

Word length effects in object naming: The role of a response criterion

Antje S. Meyer,^{a,*} Ardi Roelofs,^b and Willem J.M. Levelt^b

^a *Behavioral Brain Sciences Centre, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

^b *Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands*

Received 20 September 2001; revision received 12 April 2002

Abstract

According to Levelt, Roelofs, and Meyer (1999) speakers generate the phonological and phonetic representations of successive syllables of a word in sequence and only begin to speak after having fully planned at least one complete phonological word. Therefore, speech onset latencies should be longer for long than for short words. We tested this prediction in four experiments in which Dutch participants named or categorized objects with monosyllabic or disyllabic names. Experiment 1 yielded a length effect on production latencies when objects with long and short names were tested in separate blocks, but not when they were mixed. Experiment 2 showed that the length effect was not due to a difference in the ease of object recognition. Experiment 3 replicated the results of Experiment 1 using a within-participants design. In Experiment 4, the long and short target words appeared in a phrasal context. In addition to the speech onset latencies, we obtained the viewing times for the target objects, which have been shown to depend on the time necessary to plan the form of the target names. We found word length effects for both dependent variables, but only when objects with short and long names were presented in separate blocks. We argue that in pure and mixed blocks speakers used different response deadlines, which they tried to meet by either generating the motor programs for one syllable or for all syllables of the word before speech onset. Computer simulations using WEAVER++ support this view.

© 2002 Elsevier Science (USA). All rights reserved.

Keywords: Language production; Word length effects; Response criterion; Phonological encoding; Eye tracking

Though adults often tell children that they must think *before* they speak, adults rarely follow this rule themselves; at least they usually do not fully plan every word of an utterance before they start to speak. Instead, speaking and speech planning occur in parallel. When speakers converse or describe events or scenes, they may have a global plan for their entire utterance and perhaps a more specific plan for the materials to be covered in the first clause. However, lexical access, the retrieval of

individual words, is usually confined to the first few words of the first sentence. The other words are retrieved while the preceding ones are being said or in pauses between words (Clark, 1996; Levelt, 1989).

But what happens at the level of individual words? Is a single word fully planned before speech onset, or do speakers begin to speak as soon as they have planned a single syllable or segment? How do speakers orchestrate speech planning and speech output at the word level? How do they decide when it is time to begin to say a word? Lashley (1951) argued that in fluent rapid action, such as producing speech, typing, and playing music, there is too little time to specify each elementary action component after execution of the previous element, as in

* Corresponding author.

E-mail addresses: a.s.meyer@bham.ac.uk (A.S. Meyer), ardi@mpi.nl (A. Roelofs), pim@mpi.nl (W.J.M. Levelt).

an associative chain mechanism. He argued that, instead, motor programs for sequences of elementary actions must be established before execution onset. This raises the issue of how these sequences are prepared and how many elementary actions they comprise. Speakers could, for instance, generate a plan for an entire word, for one syllable, or part of a syllable before speech onset.

In the experiments reported below, we compared the preparation times for longer and shorter words. If words are fully planned before speech onset, and if planning takes longer for long than for short words, the preparation times should also be longer for long than for short words. To capture preparation times, we measured speech onset latencies and, in one experiment, how long the speakers looked at the objects they named. The speakers produced the long or short target words in isolation or as the first word of a phrase. We presented the materials in “pure” test blocks, i.e., blocks including only objects with long names or with short names, and in “mixed” blocks including a mixture of objects with long and short names. As we will explain below, we expected speakers to base their decision when to begin to speak on different criteria in the mixed and pure blocks and we anticipated sizable word length effects only in the pure blocks. In order to motivate the experiments, we will first outline our model of lexical access and explain its predictions concerning word length effects. We will then summarize the existing evidence concerning the effects of word length and stimulus blocking on word production latencies.

A working model of lexical access

According to our model of lexical access in speech production (Levelt et al., 1999), the main stages of lexical access to a word are (1) the retrieval of a lexical concept defining the meaning to be expressed, (2) the selection of a lemma, which is a syntactic representation of a word, and (3) the generation of the corresponding word form (for a discussion of alternative models see Levelt, 1999; Rapp & Goldrick, 2000). Word form generation encompasses three steps, namely the retrieval of one or more morphemes (morphological encoding), the generation of the phonological representation (phonological encoding), and the generation of an articulatory program (phonetic encoding). Speakers often commit segmental errors such as “teep a cape” (instead of “keep a tape,” Fromkin, 1971, p. 31), in which individual segments or clusters are selected or ordered incorrectly. The error outcomes are usually phonetically well-formed (as, for instance, evidenced by the error “a meeting_arathon” instead of “an eating marathon,” Fromkin, 1971, p. 41). This shows that most sound errors do not arise during articulation or phonetic encoding, but earlier, during phonological planning. Thus, the occurrence of sound errors, along with a substantial body of experimental

evidence (for a review see, for instance, Levelt et al., 1999), strongly suggest that speakers do not retrieve the phonological forms of words or morphemes as units from their mental lexicon, but generate them by selecting and combining phonological segments.

We assume that the segments of a morpheme are retrieved in parallel. Their serial order is coded in labeled links to the super-ordinate morpheme node. However, word forms are not just strings of segments, but consist of stressed and unstressed syllables. Thus, the retrieved set of segments must be parsed into syllables. In our model, the syllabic structure of words is not stored but is computed during phonological encoding following universal and language-specific syllabification rules. Similarly, the lexical entries of most words do not include information about the words’ stress pattern. Instead, stress is derived by rule. For instance, in English and Dutch most words are stressed on the first syllable with a full vowel. For words deviating from this rule, the stress pattern must be stored as part of the lexical entry. For the present purposes, it is important that the parsing of the string of segments into syllables is taken to be a sequential process running from the beginning of the word to its end. As the phonological syllables are successively composed, the corresponding articulatory programs are retrieved. Speakers can monitor their developing speech plan. In our model, the representation that is scrutinized in self-monitoring is the phonological representation (see Wheeldon & Levelt, 1995).

Other models of word production (e.g., Dell, 1986; Dell, Burger, & Svec, 1997; MacKay, 1987; Stemmer, 1985; Vousden, Brown, & Harley, 2000) conceptualize syllabification and stress assignment in different ways. Most commonly, richer lexical entries are assumed, which include metrical information for all words. Often there are metrical frames capturing the number of syllables, their stress values, and, in some models, syllable-internal structure. During word form encoding, the string of segments is mapped onto the positions of the metrical frame. The result is a syllabified phonological representation, in which each segment is assigned to at least one syllable. However, all major models share the assumption that the assignment of segments to successive syllables occurs in sequence. Some models, including our working model, assume that segments within syllables are assigned to syllable positions in sequence as well. Evidence for the sequential encoding assumption comes from a variety of sources including reaction time experiments using a word preparation task (Meyer, 1990, 1991; Roelofs, 1996, 1998), primed picture naming (Meyer & Schriefers, 1991), combined primed/prepared word production (Roelofs, 2002a), self-monitoring (Wheeldon & Levelt, 1995), repeated pronunciation (Sevold & Dell, 1994), and electrophysiological studies (van Turenhout, Hagoort, & Brown, 1997). All of these studies support the view that the segments at the

beginning of a word are associated to their positions slightly before the segments at the end of the word.

Word length effects in speech production

If the mapping of segments onto syllables and the retrieval of the corresponding articulatory programs are sequential processes, it should take speakers longer to generate the phonological forms of long words than of short ones. If speakers phonologically encode the entire word and retrieve all articulatory programs before beginning to speak, speech onset latencies should be longer for long than for short words. By contrast, if they only encode part of a word, for instance the initial syllable, before speech onset, latencies for long and short words should not differ. Provided that speakers indeed generate the phonological and articulatory codes of successive syllables in sequence, the speech onset latencies for words differing in length should reveal whether speakers prepare a constant fragment of the form representation for all words, or whether they prepare a longer fragment for long than for shorter words.

