech production to speech recognition allows for the rehearsal of a planned utterance prior to actual production. A subvocal rehearsal loop is a standard and central component of models of working memory, such as Baddeley's (1986, 2003) phonological loop model. According to this model, phonological short-term memory includes a phonological buffer that can hold memory traces for a few seconds and a subvocal rehearsal process used to refresh the memory traces. Baddeley (2003), R. Martin, Lesch, and Bartha (1999), and Jacquemot and Scott (2006) proposed that phonological short-term memory arises from the recycling of information between two phonological buffers, one involved in speech perception and one in speech production. Page, Madge, Cumming, and Norris (in press) demonstrated that those errors in immediate serial recall (a standard working memory task) that are attributable to phonological similarity resemble phonemic errors in normal speech production. This suggests a close relation between the phonological loop component of working memory and phonological encoding in speech production. Information can only be transferred from production to perception as assumed by the phonological loop model if a pathway exists that links the two systems.

Second, a pathway from speech production to speech recognition may be important in language learning. Monsell (1987) stated that if such a pathway exists, "the phonological products of the output lexicon can be checked against entries in the input lexicon and reinforced, or modified, as necessary, to bring the systems into closer correspondence" (p. 283). In line with this suggestion, Baddeley, Gathercole, and Papagno (1998) argued that the phonological loop of working memory serves as a language learning device. In particular, according to them, the phonological loop plays a crucial role in learning the novel phonological forms of new words. This is achieved by temporarily maintaining unfamiliar sound patterns while more permanent phonological representations are being constructed for the words in memory. Gathercole (1999) argued that the phonological loop also plays a crucial role in learning a foreign language.

Third, a pathway from speech production to speech recognition allows for the self-monitoring of a planned utterance prior to actual production, as assumed by the perceptual loop theory of verbal self-monitoring (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999; Roelofs, 2004). Speakers can monitor their utterances by listening to their own overt speech. Production and perception are linked via the overt speech signal in that situation. The perceptual loop theory also holds that an internal pathway from production to perception is used for self-monitoring. An internal route for self-monitoring of speech implies links running from speech planning to speech recognition. Supporting the perceptual loop theory, functional brain imaging studies have suggested that verbal self-monitoring and speech recognition are served by the same or closely linked neural structures (e.g., McGuire, Silbersweig, & Frith, 1996; Paus, Perry, Zatorre, Worsley, & Evans, 1996). Chronometric evidence that phonological representations mediate between speech planning and perception and that the speech comprehension system is engaged in internal self-monitoring comes from phoneme monitoring studies. Next, we discuss these studies in some depth because the phoneme monitoring task has also been used in the experiments we report.

Evidence From Phoneme Monitoring

Wheeldon and Levelt (1995) provided evidence that phonological representations underlie the self-monitoring of internal speech. Their participants were native Dutch speakers who fluently spoke English. They had to monitor for target phonemes in the Dutch translation equivalent of visually presented English words. For example, they had to indicate by means of a button press (*yes*/*no*) whether the phoneme /n/ is part of the Dutch translation equivalent of the English word *waiter*. The Dutch word is *kelner*, which has /n/ as the onset of the second syllable, thus requiring a positive response. All Dutch target words were disyllabic. There is evidence that phonological word representations are planned from the beginning of a word to its end (e.g., A. S. Meyer & Schriefers, 1991). To examine the time course of phonological encoding, Wheeldon and Levelt manipulated the serial position of the critical phonemes in the Dutch words. The target phoneme could be the onset or coda of the first syllable or the onset or coda of the second syllable. Monitoring latencies increased with the serial position of the phonemes within the word. To experimentally verify whether phonological rather than phonetic representations were monitored, the authors had participants perform the phoneme monitoring task while simultaneously counting aloud, which is known to suppress the maintenance of phonetic representations (cf. Baddeley, 1986, 2003). The monitoring latencies were longer with the counting task, but the serial position effect was replicated. In another experiment, it was observed that internal monitoring is sensitive to a word's syllable structure. These findings suggest that self-monitoring involves a phonological rather than a phonetic representation. Wheeldon and Morgan (2002) replicated the serial position effect in internal phoneme monitoring in English, and Van Turennout, Hagoort, and Brown (1997) and Özdemir, Roelofs, and Levelt (in press) replicated the effect with monitoring for phonemes in picture names.

Özdemir et al. (in press) also provided evidence that internal self-monitoring of speech production planning is achieved via the speech comprehension system. According to the perceptual loop theory, speech perception-specific effects should be obtained on internal self-monitoring. One such perception-specific effect is the uniqueness point effect (e.g., Marslen-Wilson, 1990). The uniqueness point of a word is defined as the phoneme in the word where it diverges from all other words in the language, going from the beginning of the word to its end. The uniqueness point influences the speed of spoken word recognition. For example, Marslen-Wilson (1990) observed that listeners are faster in deciding whether a spoken item is a word (auditory lexical decision) when the uniqueness point is early in a word than when it is late in a word. Moreover, in phoneme monitoring experiments, participants were faster in detecting a target phoneme in a spoken word when the phoneme followed the uniqueness point of the word than when it preceded the uniqueness point (Frauenfelder, Segui, & Dijkstra, 1990). Furthermore, when the target phoneme followed the uniqueness point, phoneme monitoring was faster when the distance of the phoneme to the uniqueness point was long than when it was short (Frauenfelder et al., 1990).

Özdemir et al.'s (in press) study tested for effects of the perceptual uniqueness point of a word in monitoring internal speech. Participants were presented with pictured objects, and they indicated by pressing a button whether the picture name contained a prespecified target phoneme. The critical manipulation concerned the position of the target phonemes relative to the uniqueness point of the picture names. All target phonemes followed the uniqueness point, and the distance could be short or long. According to the perceptual loop theory, monitoring latencies should depend on the distance of the phoneme from the uniqueness point of the picture name. Moreover, effects of uniqueness point should be present in the monitoring of internal speech but not in naming the pictures. The experimental results showed an effect of the perceptual uniqueness point of a word in internal phoneme monitoring in the absence of such an effect in picture naming. Phoneme monitoring latencies were smaller when the distance from the uniqueness point was long than when it was short, just as Frauenfelder et al. (1990) observed for the monitoring of external speech. These results support the perceptual loop theory of self-monitoring.

A Working Model for the Present Experiments

The existence of internal phonological links between speech planning and speech comprehending entails phonological influences of speech planning on the recognition of external speech. In the present article, we report a series of experiments that examined the presence and nature of these influences. As a working model for the present research, we used the WEAVER++ model of spoken word planning and self-monitoring (Levelt et al., 1999;

Roelofs, 1992, 1997, 2003a, 2003b, 2004, 2006), illustrated in Figure 1.