Our working model of lexical access and the related computational model *WEAVER++* (Roelofs, 1997a, b) assume that speakers usually prepare at least one phonological word before beginning to speak. This assumption is based on linguistic considerations and the results of a number of word production studies. Klapp, Anderson, and Berrian (1973) found a word length effect for object naming and Eriksen, Pollack, and Montague (1970) found such an effect for number naming. Wheeldon and Lahiri (1997) tested whether the minimal production unit in connected speech was a lexical or a phonological word. They presented Dutch speakers with questions (e.g., *Wat zoek je?* ‘what are you looking for?’) and cues to the answers (e.g., *water* ‘water,’ *het water* ‘the water,’ or *vers water* ‘fresh water’). The speakers answered in short sentences (e.g., *Ik zoek water* ‘I am looking for water’). The mean speech onset latency was longer for sentences including a determiner (*Ik zoek het water*) than for the other two sentence types (*Ik zoek water* and *Ik zoek vers water*). Most likely, this difference arose because the determiner attached itself to the preceding verb to form one phonological word. Thus, in the determiner condition the first phonological word of the utterance included three syllables (*ik.zoe.ket*), whereas in the other two conditions it only included two syllables (*ik.zoek*). Apparently, speakers prepared the first phonological word of the utterance before initiating the response, which took longer when the word included three syllables than when it included only two. Finally, Santiago, MacKay, Palma, and Rho (2000) (see also Roelofs, 2002b; Santiago, MacKay, & Palma, 2002) also reported longer naming latencies for objects with disyllabic names than for objects with monosyllabic names. All of these results support the assumption that speakers

encode at least one full phonological word before speech onset. In fact, Costa and Caramazza (2002) have recently shown that speakers producing adjective noun phrases tended to encode both phonological words of the phrase before speech onset.

However, Bachoud-Lévi, Dupoux, Cohen, and Mehler (1998) noted that in the object naming experiment reported by Klapp et al. the targets with long and short names were matched for word frequency but not for object familiarity. In addition, they noted that in the number naming experiment reported by Eriksen et al., most disyllabic words were “reference numerals,” such as *15* or *20*, whereas the longer words were non-reference numerals, such *17* or *28*, and tended to denote larger quantities than the disyllabic numerals. Results of a study by Dehaene and Mehler (1992) suggest that these variables, rather than word length, may be responsible for the faster latencies for the disyllabic items. Hence, Bachoud-Lévi et al. re-examined the effect of word length on the naming latencies for simple nouns. Their Experiment 1, carried out in French, compared the naming latencies for objects with mono- and disyllabic names. Frequency and familiarity were balanced across conditions, and mono- and disyllabic items were pairwise matched for initial segment. In this carefully designed experiment, there was no hint of a word length effect. Experiment 3 had a similar design but was carried out in English. The result was the same: no word length effect. Experiments 2 (with French speakers) and 4 and 5 (with English speakers) used the symbol naming task introduced by Levelt and Wheeldon (1994). The participants first learned symbol–word pairs (e.g., “@@ = doctor”; “## = soldier”) and later produced the words as quickly as possible upon presentation of the symbols. Number of syllables in the response words (one, two, or three syllables) and response word frequency were varied orthogonally. In all experiments, robust frequency effects were obtained, but there were no significant word length effects. The absence of a length effect is, of course, a null-result and must be interpreted with caution. Perhaps a small existing word length effect was not detected due to lack of experimental power. However, given that there was a robust frequency effect, this does not seem very likely. Instead, the results suggest that the speakers either did not prepare successive syllables in sequence, but did so in parallel, or that they began to speak before having prepared all syllables of the target words. Given the substantial body of evidence showing that syllables are generated in sequence, the latter assumption is more plausible. Thus, Bachoud-Lévi et al. concluded that speakers often began to speak before having encoded an entire phonological word.

Based on results of a picture–word interference study Schriefers and Teruel (1999) made a similar suggestion. In their experiments, German speakers described objects

in adjective noun phrases such as *rosa Sofa* ‘pink couch,’ while hearing distracter syllables that were identical to one of the syllables included in the phrase or that were phonologically unrelated. The utterances were initiated faster when the distracter was identical to the first syllable of the utterance than when it was unrelated. Distracters that were identical to the second syllable led to facilitation for some speakers, who could be classified on the basis of independent criteria as “careful speakers.” For the other speakers, the “hasty” ones, Schriefers and Teruel found no priming effect. They concluded that the speakers differed in their preferences for smaller or larger phonological planning units. However, these results are also compatible with the assumption that speakers always completed the *phonological* encoding of the first word of their utterance before speech onset, but did not always complete the following processing step, which was the generation of articulatory programs for the syllables (Roelofs, 2002a). Thus, flexibility in the choice of planning units may arise at the articulatory rather than the phonological level. We will return to these options in the General discussion (for a Discussion of speech planning units in reading see Kawamoto, Kello, Jones, & Bame, 1998; Kawamoto, Kello, Higareda, & Vu, 1999).

The role of response criteria in determining speech latencies

It has long been known that response latencies in psychological experiments are not determined exclusively by the time minimally required to process a stimulus and generate a response. Instead they also depend on the time criteria participants adopt in deciding when to respond (e.g., Lupker, Brown, & Colombo, 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Sanders, 1998; Taylor & Lupker, 2001). In a typical reaction time experiment, participants will quickly determine when they should react in order to meet the requirement to respond fast without making too many errors. During the experiment, the criterion can be continuously updated depending on the difficulty of the stimuli already encountered.

An observation supporting this view is that the time to respond to a stimulus often depends on the difficulty of the preceding stimuli. For instance, Lupker et al. (1997, Experiment 3) found a word frequency effect when readers named English words with irregular grapheme–phoneme correspondence. The effect was much stronger (95 ms) when the materials were blocked by frequency than when high and low frequency words were mixed (46 ms). In the mixed blocks, the high frequency words were produced more slowly and the low frequency ones faster than in the pure blocks. In accounting for this pattern, Lupker et al. made two assumptions. The first was that the articulatory code for a word did not suddenly become available but was grad-

ually built up. The second assumption was that participants aiming for acceptable speed and an acceptable level of accuracy set a time criterion for when to begin to speak. In pure blocks, in which all stimuli were of similar difficulty, optimal criteria could be set for each stimulus type. Thus, an earlier criterion was set for the high frequency words than for the low frequency ones. By contrast, in mixed sets, an intermediate criterion was set, which was not optimal for either type of stimuli but acceptable for both. As Lupker et al. pointed out (see also Taylor & Lupker, 2001), these criteria were flexible, in that they could be continuously adjusted based on further experience with the stimuli, and in that they did not determine the response onset on all trials. For instance, occasionally the articulatory code for a word might be available well before the set time, which might trigger an early response. On other trials, processing might not have advanced far enough to permit articulation when the criterion time was reached such that articulation could only be initiated later. In other words, the external time criterion was one of several variables affecting response speed (for a different account of blocking effects see Kello & Plaut, 2000; Kello, Plaut, & MacWhinney, 2000).

Though Lupker et al. (1997) and Taylor and Lupker (2001) discussed word reading, their argument applies to performance in all speeded tasks. Therefore, a word length effect on object naming latencies should be more likely to be detected in blocks in which all object names have the same length than in blocks including objects with long and short names. In pure blocks, speakers could set tailored response criteria, which would give them just enough time to optimally prepare for each word type. By contrast, in mixed blocks, they would use an intermediate criterion, which would be acceptable for both types of response words but not optimal for either of them.

Speakers could use different strategies to meet varying response criteria. Kello and Plaut (2000, see also Kello et al., 2000) proposed that participants in speeded tasks could strategically increase or decrease their rate of stimulus processing, response preparation, and execution. We will return to this proposal in the General Discussion. Alternatively, speakers may initiate responses on the basis of more or less complete phonological or articulatory representations. Thus, in addition to the inter-individual differences in the preferred coordination of speech planning and speech output invoked by Schriefers and Teruel (1999), there may be intra-individual variability (see also Ferreira & Swets, 2002). In pure blocks of monosyllabic or disyllabic items, speakers would have time to generate complete phonological and articulatory representations for long and short words. However, in mixed blocks they would fully plan the short words, but would initiate long words on the basis of a partial phonological representation or after having retrieved only the first syllable program.

Overview of the experiments

The main empirical question addressed in the experiments described below was whether naming latencies would be longer for objects with disyllabic names than for objects with monosyllabic names. Given the inconsistency of the existing evidence, it seemed appropriate to examine again whether a word length effect could be readily observed. We were interested in the word length effect because our working model of lexical access (Levelt et al., 1999) and the associated computational model WEAVER++ (Roelofs, 1997a, b) predict such an effect and because the presence or absence of a word length effect reveals how speakers co-ordinate speech planning and speech output.

In all experiments, we presented objects with long and short names in pure and in mixed blocks. We included pure blocks in order to maximize our chances of observing a word length effect, should it exist. We included mixed blocks in order to examine whether stimulus blocking has the same effect in picture naming as it has been reported to have in reading. An interaction of blocking and length effects would be of methodological interest as it would demonstrate that object naming latencies are not indicators of invariant stimulus processing and response preparation processes. It would also be of theoretical interest as it would show that speakers can co-ordinate speech planning and speech output in different ways and that their preferred co-ordination can be experimentally manipulated.

In Experiments 1 and 3, speakers named single objects in bare nouns. These experiments differed in their design, as will be explained below. Experiment 2 was a control experiment carried out to demonstrate that the latency difference for objects with long and short names discovered in Experiments 1 and 3 arose during object naming rather than object recognition. In Experiment 4, the participants named object pairs and we examined whether we would obtain a word length effect when the targets appeared as the first word of phrases such as “cat and chair.” In this experiment, we also recorded how long the participants looked at the objects before naming them. We will explain the reasons for carrying out this experiment below.

Experiment 1

Method

Participants. All experiments were carried out with paid participants, who were undergraduate students of Nijmegen University and native speakers of Dutch. There were 16 participants in Experiment 1.

Materials. We selected 115 pictures of objects with monosyllabic or disyllabic names from the picture gal-

lery available at the Max Planck Institute. In a pre-test, 14 participants, who did not take part in the main experiments, named these objects. We chose 16 objects with monosyllabic and 16 objects with disyllabic names for which at least 11 participants agreed on a name. The two sets of objects were matched for average name frequency (mean word form frequencies per million in the CELEX data basis for monosyllabic targets: 7.0, $SD = 1.8$; for disyllabic targets: 6.45, $SD = 2.29$) and were pair-wise matched for the initial segment. The disyllabic object names were stressed on the first syllable and were monomorphemic (see Appendix A). In addition to these experimental pictures, there were eight practice pictures, four of which had monosyllabic and four disyllabic names. The pictures were digitized and scaled to just fit into virtual frames of five by five degrees of visual angle when viewed from a distance of 65 cm. They were presented as black line drawings on a gray background.

Design. The materials were tested under two conditions: in pure blocks, where all object names had the same length, and in mixed blocks, where half the object names were monosyllabic and half disyllabic. Four lists of materials were created. The first list included the targets with monosyllabic names, the second those with disyllabic names, and the third and fourth list (the mixed lists) each included eight targets with monosyllabic and eight targets with disyllabic names.

Blocking by length was tested between participants. A between-participants design was used to render the mixed condition fully comparable with the experiments by Bachoud-Lévi et al. (1998), where all participants received the mixed condition only. Eight participants were only tested on the monosyllabic and the disyllabic lists, with four participants beginning with the monosyllabic list and four with the disyllabic list. Eight other participants were only tested on the two mixed lists, with four persons working on the lists in each order. Each list was tested three times before another list was tested. The order of presenting the objects within the lists was random and different for each presentation of a list, except that the practice items were always tested first.

Apparatus. The experiments were controlled by a Compaq 486 computer. The pictures were presented on a ViewSonic 17PS screen. The participants' speech was recorded using a Sennheiser ME400 microphone and a SONY DTC55 DAT recorder. Speech onset latencies were measured using a voice key.

Procedure. Participants were tested individually in a quiet room. At the beginning of the session they read the instructions. They received a booklet displaying the 20 objects of the first list (16 experimental and four practice objects) together with the expected names. They were told that only these objects would be shown and that they should name them as rapidly as possible using the names printed in the booklet. Then the first block of

trials began. There were short breaks after 20 and 40 trials (i.e., after the first and second presentation of the materials) and a longer break after 60 trials. During this break the participants received another booklet displaying the materials of the second list, which was subsequently tested in the same way as the first list.

At the beginning of each test trial a fixation point was shown for 800 ms. Following a blank interval of 350 ms, an object was shown for 1500 ms. After another blank interval of 350 ms the next trial began. Naming latencies were measured from picture onset.

Results

Errors. On 3.5% of the trials participants stuttered, repaired their utterance, or used an incorrect object name. These trials were excluded from the analyses of speech onset latencies. Analyses of variance showed that the error rates were not systematically affected by the experimental variables.

Latencies. Table 1 displays the mean speech onset latencies for monosyllabic and disyllabic words in the pure and mixed conditions. Overall, there was a word length effect of 20 ms (means: 573 vs. 593 ms) favoring the monosyllabic words. This effect was significant by participants, $F(1, 14) = 9.72$, $p < .01$, and marginally significant by items, $F(2, 30) = 2.97$, $p < .10$. Most importantly, the interaction of blocking and length was significant, $F(1, 14) = 5.88$, $p < .05$, $F(2, 30) = 4.70$, $p < .05$, with the length effect being strong in the pure condition (36 ms) and very weak in the mixed condition (4 ms). Analyses of simple effects showed that only the former effect was significant, $F(1, 14) = 15.36$, $p < .05$, $F(2, 30) = 5.92$, $p < .05$.

Discussion

In pure blocks, the naming latencies were longer for the pictures with disyllabic names than for those with monosyllabic names. Our model of word form retrieval predicts this result, as successive syllables of a word are assumed to be generated in sequence and the articulation of a word should only begin after the phonological form of the entire word has been generated and the corresponding articulatory commands have been retrieved.

There was no word length effect in the mixed blocks. The naming latencies for the monosyllabic words were longer in mixed than in pure blocks, while the latencies for disyllabic words were shorter in mixed than in pure blocks. We can explain this convergence of the naming latencies by assuming that the speakers used different response deadlines in pure monosyllabic, pure disyllabic, and mixed blocks. The latencies for the monosyllabic targets were affected more strongly by blocking than the latencies for the disyllabic targets, though the blocking

Table 1
Results of Experiment 1

Condition	Latencies		Errors %
	<i>M</i>	<i>SD</i>	
Pure			
Monosyllabic targets	561	35	2.6
Disyllabic targets	597	52	3.4
Length effect	36		
Mixed			
Monosyllabic targets	585	62	3.4
Disyllabic targets	589	53	4.7
Length effect	4		

Means (*M*) and standard deviations (*SD*, by participants) of speech onset latencies and error rates (%) for monosyllabic and disyllabic targets in the pure and mixed conditions.

effect was not significant for either type of targets.¹ Taylor and Lupker (2001) noted that the latencies for “fast” stimuli are often affected more strongly by blocking than the latencies for “slow” stimuli. The likely reason is that participants can always react more slowly to relatively easy stimuli in the mixed than in the pure condition, whereas they may not always be able respond any faster to difficult stimuli in the mixed condition than they do in the pure condition.