WEAVER++ assumes that word planning is a staged process, moving from conceptual preparation (i.e., the identification of a pictured object in picture naming), via lemma retrieval (recovering the word as syntactic entity, including its syntactic properties, crucial for the use of the word in phrases and sentences), to word-form encoding. Dell (1986) assumed similar planning levels. Word-form encoding includes morphological, phonological, and phonetic encoding. In morphological encoding, a lemma is used to recover the corresponding morphemes and to assemble a morphological representation for the word. In phonological encoding, the phonemes of the morphemes are retrieved and used to construct a phonological word representation. A phonological word representation specifies the syllables and, for polysyllabic words, the stress pattern across syllables. In phonetic encoding, the phonological word representation is used to generate an articulatory program, which makes explicit articulatory tasks such as lip protrusion, lowering of the jaw, and the timing of voice onset by the vocal cords (cf. Gordon & Meyer, 1984).

Comprehending spoken words traverses from word-form perception to lemma retrieval and conceptual identification. Perceived words activate lemmas and word forms in parallel. In the model, concepts and lemmas are shared between production and comprehension, whereas there are separate input and output representations of word forms. Consequently, the flow of information between the conceptual and the lemma level is bidirectional, whereas it is unidirectional between lemmas and forms (top-down for production and bottom-up for comprehension). Internal monitoring of the speech plan involves feeding a rightward incrementally generated phonological word representation into the word-form perception system. The phonological word is sent to the perception system as it becomes available over time. Phoneme decisions in monitoring are based on phonological information activated in the word-form perception system. McClelland and Elman (1986) and Norris, McQueen, and Cutler (2000) presented models of how phoneme monitoring may be achieved by the speech perception system. These models assume explicit representations of phonemes in the recognition system, but this is not essential (Lahiri & Marslen-Wilson, 1991; Norris et al., 2000). The links from word-form encoding to word-form perception and vice versa make up a phonological loop (cf. Baddeley, 1986, 2003; Kieras, Meyer, Mueller, & Seymour, 1999; Page et al., in press).

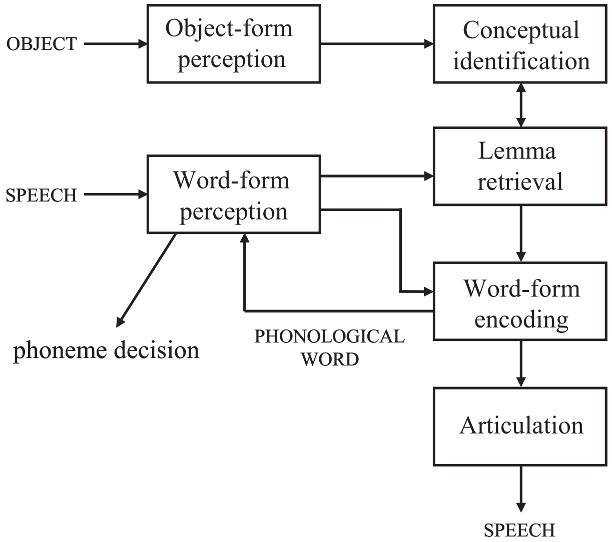

Figure 1. Flow of information in the WEAVER++ model during object naming and spoken word recognition. Phoneme decisions in phoneme monitoring are based on information provided by the word-form perception process.

Overview of the Present Experiments

The existence of an internal perceptual loop in verbal self-monitoring (Levelt, 1983; Levelt et al., 1999) and a phonological loop in working memory (Baddeley, 1986, 2003) entails an internal phonological pathway running from speech planning to speech recognition. If such an internal phonological pathway exists, then phonological influences of speech planning on concurrent speech perception should occur. We report a series of four experiments that tested for these influences. In the experiments, participants were shown pictures and heard a tone or a spoken word presented shortly after picture onset. When a spoken word was presented, participants had to indicate whether it contained a prespecified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme (Experiment 1) or they had to name the picture (Experiments 2–3). We measured the phoneme monitoring latencies for the spoken words when the picture name contained the prespecified phoneme and when it did not, referred to as the match and the nonmatch conditions, respectively. Faster monitoring responses in the match than in the nonmatch condition would suggest facilitation of the recognition of the target phoneme in the spoken word. Such a phonological priming effect would be evidence for the existence of phonological links running from speech planning to speech recognition, affecting the recognition of external speech.

Chronometric evidence in the literature has suggested that phonological activation in picture naming starts around 400 ms after picture onset (e.g., Indefrey & Levelt, 2004; Roelofs, 2007). Thus, a stimulus onset asynchrony (SOA) of 300 ms between picture and spoken word should be too short for obtaining a phonological priming effect, whereas an interval of 600 ms should yield a priming effect. The predicted effect of SOA was tested in Experiment 1. The WEAVER++ model (Levelt et al., 1999) implements the claim that only selected lemmas spread activation to the phonological level. Thus, the model predicts that priming should be obtained only when the picture names need to be phonologically encoded. Phonological priming should occur regardless of whether the response to the picture is phoneme monitoring (tested in Experiment 1) or naming (tested in Experiment 2). However, priming should not be obtained with passive picture viewing (tested in Experiment 4). Moreover, phonological priming should occur regardless of whether the target phoneme is part of a spoken word (tested in Experiments 1 and 2) or a spoken nonword (tested in Experiment 3).

Experiment 1

Experiment 1 examined whether phonemes are preactivated in the speech perception system as a result of phonological encoding in response to a picture. Participants were presented with pictures and a tone or spoken word occurring 300 or 600 ms after picture onset. These SOAs correspond to, respectively, points in time before and after phonological information is activated in the production system. When a spoken word was presented, participants had to indicate whether it contained a prespecified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme. Trial types were randomly intermixed.

Method

Participants. Forty-two native Dutch speakers (37 women; mean age = 20 years) from the participant pool of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands took part in the experiment. We tested 28 participants with an SOA of 300 ms and 14 participants with an SOA of 600 ms (we predicted a null effect at SOA = 300 ms, hence the larger number of participants). Participants all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials and design. Participants had to monitor for the phonemes /p/ and /k/ in existing Dutch spoken words. There were 10 items per phoneme, 5 monosyllabic and 5 disyllabic words. The target phoneme was always in the initial position. Each spoken word was preceded by a picture prime presented either 300 or 600 ms before the onset of the word. There were two critical contexts: match and nonmatch. In the match context, participants saw a picture that also had the target phoneme in the initial position (e.g., picture *peer*, word "paal;" picture *kist*, word "kam"). In the nonmatch context, the picture name did not begin with the target phoneme (e.g., *peer*–"kam," *kist*–"paal") nor did it contain this phoneme in any other position of its name. The pictures were recombined in such a way that the /k/ match pictures were the pictures for the nonmatch context of the target phoneme /p/ and vice versa. The full set of the critical materials can be found in the Appendix.

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

Procedure. Participants were tested individually in a quiet room. They sat in front of a computer screen wearing headphones. They received written instructions to react as quickly and accurately as possible. The participants were instructed to press a response button with their dominant hand if the word began with the target phoneme or, in case they heard the tone, if the picture name began with this phoneme. The participants were familiarized with the pictures and their names before the beginning of the experimental trials. The familiarization happened before the participants received the instructions for the experiment.