If participants continuously update their response criterion depending on the difficulty of the items encountered before, one might expect the latency on a given trial to depend on the difficulty of the stimuli presented on the immediately preceding trials. Taylor and Lupker (2001) showed that this was indeed the case for their reading data. Readers responded faster on “fast” word trials immediately preceded by other word trials than on word trials preceded by non-word trials, and they responded more slowly on “slow” non-word trials preceded by other non-words trials than by word trials. We carried out post-hoc analyses of the data obtained in the mixed blocks to examine sequential effects. However, we obtained no systematic first- or second-order sequential effects. The latencies for monosyllabic targets preceded by mono- or disyllabic items were exactly the same (589 ms), whereas the latencies for disyllabic targets were slightly faster (by 9 ms) when the preceding item was disyllabic than when it was monosyllabic. Taylor and Lupker’s analyses were based on approximately 1600 observations per cell, whereas our cell frequencies were less than 200. Thus, sequential

¹ This pattern of results was not replicated in the following experiments.

effects in picture naming may be discovered in more powerful designs. Alternatively, speakers may simply not update their response criterion on a trial-by-trial basis but use a stable criterion for a block of trials (see Lupker et al., 1997).

If speakers set different response criteria in mixed and pure blocks, one may expect the error rates for the more difficult items to be higher in mixed than in pure blocks, whereas the reverse should be true for the easier items. As Table 1 shows, we did not find such a pattern. Instead the error rates for both target types were slightly higher in the mixed than in the pure blocks, though this difference did not approach significance. Apparently, the participants “played safe” in both conditions, adjusting their response criterion only as much as possible while maintaining a high level of accuracy (see Taylor & Lupker, 2001, for a similar conclusion).

Experiment 2

As Bachoud-Lévi et al. (1998) pointed out, the length of the names of objects is likely to be correlated with the frequency of the names and the familiarity of the objects, and these latter variables may be the true source of a naming latency difference observed for objects with short and longer names. Our materials were controlled for word frequency, but object familiarity ratings from a study by van Schagen, Tamsma, Bruggemann, Jackson, and Michon (1982) were only available for three monosyllabic and four disyllabic items. Seven other monosyllabic and seven disyllabic targets were very similar to items in that study. The mean familiarity ratings for these ten monosyllabic and 11 disyllabic items were 2.81 and 2.55 on the five-point familiarity scale, respectively.

To obtain additional evidence about the ease of recognizing the objects with monosyllabic and disyllabic names, we carried out a control experiment using an object/non-object categorization task, which has been used in other research on picture naming and categorization (e.g., Humphreys, Lamote, & Lloyd-Jones, 1995; Meyer, Sleiderink, & Levelt, 1998). In addition to the experimental items of Experiment 1, the participants saw drawings of pseudo-objects, i.e., entities that could be real objects but are judged by most people not to be. The participants categorized each drawing as representing or not representing an existing object by pressing one of two response buttons. To carry out this task, participants must recognize the objects, but they do not have to retrieve their names. If the objects with monosyllabic and disyllabic names differ in the ease of recognition, their categorization latencies should differ. By contrast, if they only differ in the time needed to encode their names, there should be no difference in categorization times.

Method

Participants. The experiment was carried out with 16 participants who had not taken part in Experiment 1.

Materials. The experimental pictures were the same as in Experiment 1. As foils we used 40 drawings of pseudo-objects, also available in the picture gallery of the Max Planck Institute. The pseudo-objects were closed figures with an object-like appearance. Most of them were simplified versions of the pseudo-objects provided by Kroll and Potter (1984). The objects and pseudo-objects were matched for size and number of black pixels.

Design. Four lists of materials were created. Each list included the practice and target objects of one list of Experiment 1. Thus, one list included the 20 objects with monosyllabic names (four practice objects and 16 experimental objects), another list included the 20 objects with disyllabic names, and each of the remaining two lists included 10 objects with monosyllabic names and 10 objects with disyllabic names. The pseudo-objects were arbitrarily divided into two lists of 20 items, each consisting of four arbitrarily selected practice items and 16 experimental items. One 20-item-list was added to the list of objects with monosyllabic names and the other to the list of objects with disyllabic names. Finally, 10 pseudo-objects from each list were added to each mixed list. Thus, each pseudo-object appeared once in a pure and once in a mixed list, and each list included 20 objects and 20 pseudo-objects.

As in Experiment 1, there were two groups of participants. One group saw the two mixed lists and the other group saw the two pure lists. Each list was tested three times before another list was tested. Four participants in each group saw the two lists in each order. The order of presenting the items within a list was random and different for each repetition of a list. The practice items preceded the experimental items.

Apparatus. The same equipment was used as in Experiment 1, except that a two-button push-button panel was installed, which the participants used to indicate whether the drawings represented objects or pseudo-objects.

Procedure. The participants were instructed that they would see drawings of objects and pseudo-objects, i.e., things that resembled real objects but did not exist. For each drawing, they should decide as quickly as possible whether it represented an existing object by pressing the yes-button (right) or the no-button (left) of the two-button panel. As in Experiment 1, the participants received booklets showing all drawings used in the upcoming block of trials. However, the booklets did not include the names of the objects.

Results and discussion

Errors were “object” responses to pseudo-objects, “pseudo-object” responses for objects, and missing

responses. The error rates were 2.9% for pseudo-objects and 4.0% for objects. Analyses of variances showed that the error rates for objects or pseudo-objects did not differ significantly across experimental conditions. Incorrect responses were excluded from the analyses of categorization latencies.

The mean latency for the categorization of pseudo-objects was 453 ms (448 ms in the mixed and 457 in the pure condition). The latencies for the categorization of the objects are shown in Table 2. The latencies for objects with monosyllabic and disyllabic names differed by 11 ms, favoring the objects with shorter names. This difference did not approach significance, $F(1, 14) = 1.78$, $F(1, 30) < 1$. The effect of blocking and the interaction of blocking and length were not significant either.

The absence of a significant latency difference between objects with monosyllabic and disyllabic names in this experiment suggests that the length effect obtained in Experiment 1 arose during the retrieval of the object names rather than during object recognition. To further examine whether differences in the ease of object recognition contributed to the effects found in Experiment 1, we carried out by-item analyses of covariance, using the item-means of Experiment 1 as the dependent variable and the item means of Experiment 2 as co-variables. There were two co-variables per item, which were the means obtained in the pure and mixed conditions of the recognition experiment. In these analyses, the significant length effect in the pure condition was maintained, $F(1, 29) = 5.64$; $p < .05$, while there was no length effect in the mixed condition, $F(1, 29) < 1$. Thus, it appears that the objects with monosyllabic names were named faster in the pure condition of Experiment 1 than the objects with disyllabic names because the shorter names could be encoded more quickly, not because the objects with shorter names were faster to recognize.

Table 2
Results of Experiment 2

Condition	Latencies		Errors %
	<i>M</i>	<i>SD</i>	
Pure			
Monosyllabic targets	452	32	3.9
Disyllabic targets	461	29	3.9
Length effect	9		
Mixed			
Monosyllabic targets	441	56	2.3
Disyllabic targets	454	51	5.9
Length effect	13		

Means (*M*) and standard deviations (*SD*, by participants) of the categorization latencies and error rates (%) for monosyllabic and disyllabic targets in the pure and mixed conditions.

Experiment 3

In Experiment 1, a between-participants design was used to render the mixed condition fully comparable to the experiments by Bachoud-Lévi et al. (1998), where all participants received the mixed condition only. The aim of Experiment 3 was to replicate *within participants* the critical finding of the effect of blocking observed in Experiment 1. If it can be shown that the length effect comes and goes for the same group of participants as a function of blocking, the claim that blocking plays a critical role in determining response latencies would be considerably strengthened.

Method

Participants. The experiment was carried out with 24 participants, who had not participated in Experiments 1 or 2.

Materials, procedure, and apparatus. These were the same as in Experiment 1.

Design. The materials were again tested in pure blocks, where all object names had the same length, and in mixed blocks, where half the object names were monosyllabic and half disyllabic. There were two groups of 12 participants each. In contrast to Experiment 1, where blocking was tested between participants, all participants were initially, i.e., in the control phase of the experiment, tested on pure lists. In the second half of the experiment (the test phase), 12 participants (group 1) continued seeing pure lists, whereas 12 other participants (group 2) worked on mixed lists. Thus, there were three crossed independent variables: group (pure only versus pure, then mixed), phase (control versus test), and length of targets (monosyllabic versus disyllabic).

Using the stimuli selected for Experiment 1, eight lists of materials were created. Lists 1 and 2 each included eight targets with monosyllabic names. Lists 3 and 4 each included eight items with disyllabic names. List 5 included all 16 items with monosyllabic names and list 6 all 16 items with disyllabic names. Finally, list 7 included list 1 (monosyllabic targets) and list 3 (disyllabic targets), and list 8 included list 2 (monosyllabic targets) and list 4 (disyllabic targets). Lists 5 to 8 were identical to the lists used in Experiment 1.

In the control phase of the experiment, all participants were tested on the four pure lists (lists 1–4). Each list was tested in a different block of trials. Each object was repeated four times within a block of trials. The order of presenting the lists was counterbalanced across participants. In the test phase of the experiment, the participants of group 1 were tested on the two remaining pure lists (lists 5 and 6), whereas the participants of group 2 were tested on the two remaining mixed lists (lists 7 and 8). Again, each list was tested in a different block of trials, and each object was repeated four times within a trial block.

Results and discussion

Errors. On 1.1% of the trials participants stuttered, repaired their utterance, or used an incorrect object name. These trials were excluded from the analyses of speech onset latencies. Analyses of variance showed that the error rates were not systematically affected by the experimental variables.

Latencies. Table 3 displays the mean speech onset latencies for the monosyllabic and disyllabic words in the control and test phase for each group of participants. Overall, there was a word length effect of 20 ms favoring the short items, $F(1, 22) = 51.64$, $p < .001$, $F(1, 30) = 5.47$, $p < .05$. There was a 28-ms-effect of group, favoring the group working on pure lists only, $F(1, 22) = 3.85$, $p < .062$, $F(1, 30) = 78.34$, $p < .001$. This group difference was observed in both phases of the experiment; thus it was independent of the experimental conditions. Contrary to what one might expect, the latencies in the two phases were almost identical, 600 and 598 ms, $F(1, 22) < 1$, $F(1, 30) < 1$. Most importantly, there was a triple interaction between length, group, and phase, $F(1, 22) = 4.50$, $p < .05$, $F(1, 30) = 7.65$, $p < .01$.

In the control phase, where all participants saw pure lists, there was a main effect of length, $F(1, 22) = 25.67$, $p < .001$, $F(1, 30) = 4.87$, $p < .05$. The size of the effect did not differ between the two groups of par-

ticipants (18 and 24 ms; $F(1, 22) < 1$, $F(1, 30) = 1.13$ for the interaction). In the test phase, there was also a main effect of length, $F(1, 22) = 22.41$, $p < .001$, $F(1, 30) = 4.91$, $p < .05$. However, now the length effect was significantly larger for group 1, who received pure lists in the test phase, than for group 2, who received mixed lists, $F(1, 22) = 5.40$, $p < .05$, $F(1, 30) = 4.02$, $p < .054$. The effect sizes were 28 and 9 ms, respectively. Tests of simple effects showed that only the effect found in group 1 was significant, $F(1, 11) = 45.73$, $p < .001$, $F(1, 30) = 9.61$, $p < .01$. Thus we replicated the main result of Experiment 1—the confinement of the length effect to pure blocks—using a within-participants design. In the group working only on pure lists, the length effect was maintained throughout the experiment, whereas in the group switching from pure to mixed lists, the effect disappeared after the switch. This demonstrates that it is possible to induce changes in the speakers' preferred way of preparing for the production of words.

Experiment 4

In Experiment 4, the speakers named object pairs in noun phrase conjunctions such as *peer en schaar* 'pear and scissors.' The left object of each pair, which was to be named first, was one of the objects used in Experiments 1–3. The right objects were newly selected objects. As in the preceding experiments, the independent variables were the length of the target nouns and whether monosyllabic and disyllabic targets were tested in mixed or pure blocks.

One goal of the experiment was to examine whether the speech onset latencies in pure blocks would depend on the length of the first noun. As noted in the Introduction, Wheeldon and Lahiri (1997) showed that speakers producing short sentences such as *ik zoek water* ('I am looking for water') phonologically encoded the complete first phonological word before speech onset. The noun phrase conjunctions of Experiment 4 could be parsed into phonological words in different ways, and it was not clear whether speakers would incorporate the conjunction *en* ('and') into the first phonological word. However, unless speakers systematically combined monosyllabic nouns and *en* to form one phonological word, but did not combine disyllabic nouns and *en*, the first phonological word should be longer when the noun was disyllabic than when it was monosyllabic. Thus, based on the results of Experiments 1 and 3 and on those obtained by Wheeldon and Lahiri, one should expect the speech onset latencies to be longer for disyllabic than for monosyllabic targets. However, Schriefers and Teruel (1999) obtained evidence suggesting that speakers producing adjective–noun phrases often only retrieved the phonological code of the first syllable of the adjective,

Table 3
Results of Experiment 3

Condition	Latencies		Errors %
	<i>M</i>	<i>SD</i>	
Control Phase			
Group 1			
Monosyllabic targets	576	32	.9
Disyllabic targets	594	30	1.4
Length effect	18		
Group 2			
Monosyllabic targets	603	37	1.2
Disyllabic targets	627	47	2.0
Length effect	24		
Test phase (pure vs. mixed)			
Group 1 (pure)			
Monosyllabic targets	571	32	1.2
Disyllabic targets	599	32	1.2
Length effect	28		
Group 2 (mixed)			
Monosyllabic targets	607	44	.3
Disyllabic targets	616	45	.8
Length effect	9		

Means (*M*) and standard deviations (*SD*, by participants) of speech onset latencies and error rates (%) for monosyllabic and disyllabic targets in the pure and mixed conditions.

rather than the codes for both syllables before speech onset. They proposed that speakers might be more likely to use small advance planning units when they generated longer and more complex utterances than when they generated shorter utterances. Small planning units may be preferred over longer ones for longer utterances because other planning processes demand capacity as well. Schriefers and Teruel based their suggestion on the comparison of their own results to those obtained in other studies using different materials and trial structures. In the present series of experiments, we performed a more stringent test of their proposal, as we presented the same set of target objects and elicited either bare nouns (Experiments 1 and 3) or noun phrase conjunctions (Experiment 4).

In addition to the speech onset latencies, we measured how long the speakers looked at the target objects. Earlier research has shown that during the description of scenes and events, the speakers' eye movements and their speech planning processes are tightly coordinated in time (e.g., Griffin & Bock, 2000; Levelt & Meyer, 2000). Speakers almost always look at the objects they name in the order of mention. Moreover, the time spent looking at an object (the viewing time) depends on the time necessary for identifying the object and on the time necessary for planning its name. When speakers name two objects, their gaze usually shifts from the first to the second object *before* the name of the first object is initiated but *after* its phonological form has been retrieved. Evidence supporting this ordering of events comes from studies showing that speakers look longer at objects with low frequency than with high frequency names (Griffin, 2001; Meyer et al., 1998). In addition, Meyer and van der Meulen (2000) found in a picture–word interference experiment that objects to be named were fixated for shorter periods when simultaneously presented auditory distracters were phonologically related to their names than when the distracters were unrelated. These results suggest that speakers generate utterances referring to several objects in a very sequential fashion. Apparently, they focus on one object until they have planned its name at least down to the phonological level (possibly even to the level of articulatory commands) and only then turn to a new object.