The structure of an experimental trial was as follows. Participants saw a picture on the screen for 1,000 ms. The presentation duration of the pictures was relatively long because we wanted to make sure that the participants identified the pictures, especially when a trial required a response to the spoken words (this was particularly important in Experiment 4, which tested for priming effects with passive picture viewing). With a certain SOA (300 ms or 600 ms, depending on the participant group), the participants heard either a word or the tone over headphones. The next trial started 2.0 s after word or tone onset. The experiment was controlled by the Nijmegen Experimental Setup (NESU) software developed at the Max Planck Institute for Psycholinguistics. A push button box with one button was used to register the phoneme monitoring latencies, which were written to hard disk after each trial.

The experiment consisted of two blocks of trials (with /p/ and /k/ as targets), with every word occurring twice per block (preceded by a picture from the match or nonmatch context) and every picture occurring three times per block (once combined with a word containing the target phoneme, once with a word not containing the target phoneme, and once with a tone requesting a monitoring response to the picture name). The order of blocking was counterbalanced across participants. The order of items within a block was randomized. There was a short break between the trial blocks. An experimental session lasted about 30 min.

Analysis. Trials on which participants missed the targets or on which monitoring latencies exceeded 1,500 ms were regarded as errors and were excluded from the analysis of the response latencies. Repeated measures of variance were performed on the latencies of the correct responses and on the error rates. Context and phoneme were tested within participants, and SOA was tested between participants. Furthermore, context and SOA were tested within items, and phoneme was tested between items. In all analyses, an alpha level of .05 was adopted.

Results and Discussion

The mean monitoring latencies, their standard deviations, and the error percentages for each context, phoneme, and SOA are given in Table 1. The table shows that there was a facilitation effect for the phoneme /k/ as well as for the phoneme /p/ at the 600-ms SOA, whereas there was no effect for either of the two phonemes at the 300-ms SOA. The statistical analysis of the phoneme monitoring latencies yielded a main effect of context, $F_1(1, 40) = 12.83$, $p = .001$; $F_2(1, 18) = 8.96$, $p = .008$, but not of phoneme, $F_1(1, 40) < 1$, $p = .71$; $F_2(1, 18) < 1$, $p = .77$, or of SOA, $F_1(1, 40) = 1.24$, $p = .27$; $F_2(1, 18) = 5.70$, $p = .03$. There was no interaction of phoneme and context, $F_1(1, 40) = 1.54$, $p = .22$; $F_2(1, 18) < 1$, $p = .37$. Also, there was no interaction of phoneme, context, and SOA, $F_1(1, 40) < 1$, $p = .44$; $F_2(1, 18) = 1.0$, $p = .33$. However, context and SOA interacted, $F_1(1, 40) = 13.00$, $p = .001$; $F_2(1, 18) = 19.89$, $p = .001$. There was an effect of context for the 600-ms SOA, $F_1(1, 13) = 14.28$, $p = .002$; F_2(1,

Table 1
Mean Phoneme Monitoring Latencies (in Milliseconds), Their Standard Deviations (in Milliseconds), and the Error Rates (Percentages) per Phoneme and Context for Experiments 1–4

	Target phoneme								
	/k/			/p/			Total		
Experiment and context	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%
Experiment 1: 300-ms SOA									
Match	699	177	1.8	704	180	3.2	702	178	2.5
Nonmatch	701	166	3.2	699	198	3.6	700	182	3.4
Experiment 1: 600-ms SOA									
Match	686	205	5.0	711	192	4.3	699	198	4.6
Nonmatch	774	218	8.6	762	201	8.6	768	210	8.6
Experiment 2									
Match	727	181	1.7	751	201	1.7	739	191	1.7
Nonmatch	776	174	7.5	797	214	5.8	787	195	6.7
Experiment 3									
Match	733	183	3.3	762	209	3.3	747	197	3.3
Nonmatch	793	183	5.8	780	211	6.7	787	197	6.3
Experiment 4									
Match	608	189	3.5	644	191	3.0	626	190	3.3
Nonmatch	588	166	0.5	649	163	3.0	618	167	1.8

Note. SOA = stimulus onset asynchrony.

19) = 21.86, $p = .001$, but not for the 300-ms SOA, $F_1(1, 27) < 1$, $p = .98$; $F_2(1, 19) < 1$, $p = .90$.

Table 1 shows that more errors were made in the nonmatch than in the match context and more at the 600-ms SOA than at the 300-ms SOA. The statistical analysis of the errors yielded an effect of context, $F_1(1, 40) = 7.16$, $p = .011$; $F_2(1, 18) = 4.56$, $p = .047$, and of SOA, $F_1(1, 40) = 4.96$, $p = .03$; $F_2(1, 18) = 6.47$, $p = .02$, but not of phoneme, $F_1(1, 40) < 1$, $p = .81$; $F_2(1, 18) < 1$, $p = .78$. There was no interaction of phoneme and context, $F_1(1, 40) < 1$, $p = .93$; $F_2(1, 18) < 1$, $p = .91$, or of SOA and context, $F_1(1,40) = 2.84$, $p = .10$; $F_2(1, 18) = 3.21$, $p = .09$. Also, there was no interaction of phoneme, context, and SOA, $F_1(1, 40) < 1$, $p = .68$; $F_2(1, 18) < 1$, $p = .57$. Given that most errors were made in the conditions with the slowest responses, there is no evidence for a speed–accuracy trade-off in the data.

The stability of the picture context effects on phoneme monitoring observed in the means was checked by examining the latency distributions for the monitoring responses in the match and nonmatch conditions for each SOA (cf. Wheeldon & Morgan, 2002). To obtain the latency distributions, we divided the rank-ordered latencies for each participant into deciles (10% quantiles) and computed mean latencies for each decile, separately for the match and nonmatch contexts and separately for the two SOAs. Given that no effects of phoneme were obtained, the latencies were collapsed across phoneme. By averaging the decile means across participants, so-called Vincentized cumulative distribution functions are obtained (Ratcliff, 1979). Vincentizing the latency data across individual participants provides a way of averaging data while preserving the shapes of the individual distributions. Figure 2 shows the means by participants of the deciles plotted as cumulative distributions for each of the contexts and each of the SOAs for Experiment 1.

The left panel of Figure 2 shows that the context effect was absent throughout the monitoring latency distribution for the 300-ms SOA. The right panel of Figure 2 shows that the context effect was present over the entire latency range for the 600-ms SOA. The statistical analysis of the latencies at the 300-ms SOA yielded no interaction of context and decile, $F(9, 243) = 1.17$, $p = .32$. The statistical analysis of the latencies at the 600-ms SOA also yielded no interaction of context and decile, $F(9, 117) = 1.39$, $p = .20$. Thus, the context effect was stable throughout the monitoring latency distribution and did not depend on the absolute monitoring latency.