Based on these findings we expected parallel results for the speech onset latencies and the viewing times for the targets in the pure blocks of Experiment 4. Latencies and viewing times should be longer for disyllabic than for monosyllabic targets. The results for the mixed blocks were more difficult to predict because it was not known from the earlier studies whether the shift of gaze from one object to the next occurs as soon as a particular internal representation of the target name has been generated, or whether the shift of gaze is bound to the onset of articulation for the object name. Based on the results of Experiments 1 and 3 we did not expect a length

effect on the speech latencies in mixed blocks. If speakers use the same criteria in deciding when to begin to speak and when to shift gaze to a new object, there should not be a length effect for the viewing times either. Alternatively, the decisions when to begin to speak and when to shift gaze could be governed by different criteria. For instance, speakers could begin to speak as soon as they have retrieved the articulatory commands for one syllable but initiate the shift of gaze only after having retrieved the articulatory commands for all syllables of the target word. In that case, there should be a word length effect for the viewing times, but not for the speech onset latencies. Thus, the experimental results should show whether the viewing time for an object depends on the time required to complete the articulatory planning for the object name or is tightly linked to the onset of the articulation of the corresponding word. In addition, they should contribute to our understanding of the effects of stimulus blocking as they should show whether blocking affects only the speakers' decision when to begin to speak or affects their decision to terminate the visual processing of the referent objects as well.

Method

Participants. The experiment was carried out with 16 participants who had not taken part in any of the other experiments reported here.

Materials. The target pictures were the same as in Experiments 1 and 3, but they were now presented in the center of the left half of the computer screen. In addition, 40 new pictures ("right pictures" hereafter) were selected from the picture gallery to be shown on the right side of the display (see Appendix). Twenty of them had monosyllabic names and 20 had disyllabic names. They were very similar in style and complexity to the target pictures and were scaled to fit into a frame of the same size (five by five degrees).

Design. The design was the same as in Experiment 1, except that object pairs were presented instead of single objects. Each participant was tested on two lists of items, either a pure monosyllabic and a pure disyllabic list or two mixed lists. Each list was presented three times before a new list was shown. The right picture accompanying a given target was different in each presentation of the materials. For instance, the target picture *taart* ('cake') was shown together with a chair (*stoel*), a button (*knoop*), and a fish (*vis*). The names of the two objects shown together were semantically and phonologically unrelated, but they had the same number of syllables. We matched targets and right objects for length in order to be able to present the materials in the pure condition, where all pictures to be named should have monosyllabic or disyllabic names.

Apparatus. Eye movements were monitored using an SMI EyeLink-Hispeed 2D head-mounted eye tracking

system (SensoMotoric Instruments GmbH, Teltow, Germany). The position of the right eye was determined every 4 ms. The spatial accuracy of the eye tracker is better than 0.1° .

Procedure. The participants were tested individually. At the beginning of the experiment they read the instructions, which explained to them how the eye movements would be monitored and what their task would be. They were told that they would see pairs of objects, which they should name in noun phrases such as *taart en stoel* ('cake and chair'). The participants received a booklet showing all objects used in the experiment together with the expected names.

When the participants had read the instructions and studied the picture booklet, a training block began. The participants saw the objects of the first list to be tested one-by-one centered in the middle of the screen and were asked to name them. The experimenter immediately corrected all naming errors. The participants were informed that the upcoming three blocks of trials would only include the materials seen in the training block.

After training, the headband of the eye-tracking system was placed on the participant's head and the system was calibrated. For the calibration, a grid of three by three positions had been defined. During a calibration trial a fixation target appeared once, in random order, in each of these positions for one second. The participants were asked to fixate upon each target until the next target appeared. After the calibration trial, the estimated positions of the participant's fixations and the distances from the fixation targets were displayed to the experimenter. Calibration was considered adequate if there was at least one fixation within 1.5° of each fixation target. When calibration was inadequate, the procedure was repeated, sometimes after adjustment of the cameras. Successful calibration was followed by a validation trial. For the participants, this trial did not differ from the calibration trial, but the data collected during the validation trial were used to estimate the

participants' gaze positions, and the error (i.e., the distance between the estimated gaze position and the target position) was measured. The validation was considered adequate if the average error was below one degree and the worst error below 1.5 degree. Depending on the result of the validation trial, the calibration and validation trials were repeated or the main part of the experiment began. Calibration and validation were repeated after each test block.

At the beginning of each test trial, a fixation point appeared in the center of the frame for the left object for 800 ms. Earlier experiments had shown that speakers naming object pairs usually (i.e., on more than 90% of the trials) first looked at the left and then at the right object and named the objects in the same order (e.g., Meyer et al., 1998). This strong tendency to inspect and name the left object first was reinforced by the presentation of the fixation point. Following a blank interval of 200 ms, an object pair was presented for 2000 ms. The participant named the objects in a noun phrase conjunction. After a blank interval of 500 ms the next trial began. There were short breaks after the first and second test block (i.e., after 20 and 40 trials). After the third block (i.e., after 60 trials), the experimenter told the participant that a new set of materials (the second list) would be tested. This new set was then presented in a training block followed by three experimental blocks.

Results

The speakers occasionally used incorrect object names, stuttered or repaired an utterance or began an utterance with a non-speech sound that triggered the voice key. Such responses were coded as errors. As Table 4 shows, there were more errors in mixed than in pure blocks and more for disyllabic than for monosyllabic targets. However, in analyses of variance of the error rates no main effect or interaction approached

Table 4
Results of Experiment 4

Condition	Latencies		Viewing times		Errors %
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Pure					
Monosyllabic targets	685	79	416	44	4.4
Disyllabic targets	719	63	480	53	3.1
Length effect	34		64		
Mixed					
Monosyllabic targets	712	91	470	70	5.2
Disyllabic targets	694	84	480	62	8.1
Length effect	-18		10		

Means (*M*) and standard deviations (*SD*, by participants) of speech onset latencies and viewing times and error rates (%) for monosyllabic and disyllabic targets in the pure and mixed conditions.

significance. Error trials were discarded from the analyses of eye movements and speech onset latencies.

Table 4 shows that in the blocked condition, speech was initiated faster for monosyllabic than for disyllabic targets. The reverse was true in the mixed condition. The analyses of variance revealed no main effect of length (both $F < 1$), but a significant interaction of length and blocking, $F(1, 14) = 10.57$; $F(1, 30) = 11.68$, both $p < .01$. In analyses of simple effects only the 34-ms-length effect in the pure condition reached significance, $F(1, 14) = 8.61$; $p < .02$; $F(1, 30) = 8.79$, $p < .01$. Thus, as in Experiments 1 and 3, a word length effect was obtained only in the pure condition. Monosyllabic targets were produced more slowly and disyllabic ones faster in mixed than in pure blocks, but the effect of blocking was not significant for either type of targets.

To analyze the speakers' eye movements, we first classified their fixations as falling on the left or right object or elsewhere. A fixation was counted as on an object, if its coordinates lay within or on the outer contours of the object. Next, the speakers' gaze patterns were examined. At the beginning of each trial a fixation point had appeared at the location where the left object would appear a little later. On all but one trial, which was eliminated from the further analyses, the speakers were fixating upon the left object at picture onset. Although the speakers had not been specifically instructed to fixate upon the right object, they almost always did so. The data from seven trials on which this was not the case were not analyzed. On the remaining trials, the participants first fixated upon the left and then on the right object.

On average, the participants looked at the left object for 463 ms before turning to the right object. The object-to-object saccade took on average 67 ms. Sometimes, the participants' gaze remained on the right object until the end of the trial, but on 87% of the trials, the speakers looked at the left object again before the end of the trial. This second inspection of the left object began, on average, 1366 ms after picture onset and 903 ms after speech onset. The participants probably returned to the left object to check the correctness of their utterance or to prepare for the next trial, which would begin with the presentation of a fixation point on the left side of the screen. Since we were interested in the inspection of the target objects accompanying utterance planning, these returns to the left object were not analyzed.

Table 4 displays the mean viewing times for the left object. The viewing time was defined as the time interval between the beginning of the first fixation on the left object and the end of the last fixation before the shift of gaze to the right object.² There was a significant main

effect of target length, $F(1, 14) = 20.06$, $F(1, 30) = 9.64$, both $p < .01$. The mean viewing time was shorter by 37 ms for monosyllabic than for disyllabic targets. The interaction of blocking and length was also significant, $F(1, 14) = 10.56$, $F(1, 30) = 11.74$, both $p < .01$. The viewing times for the two types of targets differed by 10 ms in the mixed condition, but by 64 ms in the pure condition. Analyses of simple effects revealed that only the latter effect was significant, $F(1, 14) = 29.85$, $F(1, 30) = 29.52$, both $p < .01$.