At first sight, the absence of a context effect for the slow phoneme monitoring responses at the 300-ms SOA seems incompatible with the presence of a context effect for the fast responses at the 600-ms SOA. Phoneme monitoring latencies were measured from the onset of the spoken words. The average latency of the 10% slowest responses at the 300-ms SOA was 1,017 ms. This means that the monitoring response actually occurred 1,317 ms (i.e., 300 ms + 1,017 ms) after picture onset. The average latency of the 10% fastest responses at 600-ms SOA was 519 ms. This means that the monitoring response actually occurred 1,119 ms (i.e., 600 ms + 519 ms) after picture onset. Thus, relative to picture onset, the slowest responses at 300-ms SOA occurred later than the fastest responses at 600-ms SOA. It would seem that for the slowest monitoring responses at 300-ms SOA, activation from the picture might reach the phonological level in time and cause a priming effect, as it did for the fastest responses at the 600-ms SOA. Still, a context effect was present for the fastest responses at the 600-ms SOA but not for the slowest responses at the 300-ms SOA. However, if phonological activation in the production system depends on lemma selection (Levelt et al., 1999) and the internal monitoring for picture name phonemes is interrupted earlier at the 300-ms SOA than at the 600-ms SOA, then the phonological activation at the 300-ms SOA may be insufficient to yield a priming effect, even for the slow responses. This early interruption account entails that what is done in response to the

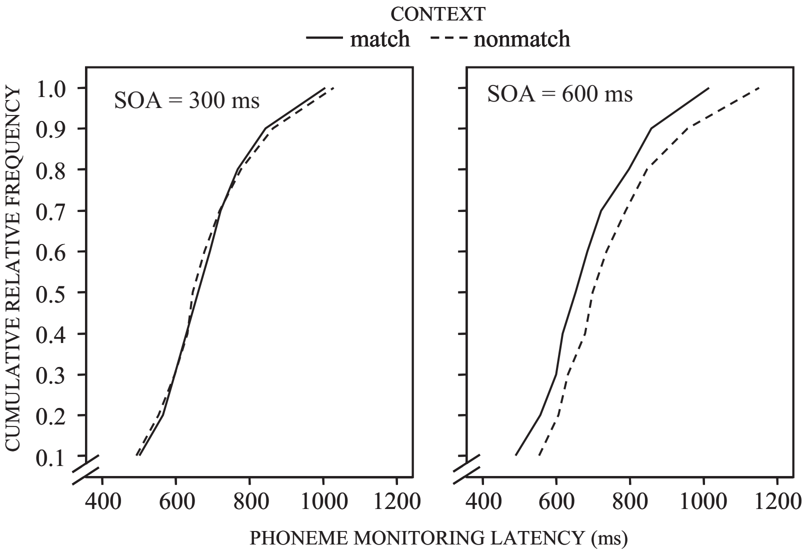

Figure 2. Vincentized cumulative distribution curves for the phoneme monitoring latencies in the matching and nonmatching picture contexts at stimulus onset asynchronies (SOAs) of 300 ms and 600 ms in Experiment 1. The monitoring latencies were measured from spoken word onset.

pictures in the experiment should be important, which is further examined in Experiments 2 and 4. In particular, it should matter whether active phonological encoding is required (tested in Experiment 2) or not (tested in Experiment 4).

To summarize, the statistical analysis showed that there were no significant context effects for either of the two phonemes at the short SOA (300 ms), whereas there were facilitation effects for both phonemes at the long SOA (600 ms). The absence and presence of priming effects was independent of the absolute monitoring latency. Whether facilitation was obtained appeared to depend on the onset of phonological encoding in the production system, leading to an absence of phonological priming at an SOA preceding the encoding onset (300 ms) and to a facilitation effect at an SOA following the encoding onset (600 ms). These findings suggest that active phonological encoding in response to the pictures helped the participants to recognize matching phonemes in the speech signal. The results suggest that there are internal phonological links running from speech planning to speech recognition.

It is important to exclude that the priming effect found at the long SOA is due to response preparation rather than due to a preactivation of the phoneme in the perception system. By including the internal monitoring condition, we gave the participants the opportunity to prepare for a response. If participants saw a picture whose name contained the target phoneme, then they could, in principle, already prepare the button press response. If the tone came, then participants only had to execute the response. When a word was presented instead of a tone, the response was already prepared. This could have led to a speeding up of the monitoring response in the match context.

Moreover, it is important to exclude that the links between the systems causing the preactivation had been established because of the internal monitoring task. One could argue that the phonological links do not exist in normal speech processing without involvement of an explicit internal monitoring task. We tried to rule out the response preparation and task-dependent-links objections by running Experiment 2.

Experiment 2

Our intention with Experiment 2 was to exclude an interpretation of the facilitation effect found in Experiment 1 in terms of response preparation or links that are set up only for the purpose of an internal phoneme monitoring task. We replaced the internal monitoring task with a simple picture naming task. Each time the participants heard a tone, they had to simply name the picture. This situation rules out the contribution of response preparation. Moreover, there is also no need for phonological links serving only an explicit internal monitoring task. If the priming effect remains in the present experiment, this would strongly suggest that the effect is not due to response preparation and task-dependent links.

Method

Participants. Twelve new participants (10 women; mean age = 20 years) from the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, design, procedure, and analysis. These were the same as for Experiment 1, except that the internal monitoring task was replaced by a picture naming task. We tested only the 600-ms SOA. The statistical analysis performed was exactly like the analysis in Experiment 1 but without the between-participants factor of SOA.

Results and Discussion

The mean monitoring latencies, their standard deviations, and the error percentages for each context and phoneme are given in

Table 1. The table shows that there was a facilitation effect for the phoneme /k/ as well as for the phoneme /p/.

The statistical analysis of the phoneme monitoring latencies yielded a main effect of context, $F_1(1, 11) = 8.74$, $p = .01$; $F_2(1, 18) = 9.07$, $p = .007$, but not of phoneme, $F_1(1, 11) < 1$, $p = .46$; $F_2(1, 18) = 1.39$, $p = .25$. There was no interaction of phoneme and context, $F_1(1, 11) < 1$, $p = .99$; $F_2(1, 18) < 1$, $p = .85$.

Table 1 shows that more errors were made in the nonmatch than in the match context. The statistical analysis of the errors yielded an effect of context, $F_1(1, 11) = 7.33$, $p = .02$; $F_2(1, 18) = 5.88$, $p = .026$, but not of phoneme, $F_1(1, 11) < 1$, $p = .66$; $F_2(1, 18) < 1$, $p = .65$. There was no interaction of phoneme and context, $F_1(1, 11) < 1$, $p = .69$; $F_2(1, 18) < 1$, $p = .69$. Given that most errors were made in the condition with the slowest responses, there is no evidence for a speed–accuracy trade-off in the data.

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

In our working model (illustrated in Figure 1), the influence from planning on recognition is mediated by phonological representations. However, the influence from production planning on speech recognition could also come from links between production and comprehension at higher levels, such as at the level of lemmas or lexical forms. Lexical influences imply backward links from lemmas to word-form perception or from lexical forms in morphological encoding to word-form perception, which differs from the priming route assumed by the working model illustrated in Figure 1. However, the results of Levelt et al. (1991) concerning the effect of spoken word planning on auditory lexical decision suggest that lexical influences already occur much earlier than the effect observed in Experiments 1 and 2. In Levelt et al.'s experiments, the responses to the spoken words were slower for phonologically related than phonologically unrelated picture names at SOAs of 73 and 373 ms, whereas the phonological effect in the present experiments happened at an SOA of 600 ms. Moreover, Levelt et al. observed phonological interference, whereas in the present experiment, phonological facilitation was obtained. Thus, the direction and timing of phonological effects differs between auditory lexical decision (Levelt et al., 1991) and phoneme monitoring (the present experiments). This difference suggests that the routes mediating the lexical and phonemic effects are distinct.

Experiment 3

In Experiments 1 and 2, participants monitored for phonemes in spoken words. Our working model predicts that the phonological influences of speech planning on speech recognition should be obtained regardless of whether the critical phonemes are part of spoken words or nonwords. To test the latter prediction, we ran a phoneme monitoring experiment with target phonemes embedded in spoken nonwords (cf. Connine & Titone, 1996). The target phonemes were again all in initial position. All spoken words of Experiments 1 and 2 were replaced by spoken nonwords. Finding the facilitation effect again with the target phonemes in nonwords would suggest that the lexical status of the spoken item is not critical. The task performed on the pictures was again picture naming, as in Experiment 2.

Method

Participants. Twelve new participants (10 women; mean age = 21 years) from the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, design, procedure, and analysis. These were the same as for Experiment 2. We used the same picture primes as in Experiments 1 and 2 and created nonwords from the spoken words used in Experiments 1 and 2. In the monosyllabic words, we changed the coda (e.g., "paal" was replaced by "paag"). For the disyllabic words, we changed the second syllable (e.g., "kasteel" was replaced by "katroog"). The segmental structure was kept the same. The spoken items were newly recorded to avoid splicing artifacts. Again, only an SOA of 600 ms was tested. The analysis of the monitoring latencies and errors was performed exactly as in Experiment 2.

Results and Discussion

The mean monitoring latencies, their standard deviations, and the error percentages for each context and phoneme are given in Table 1. The table shows that there was a facilitation effect for /k/ as well as for /p/. The statistical analysis of the phoneme monitoring latencies yielded a main effect of context, $F_1(1, 11) = 5.00$, $p = .047$; $F_2(1, 18) = 7.23$, $p = .015$, but not of phoneme, $F_1(1, 11) < 1$, $p = .60$; $F_2(1, 18) < 1$, $p = .67$. There was no interaction of phoneme and context, $F_1(1, 11) = 2.64$, $p = .13$; $F_2(1, 18) = 2.11$, $p = .16$.

Table 1 shows that more errors were made in the nonmatch than in the match context. However, the statistical analysis of the errors yielded no effect of context, $F_1(1, 11) = 1.96$, $p = .19$; $F_2(1, 18) = 3.08$, $p = .10$, or of phoneme, $F_1(1, 11) < 1$, $p = .86$; $F_2(1, 18) < 1$, $p = .83$. There was also no interaction of phoneme and context, $F_1(1, 11) < 1$, $p = .83$; $F_2(1, 18) < 1$, $p = .81$.

Again, to check the stability of the patterns observed in the means, monitoring latency distributions were obtained. Figure 3 shows the distributional plots for the match and nonmatch conditions. The figure shows that the context effect was numerically present over almost the entire latency range. The statistical analysis of the latencies revealed an interaction of context and decile, $F(9, 99) = 3.19$, $p = .002$. This suggests that the magnitude of the priming effect depended somewhat on relative latency. However, further analyses using quantile–quantile plots (Thomas & Ross, 1980; Wilk & Gnanadesikan, 1968) revealed that the increase of the priming effect with latency was not disproportionate. A quantile–quantile plot is a standard technique for determining whether two distributions belong to the same distribution family. If they do, then the plot should be linear, indicating that the distributions only differ by a scale or shift factor. A perfectly linear relationship was obtained for the phoneme monitoring latencies in the match and nonmatch conditions ($R^2 = .995$).

To summarize, in Experiment 3 with phoneme targets in nonwords, the facilitation effect of context pictures was again obtained. This finding suggests that the lexical status of the spoken item is not critical, as predicted by our working model.

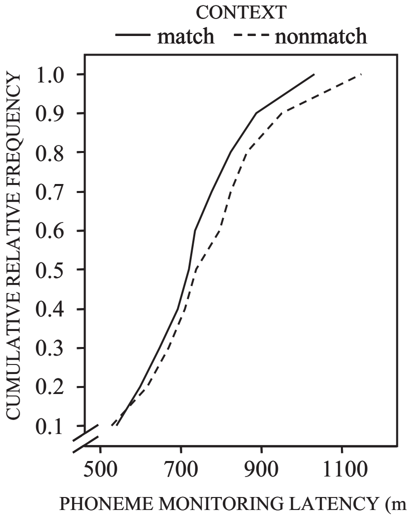

Figure 3. Vincentized cumulative distribution curves for the phoneme monitoring latencies in the matching and nonmatching picture contexts in Experiment 3. The monitoring latencies were measured from spoken word onset.

Comparison of Experiments 2 and 3

Although the target phonemes in Experiments 2 and 3 were the same, they were part of different speech tokens. This makes direct comparison of the magnitudes of the priming effects between experiments somewhat problematic. (A comparison with Experiment 1 is even more problematic, because the tasks performed in response to the pictures were different.) Still, Table 1 shows that the mean response latencies in the match and nonmatch conditions were almost the same in Experiments 2 and 3. In the nonmatch condition, the overall means were even identical. This suggests that there was no difference in the time it took to identify the target phonemes (unprimed) in the word and nonword speech tokens. We submitted the data of Experiments 2 and 3 to a combined statistical analysis to assess whether there were any statistical differences in context effect on phoneme monitoring latencies between Experiment 2 (words: 48-ms facilitation, on average) and Experiment 3 (nonwords: 40-ms facilitation, on average).

The combined analysis yielded a main effect of context, $F_1(1, 22) = 13.40$, $p = .001$; $F_2(1, 36) = 16.30$, $p = .001$, but not of experiment, $F_1(1, 22) < 1$, $p = .92$; $F_2(1, 36) < 1$, $p = .72$, or of phoneme, $F_1(1, 22) < 1$, $p = .36$; $F_2(1, 36) = 1.36$, $p = .25$. There was no interaction of experiment and context, $F_1(1, 22) < 1$, $p = .67$; $F_2(1, 36) < 1$, $p = .71$; experiment and phoneme, $F_1(1, 22) < 1$, $p = .84$; $F_2(1, 36) < 1$, $p = .56$; or context and phoneme, $F_1(1, 22) = 1.08$, $p = .31$; $F_2(1, 36) = 1.28$, $p = .27$. There was also no interaction of experiment, context, and phoneme, $F_1(1, 22) = 1.11$, $p = .30$; $F_2(1, 36) < 1$, $p = .41$. To conclude, a combined analysis of the phoneme monitoring latencies in Experiment 2 (words) and Experiment 3 (nonwords) revealed no difference in effects between experiments.