Discussion

In Experiment 4, the speech onset latencies were longer than in Experiments 1 and 3, probably because of additional syntactic encoding processes required for the longer utterances, but the pattern of results was very similar. As in Experiments 1 and 3, there was a word length effect, but only in the pure blocks. The finding that the word length effect was maintained in the pure blocks argues against the hypothesis that speakers use smaller phonological planning units when they produce longer utterances than when they produce shorter ones, as Schriefers and Teruel (1999) proposed. In our experiments, the speakers used the same phonological planning units for single words and phrases. It is, of course, possible that speakers use smaller units for longer or more complex utterances.

For the viewing times, we found the same pattern of results as for the latencies. The sensitivity of the viewing times to word length in the pure blocks corroborates earlier findings, which also showed that speakers fixated upon the objects they named at least until they had retrieved the phonological forms of their names (Griffin, 2001; Meyer & van der Meulen, 2000; Meyer et al., 1998). Given the size of the objects on the screen and their distance, it is unlikely that the right object was extensively processed before fixation. Therefore, the late shift of gaze from the left to the right object suggests that the speakers processed the two objects and planned their names in succession. This conclusion contrasts with the view that speakers generate utterances incrementally, which would imply that the lexical retrieval processes for the two names could overlap in time, with, for instance, the phonological form of the first name being retrieved at the same time as the lemma of the second name (for recent discussions of this view see Costa & Caramazza, 2002; Ferreira & Swets, 2002). Perhaps speakers sometimes adopt such a strategy, but they do not seem to do so when naming several objects.

The results obtained for the viewing times were different in pure and mixed blocks. Thus, blocking affected not only the speech onset latencies but another indicator of target processing time as well. The interaction of blocking and length obtained for latencies and viewing

² On average there were 1.95 consecutive fixations on the left object. Gaze duration, i.e., the sum of the fixation durations excluding saccades, correlated with viewing time by $r = .99$.

times shows that the criteria governing the speakers' decision when to begin to speak and when to end the visual inspection of an object were very similar. As noted above, one goal of this experiment was to determine whether the shift of gaze from a target object to a new object was triggered by the completion of an internal representation or whether it was time-locked to the onset of speech. The latter turned out to be the case. This is important for the interpretation of viewing times for objects; they are clearly not any "purer" indicators of the time minimally required to process objects and retrieve their names than latencies are.

If the criteria governing the onset of articulation and the shift of gaze from one object to the next are closely related, the time interval between the shift of gaze and speech onset (gaze–speech-lag, hereafter) should be fairly constant across a variety of conditions. For single-word naming this seems to be true. In the present study, the gaze–speech-lag was 241 ms. In earlier experiments it had been 262 ms (Meyer et al., 1998) and 259 ms (Meyer & van der Meulen, 2000), respectively. However, when speakers produce longer phrases to describe objects (saying, for instance, *the little brown cow* or in Italian *porta rossa* 'door red'), the co-ordination between eye movements and speech onset changes dramatically. In that case, the shift of gaze does not occur *before* speech onset, but well *after* speech onset and approximately 250 ms before the beginning of the last word of the phrase referring to the object (Levelt & Meyer, 2000). Thus, when speakers describe an object, they look at it until they are about to say the last word of the corresponding phrase.

Though the gaze–speech-lag for one-word utterances is fairly constant across conditions, we found in the present study that the lag was shorter by 29 ms for disyllabic than for monosyllabic targets, $F(1, 14) = 14.75$, $F(1, 30) = 19.65$, both $p < .01$. The difference in the lags for long and short words was 30 ms in pure blocks, where main effects of length were found for viewing times and latencies, and 28 ms in mixed blocks, where no significant length effect was obtained for either variable. Griffin (in press) also reported a shorter gaze–speech-lag before disyllabic than monosyllabic words. She suggested that in deciding when to shift gaze speakers took the length of the planned word into account. If it was long, they could afford to look at the target object a little longer than if it was short because the extra time required for the articulation of a long word offered additional planning time for the next part of the utterance. As we discussed above, it is possible that the speakers in both types of blocks computed the entire phonological representation of the target word before speech onset. Thus information about the length of the planned words was available and could be one of the factors determining the timing of the shift of gaze to the next object.

General discussion

The existence of word length effects

Word length effects have been obtained in some earlier production studies, but have been reported to be absent in others, most notably in the study by Bachoud-Lévi et al. (1998). One goal of the present experiments was to examine once more whether speech onset latencies depended on the length of the planned words. In three experiments, we obtained a word length effect, provided that long and short words were tested in separate blocks of trials. Speech onset latencies were longer for long than for short target words when they were produced in isolation and when they appeared at the beginning of a phrase. Given these results, it is difficult to maintain that length effects in spoken word production do not exist, as Bachoud-Lévi et al. (1998) suggested. Perhaps a word length effect could be obtained with their materials as well if long and short words were tested in separate trial blocks.

Meeting response deadlines

The second goal of our experiments was to determine whether the length effect would interact with the effect of stimulus blocking. Given the solid evidence for blocking effects obtained in studies of word reading, one would expect such effects to be present in object naming as well. Yet, to our knowledge, the present study is the first to demonstrate that naming latencies and viewing times for objects depend, among other things, on the difficulty of the items co-occurring in a block of trials. We studied the effects of blocking by word length, but we would expect blocking by other variables affecting the ease of object recognition or name retrieval to have similar effects. In all cases, speakers may set response deadlines tailored to the difficulty of the items encountered in a block of trials. In future experiments, one may want to keep this in mind when assigning materials to test blocks and selecting filler items.

Why was the length effect confined to pure blocks? Our account is based on the assumption of a temporal response criterion (e.g., Lupker et al., 1997; Taylor & Lupker, 2001). When all items of a test block are of similar difficulty, participants can set an optimal criterion, which they can just meet on most trials. By contrast, when the items vary substantially in difficulty, participants may set a criterion that is acceptable for all stimuli, but not optimal for the easiest or the most difficult items. Thus, we obtained a word length effect for the speech onset latencies in pure blocks because speakers set different response criteria for monosyllabic and disyllabic targets. We did not observe an effect in mixed blocks because the participants used an intermediate response criterion. Experiment 4 showed that these

response criteria also governed how long the speakers looked at the target objects.

How did the speakers manage to meet different response deadlines? One possibility is that deadlines affect the overall speed of response preparation processes. Kello and Plaut (2000) and Kello et al. (2000) discussed how participants dealt with experimenter-imposed response deadlines and proposed that they could strategically increase or decrease the speed of the response preparation and execution processes. With increasing response speed, latency differences between experimental conditions decrease and the transition from response preparation to execution becomes more cascaded, i.e., there is less response preparation before and more after the onset of the response. Kello and Plaut (2000) proposed that stimulus blocking effects could arise in a similar fashion. An important implication of this proposal is that blocking should affect response durations as well as latencies. On the basis of the present data we cannot evaluate this prediction.

However, we are not convinced that the model proposed by Kello et al. (2000) offers an entirely satisfactory account of their own data. Kello et al. observed that imposing a response deadline in a color–word Stroop experiment had one critical effect, which was to increase response durations in the incongruent relative to the control condition at a stimulus onset asynchrony of 100 ms. The response durations in the neutral and congruent condition did not differ from each other. Computer simulations by Kello et al. revealed that their model reproduced the observed duration difference between the incongruent and control condition. However, the model also produced differences in response durations between other conditions that were not present in the empirical data. As suggested by Kello et al. it is possible that the expected durational effects was not observed in the empirical data because of physical constraints on the minimal durations of articulatory responses; these constraints were not implemented in the model. In addition, there were discrepancies between the modeled and observed error rates and in the relative sizes of the facilitatory and inhibitory effects on response latencies. In short, it appears to us that the model does not yield an accurate account of the experimental results. In addition, Damian (in press) examined latencies and response durations in a range of commonly used speech production tasks. He replicated the standard latency effects, but failed to obtain corresponding effects on response durations.