Of course, if one accepts that there is no difference in priming effect between the monitoring for phonemes in words (Experiment 2) and nonwords (Experiment 3), then this would mean that one accepts a null hypothesis. However, the fundamental aim of Experiment 3 was to see whether we could replicate the priming effect for nonwords. Having obtained the effect for nonwords shows that a lexical contribution is not essential for the priming effect to occur and involves rejecting the null hypothesis. The fact that the magnitude of the priming effect was the same for words and nonwords further suggests that word status is entirely irrelevant, but that is not essential.

Experiment 4

In Experiments 1–3, participants had to respond to the picture primes by monitoring for phonemes in the picture names (Experiment 1) or by naming the pictures (Experiments 2 and 3). Experiment 4 examined whether priming is obtained when participants passively view the pictures. The WEAVER++ model (Levelt et al., 1999) implements the claim that only selected lemmas spread activation to the phonological level. Thus, the model predicts that no priming should be obtained with passive viewing only.

Method

Participants. Twenty new participants (15 women; mean age = 21 years) from the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. They all had normal or corrected-to-normal hearing and vision and were paid for their participation.

Materials, design, procedure, and analysis. We used the same materials and design as in Experiment 2. In the present experiment, participants only had to respond if they heard the prespecified phoneme in the spoken word. There were no tones, and there was no additional task to force them to do anything with the pictures. We tested only the 600-ms SOA. After the experiment, the participants were given a recognition test for the pictures. This was a paper sheet with the 20 test pictures and 20 pictures that were not shown in the experiment. The participants had to indicate which pictures they had seen. This was done to verify that the participants looked at the screen during the experiment. The mean recognition accuracy in the picture recognition test after the experiment was 95.25%. The analysis of the response times and errors in the experiment was performed exactly like the analysis in Experiment 2.

Results and Discussion

The mean monitoring latencies, their standard deviations, and the error percentages for each context and phoneme are given in Table 1. The table shows that there was no facilitation effect, neither for phoneme /k/ nor for phoneme /p/. If anything, monitoring latencies for /k/ were longer for the match than for the nonmatch context.

The statistical analysis of the phoneme monitoring latencies yielded no main effect of context, $F_1(1, 19) < 1$, $p = .55$; $F_2(1, 18) < 1$, $p = .53$. However, there was an effect of phoneme, $F_1(1, 19) = 8.98$, $p = .007$; $F_2(1, 18) = 15.95$, $p = .001$. The participants identified the phoneme /k/ faster than the phoneme /p/. There was no interaction of phoneme and context, $F_1(1, 19) = 2.43$, $p = .14$; $F_2(1, 18) < 1$, $p = .35$.

Table 1 shows that most errors were made in the match condition, which numerically also had the slowest responses. However, the statistical analysis of the errors yielded no main effect of context, $F_1(1, 19) = 3.35$, $p = .08$; $F_2(1, 18) = 3.08$, $p = .10$, or of phoneme, $F_1(1, 19) = 1.0$, $p = .33$; $F_2(1, 18) < 1$, $p = .83$. There was no interaction of phoneme and context, $F_1(1, 19) = 2.41$, $p = .14$; $F_2(1, 18) < 1$, $p = .81$.

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

Comparison of Experiments 2 and 4

Experiments 2 and 4 used exactly the same materials. Priming of phoneme monitoring was obtained when participants had to respond to the pictures on some of the trials in Experiment 2 but not when they only passively viewed the pictures in Experiment 4. To confirm that there were statistical differences in phoneme monitoring latencies between Experiment 2 (active) and Experiment 4 (passive), a combined statistical analysis was performed. The analysis yielded main effects of context, $F_1(1, 30) = 5.61$, $p = .024$; $F_2(1, 36) = 3.85$, $p = .057$; experiment, $F_1(1, 30) = 13.53$, $p = .001$; $F_2(1, 36) = 151.61$, $p = .001$; and phoneme, $F_1(1, 30) = 5.06$, $p = .032$; $F_2(1, 36) = 9.90$, $p = .003$. There was an interaction of experiment and context, $F_1(1, 30) = 9.21$, $p = .005$; $F_2(1, 36) = 7.58$, $p = .009$, but not of experiment and phoneme, $F_1(1, 30) < 1$, $p = .49$; $F_2(1, 36) = 1.35$, $p = .25$, or of context and phoneme, $F_1(1, 30) < 1$, $p = .40$; $F_2(1, 36) < 1$, $p = .66$. There was also no interaction of experiment, context, and phoneme, $F_1(1, 30) < 1$, $p = .41$; $F_2(1, 36) < 1$, $p = .46$. To conclude, a combined statistical analysis of Experiments 2 and 4 confirmed that priming of phoneme monitoring was obtained when participants had to respond to the pictures on some of the trials in the experiment (Experiment 2) but not when they only passively viewed the pictures in the experiment (Experiment 4), as shown by the interaction of experiment and context.

It is possible that with passive picture viewing (Experiment 4), the build up of phonological activation in the production system takes more time than with active viewing. Consequently, it might be possible that a picture context effect is obtained for the slow monitoring responses (see the *Results and Discussion* section of Experiment 1). Again, to check the stability of the patterns observed in the means, monitoring latency distributions were obtained. Figure 4 shows the distributional plots for Experiment 2 (active) and Experiment 4 (passive). The left panel of Figure 4 shows that the context effect in Experiment 2 (requiring a response to the pictures on some trials) was present over the entire latency range, as was the case for Experiment 1 (see the right panel of Figure 2). The statistical analysis of the latencies yielded no interaction of context and decile, $F(9, 99) < 1$, $p = .85$. Thus, the difference in monitoring latencies observed in the means for the match and nonmatch contexts is consistent across the whole latency distribution. The right panel of Figure 4 shows that the context effect was absent over the entire latency range in Experiment 4 (requiring no response to the pictures). The statistical analysis of the latencies yielded no interaction of context and decile, $F(9, 171) < 1$, $p = .91$. The absence of a difference in the means was stable throughout the entire latency distribution. Thus, the presence of a context effect in Experiment 2 (active) and the absence of the effect in Experiment 4 (passive) is a robust phenomenon, independent of the absolute response latency.

To conclude, priming of phoneme monitoring was obtained across the entire latency distribution when participants had to respond to the pictures on some of the trials in the experiment (Experiment 2). No such priming effect was obtained over the entire latency distribution when the participants only passively viewed the pictures in the experiment (Experiment 4). The difference between experiments suggests that responding to the pictures on some of the trials is a precondition for obtaining the priming effects.

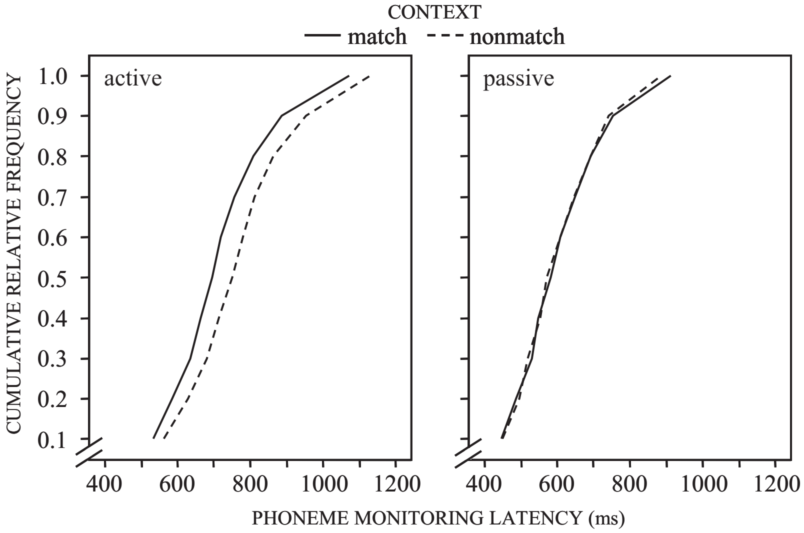

Figure 4. Vincentized cumulative distribution curves for the phoneme monitoring latencies in the matching and nonmatching picture contexts when pictures had to be responded to on some trials in the experiment (active; Experiment 2) but not others (passive; Experiment 4). The monitoring latencies were measured from spoken word onset.

General Discussion

We reported four experiments that examined the influence of spoken word planning on speech recognition. Participants were presented with pictures and a tone or spoken word 300 or 600 ms after picture onset. When a spoken word was presented, participants had to indicate whether it contained a prespecified phoneme. When the tone was presented, they had to indicate whether the picture name contained the phoneme (Experiment 1) or name the picture (Experiments 2 and 3). The phoneme monitoring latencies for the spoken words were shorter when the picture name contained the prespecified phoneme compared with when it did not. The facilitation was only obtained at the SOA of 600 ms. Facilitation was also obtained for spoken nonwords (Experiment 3). No facilitation was obtained when the pictures required no response (Experiment 4). These results suggest that there are internal phonological connections running from the word production system to the word comprehension system, as implied by an internal perceptual loop in verbal self-monitoring (Levelt, 1983; Levelt et al., 1999) and a phonological loop in working memory (Baddeley, 1986, 2003). Moreover, the data suggest that active phonological encoding in the speech production system is a precondition for engaging the connections between planning and recognition.

The priming effects in the present experiments were obtained with an SOA of 600 ms between picture and spoken word. One may argue that such long interval between picture and word is enough to subvocally produce the picture name. If this were the case, then the influence of production on perception could be at a phonetic rather than at a phonological level. However, we believe that it is unlikely that a phonetic influence of production on perception can wholly explain the priming effects in our experiments. We discuss two reasons.

First, Gordon and Meyer (1984) and D. E. Meyer and Gordon (1984) observed that fully preparing a syllable for production interfered with spoken syllable perception if the onset consonants of the prepared and heard syllables shared the voicing feature. Place of articulation had no effect. In contrast, the effect of production on perception in the present experiments was one of facilitation rather than of interference. Facilitation was obtained even though the voicing feature was shared between production primes and perception targets. The difference in direction of the effects between studies (interference vs. facilitation) suggests that the effects in the present experiments did not wholly arise at the phonetic level.

It may be that production and perception do not compete for a voicing mechanism (D. E. Meyer & Gordon, 1984) if the voicing feature is part of the same phoneme. If so, then this would mean for the present experiments that there was competition in the nonmatch context (where prime and target phonemes differed) but not in the match context (where prime and target phonemes were the same), explaining the facilitation. However, such a special status for phonemic identity would suggest that the phonological level is involved (cf. Roelofs, 1999). A special role for phonemic identity would be evidence that the priming effect in the present experiments is at least partly mediated by phonological representations.

Second, we obtained the facilitation effect regardless of whether the response to the pictures was internal phoneme monitoring (Experiment 1) or naming (Experiments 2 and 3). Phonetic preparation of the picture name happens with picture naming, but it does not underlie internal phoneme monitoring, as shown by Wheeldon and Levelt (1995). Monitoring latencies in their study of internal monitoring increased with the serial position of the target phonemes within the word. The serial position effect remained unaltered when participants had to perform the phoneme monitoring task while simultaneously counting aloud, which is known to suppress phonetic representations (Baddeley, 1986, 2003). Moreover, Wheeldon and Levelt (1995) observed that internal monitoring was sensitive to syllable structure. These results suggest that internal self-monitoring involves a phonological rather than a phonetic representation. Given that the priming effect in the present experiments occurred regardless of whether the response to the pictures involved internal phoneme monitoring or naming, it is likely that the effect is at least partly mediated by phonological representations rather than wholly by phonetic representations.

The results of Experiment 4 suggest that passive picture viewing does not lead to sufficient phonological activation to cause a significant facilitation effect in an external phoneme monitoring task, regardless of monitoring latency. Although participants were paying attention to the pictures, as indicated by the good recognition scores obtained after the experimental session of Experiment 4, the attention given to the pictures was apparently not enough to cause the priming effect. We discuss two possible reasons for why this was the case.

First, it could be that there is a discrete information flow within the production system, with only selected lemmas activating the corresponding word forms, as implemented in WEAVER++ (Levelt et al., 1999; Roelofs, 2006). With passive picture viewing, the task does not require the selection of lemmas. Consequently, word forms do not become activated, and picture-induced priming of external phoneme monitoring does not occur, as observed in Experiment 4. In contrast, in monitoring for phonemes in internal speech (Experiment 1) or in picture naming (Experiments 2 and 3), a lemma needs to be selected as part of accomplishing the task. Consequently, picture-induced priming of external phoneme monitoring should occur, as observed in Experiments 1–3.

Second, the information flow through the production system could be cascading, but the distance in the lexical network from concepts to word forms might be too long to obtain much activation at the phonological level (cf. Roelofs, 2003a). For example, Dell (1986) proposed that retrieval of information from the lexical network involves jolting the activation of target nodes at each processing level. These activation jolts are only given when the task requires phonological encoding. In the absence of the activation jolts, the network distance between concepts and phonological forms may be too long to obtain much phonological activation from irrelevant pictures. This would explain why priming effects were obtained in Experiments 1–3, where the task required phonological encoding, but not in Experiment 4, where phonological encoding was not required.

The results of Experiments 2 and 3 (with picture naming as control task) suggest that influences of speech planning on speech recognition are obtained even when there is no explicit self-monitoring task. This suggests that the encoding of a phonological representation for the picture is sufficient to yield the priming effect. This seems to disagree with an assumption made for the WEAVER++ model by Roelofs (2004). In WEAVER++, the phonological links from production to comprehension are established as a consequence of constructing a phonological word representation for the picture name. Roelofs (2004) assumed that the feeding of the phonological representation into the recognition system is under a speaker's control. Encoded phonological representations are sent to the recognition system for monitoring purposes. Similarly, the subvocal rehearsal process that is part of the phonological loop in Baddeley's (1986, 2003) model of working memory is a control process, which is by definition at the option of participants. However, the findings of Experiments 2 and 3 suggest that an explicit internal monitoring task is not required to obtain the influence from production planning on speech recognition.

It may be the case, though, that the default control setting of the production system is to allow for self-monitoring. This means that a constructed phonological representation is habitually sent to the recognition system. The priming effects in Experiments 2 and 3 suggest that speakers encoded phonological representations of the picture names on all trials. If phonological representations had only been encoded upon hearing the tone (indicating that the picture had to be named), then no priming effect should have been obtained in the phoneme monitoring task. If encoded phonological representations are habitually sent to the speech recognition system, even when there is no explicit internal monitoring task, then the priming effects on phoneme monitoring with picture naming as control task (Experiments 2 and 3) are explained.

In WEAVER++, phonological word representations are constructed incrementally from the beginning of a word to its end. Correspondingly, the phonological word representation is fed into the recognition system as it becomes available over time (Roelofs, 2004). This results in sequential activation of the perception system, as is the case with the processing of external speech. The sequential activation of the perception system by the production system agrees with the effect of serial position that was obtained by Özdemir et al. (in press), Van Turennout et al. (1997), Wheeldon and Levelt (1995), and Wheeldon and Morgan (2002) for the monitoring of phonemes in internal speech. Moreover, the sequential activation of the perception system by the production system predicts the influences of production planning on speech perception that were observed in the present experiments.

Influences of speech production on recognition should also occur if phonological representation and processes are shared between production and perception (e.g., Allport, 1984; MacKay, 1987). The present results do not distinguish between a shared phonological system, on the one hand, and separate but closely linked systems, on the other. However, using the phoneme monitoring task, Özdemir et al. (in press) obtained effects of the perceptual uniqueness point of picture names on the monitoring latencies for target phonemes in those names, whereas no such effects were obtained on the latencies of naming the pictures. These results are most consistent with the view of separate but closely linked systems.

To conclude, the reported experiments provide evidence for phonological links running from speech planning to speech recognition, as entailed by an internal perceptual loop in verbal self-monitoring (Levelt, 1983; Levelt et al., 1999) and a phonological loop in working memory (Baddeley, 1986, 2003). Speech planning influences speech recognition in the context of picture naming and internal self-monitoring tasks, but not in the context of passive picture viewing. These results suggest that the active encoding of phonological representations is a precondition for obtaining the word planning influences on speech recognition.

References

Allport, A. (1984). Speech production and comprehension: One lexicon or two? In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 209–228). Berlin, Germany: Springer-Verlag.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

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

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

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

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

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

Gathercole, S. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3,* 410–419.

Gordon, P. C., & Meyer, D. E. (1984). Perceptual–motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 153–178.

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

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92,* 101–144.

Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A.-C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology, 24,* 3–22.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10,* 480–486.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and control* (pp. 183–223). Cambridge, England: Cambridge University Press.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38,* 243–294.

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

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

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

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

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

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

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

Martin, N., & Saffran, E. M. (2002). The relationship of input and output phonological processing: An evaluation of models and evidence to support them. *Aphasiology, 16,* 107–150.

Martin, R., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41,* 3–29.

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

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain, 119,* 907–917.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1146–1160.

Meyer, D. E., & Gordon, P. C. (1983). Dependencies between rapid speech perception and production: Evidence for a shared sensory–motor voicing mechanism. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 365–378). Hillsdale, NJ: Erlbaum.

Meyer, D. E., & Gordon, P. C. (1984). *Shared mechanisms for perceiving and producing phonetic features in speech* (Tech. Rep. No. 65). Ann Arbor: University of Michigan.

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 273–311). London: Academic Press.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring, and lexicality. *Cortex, 31,* 209–237.

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

Özdemir, R. (2006). *The relationship between spoken word production and comprehension.* Unpublished doctoral dissertation, University of Nijmegen, the Netherlands.

Özdemir, R., Roelofs, A., & Levelt, W. J. M. (in press). Perceptual uniqueness point effects in monitoring internal speech. *Cognition.*

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (in press). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language.*

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience, 8,* 2236–2246.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 461–466.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142.

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

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14,* 173–200.

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

Roelofs, A. (2003a). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110,* 88–125.

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

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

Roelofs, A. (2005). The visual–auditory color–word Stroop asymmetry and its time course. *Memory & Cognition, 33,* 1325–1336.

Roelofs, A. (2006). Context effects of pictures and words in naming objects, reading words, and generating simple phrases. *Quarterly Journal of Experimental Psychology, 59,* 1764–1784.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language, 57,* 232–251.

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

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, England: Cambridge University Press.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37*(A), 507–532.

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

Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology, 21,* 136–152.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 787–806.

Wernicke, C. (1874). *Der aphasische Symptomenkomplex: Eine psychologische studie auf anatomischer basis* [The aphasic symptom complex: A psychological study on anatomical basis]. Breslau, Poland: Cohn & Weigert.

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

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

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika, 55,* 1–17.

Wise, R., Scott, S., Blank, S., Mummery, C., Murphy, K., & Warburton, E. (2001). Separate neural subsystems within “Wernicke’s area.” *Brain, 124,* 83–95.

Appendix

Experimental Stimuli for Experiments 1–4

Prime (picture)		Target (spoken item)	
Match	Nonmatch	Word	Nonword
paard (*horse*)	kers (*cherry*)	poets (*trick*)	poers
peer (*pear*)	kam (*comb*)	paal (*post*)	paag
pijl (*arrow*)	kast (*cupboard*)	puin (*rubble*)	puig
palm (*palm*)	kom (*bowl*)	pers (*press*)	pert
pen (*pen*)	kaars (*candle*)	puk (*mite*)	pun
puzzel (*puzzle*)	konijn (*rabbit*)	panter (*panther*)	pansug
penseel (*paintbrush*)	kameel (*camel*)	paleis (*palace*)	paruif
pincet (*tweezers*)	kassa (*till*)	piraat (*pirate*)	piluuf
pistool (*pistol*)	ketel (*kettle*)	parfum (*perfume*)	parsin
passer (*compass*)	kegel (*cone*)	peddel (*paddle*)	pebbor
kers (*cherry*)	paard (*horse*)	korf (*basket*)	kols
kam (*comb*)	peer (*pear*)	kist (*box*)	kirs
kast (*cupboard*)	pijl (*arrow*)	koets (*coach*)	koerl
kom (*bowl*)	palm (*palm*)	kan (*jug*)	kal
kaars (*candle*)	pen (*pen*)	kurk (*cork*)	kurl
konijn (*rabbit*)	puzzel (*puzzle*)	kasteel (*castle*)	katroog
kameel (*camel*)	penseel (*paintbrush*)	karaf (*carafe*)	kalis
kassa (*till*)	pincet (*tweezers*)	kachel (*stove*)	kafug
ketel (*kettle*)	pistool (*pistol*)	klaver (*clover*)	klasuf
kegel (*cone*)	passer (*compass*)	koffer (*suitcase*)	konnis

Received January 17, 2006
Revision received May 11, 2007
Accepted May 21, 2007 ■