Instead of affecting how quickly responses are generated and executed, different response criteria could lead speakers to plan the target words more or less completely before speech onset. For monomorphemic words, there could be incomplete planning at the phonological or phonetic level. As noted in the Introduction, Schriefers and Teruel (1999) argued for flexible

control of the planning units at the phonological level. This hypothesis accounts well for the results obtained in their picture–word interference experiments. It also accounts for the present data. However, the results of several other studies strongly suggest that speakers usually generate at least one phonological word before speech onset. For instance, Meyer and Schriefers (1991) carried out a picture–word interference experiment in which participants named objects with monosyllabic or disyllabic names. Monosyllabic and disyllabic items were mixed. The participants heard phonologically related or unrelated distracter words at variable stimulus onset asynchronies relative to the picture onset. There were facilitatory effects from distracters that shared the first or the last syllable with the disyllabic target words, and from distracters that shared the first two or last two segments with the monosyllabic targets. The last-syllable distracters had to be presented slightly later than the first-syllable ones to be maximally effective. This supports the assumption that the two syllables of the target words were generated in sequence. The finding that begin- and end-related distracters yielded facilitatory effects of comparable size demonstrates that the participants initiated target naming after having retrieved the phonological code of both syllables. Costa and Caramazza (2002) recently reported the results of picture–word interference experiments suggesting that speakers generate the phonological representations of both words of English adjective noun phrases or Spanish noun adjective phrases before speech onset. Results of implicit priming experiments (e.g., Meyer, 1990; Roelofs, 1998) in which speakers could prepare for one or more word-initial syllables, further support the view that speakers rarely initiate an utterance before having generated the form of at least one phonological word. The results of Wheeldon and Lahiri's (1997) study, which we discussed in the Introduction, also support this view. None of these studies shows that it is *impossible* for speakers to use planning units that are smaller than one phonological word. However, they demonstrate that speakers are not strongly inclined to use such units.

Finally, blocking may affect the transition from the phonological to the articulatory code of a word. In our theory of lexical access, which is implemented in *WEAVER++*, the phonological segments of a word are activated in parallel and are syllabified in a sequential left-to-right manner. As soon as a syllable has been generated, the corresponding articulatory program is accessed from memory and stored in an output buffer. So far, we have assumed that articulation only begins when the buffer contains the articulatory programs for a complete phonological word. However, it may be the case that speakers sometimes initiate the articulation earlier, for instance after having recovered the syllable program for the first syllable of a disyllabic word. This is what the speakers may have done in the mixed blocks of

the present experiments. In pure monosyllabic blocks they also selected one syllable program before beginning to speak, whereas in pure disyllabic blocks they selected two syllables. Such flexibility at the articulatory level might occur if the speech monitoring system can access the amount of information stored in the buffer, which is compatible with current theories of speech monitoring (e.g., Hartsuiker & Kolk, 2001; Postma, 2000).

WEAVER++ simulations

Santiago et al. (2002) rejected the proposal that word length effects may be hidden by a response criterion operating at the level of phonetic encoding on logical grounds: “If a whole phonological word must be planned before starting pronunciation, delays introduced before phonetic encoding should cause longer latencies for longer words. Phonetic encoding may increase these processing delays even further (access of all phonetic syllables), or may introduce a constant delay (due to accessing the first syllable) for words of different lengths. However, phonetic encoding surely cannot supersede prior delays introduced at the phonological encoding level” (p. 27). However, this argument does not take into account that phonological and phonetic encoding can occur in parallel, as they do in WEAVER++. In the model, the motor program for the first syllable of a disyllabic word is retrieved as soon as the first phonological syllable has been planned. The retrieval of the first syllable motor program and the planning of the second phonological syllable occur in parallel, but the retrieval of the first syllable motor program is the slower process. Therefore, a length effect will only be obtained when the initiation of articulation depends on the retrieval of the motor programs for the first *and* second syllable.

WEAVER++ simulations confirmed that a response criterion operating at the level of phonetic encoding can indeed hide a word length effect. The simulations used the same procedures and parameter values as earlier simulations of the model (i.e., Roelofs, 1997a). In pure blocks, all syllable motor programs were retrieved before the initiation of articulation, whereas in mixed blocks, the articulation of disyllabic words was initiated after retrieval of the first syllable program. In both pure and mixed blocks the full phonological word was planned before articulation onset. Fig. 1 shows that WEAVER++ captures the effect of blocking observed in the current study as there is a length effect in pure but not in mixed blocks. The empirical effects are means across Experiments 1, 3, and 4. There is good agreement between WEAVER++ and the empirical data. The simulations demonstrate that a word length effect can arise as long as the criterion for response initiation is that the phonological and articulatory code for the entire word have been generated. However, the length effect disap-

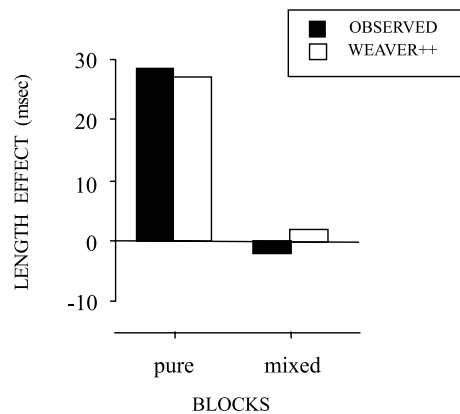


Fig. 1. The length effect (disyllabic versus monosyllabic words) as a function of block type (pure versus mixed). Observed effects (means across Experiments 1, 3, and 4) and WEAVER++ simulations. In the simulations, a phonological word was planned before articulation onset in both pure and mixed blocks. In the pure blocks, all syllable motor programs were accessed before articulation onset, whereas in the mixed blocks, only the first syllable motor program was accessed.

pears if responses are initiated as soon as one syllable program has been retrieved even if the complete phonological representation is generated before speech output. (When planning the complete phonological word takes longer than retrieving the motor program for the first syllable, a word length effect on speech onset latencies should occur. In future research, this prediction may be tested by comparing speech onset latencies for monosyllabic words and words with three or more syllables.)

The proposal that speakers generally complete the phonological encoding of the first word of their utterance but may select a single syllable as the articulatory planning unit also accounts for the data obtained by Schriefers and Teruel (1999). Simulations showed that WEAVER++ yields facilitation from related first and second syllable distracters if articulation is initiated after the phonological representation and the articulatory programs for the entire word have been generated. By contrast, there is only a first syllable priming effect if articulation is initiated upon planning of the complete phonological representation of the word and retrieval of the articulatory program for the first syllable. How quickly syllable motor programs are retrieved in the model depends on their level of activation. Priming may speed up the retrieval of syllable motor programs by enhancing activation levels. However, when the initiation of articulation only depends on the retrieval of the first syllable motor program, rather than on the retrieval of both programs, there will only be a priming effect of the first syllable of a disyllabic target word.

Summary

The main goal of our study was to determine whether naming latencies and viewing times would be longer for objects with long names than for objects with short names, as our model predicts. This turned out to be the case when long and short words were tested in separate blocks. We conclude that under some circumstances at least, speakers fully plan the word they are about to say at the phonological and articulatory level before beginning to speak and initiating the shift of gaze to a new object. A second goal was to examine the effects of stimulus blocking. We found that blocking systematically affected object naming latencies and viewing times, paralleling results obtained in studies of word reading. We propose that when objects with long and short names were presented in separate blocks, speakers set their response criterion such that they could fully plan the object names at all levels, including the level of articulatory programming, before beginning to speak and turning to a new object. By contrast, when objects with long and short names were mixed, the response criterion was set such that the articulatory planning for the longer words could often not be completed before the articulation and the shift of gaze were initiated. Consequently, no word length effect was observed.

Acknowledgments

We thank Frouke Hermens and Otto Kokke for running the experiments and Zenzi Griffin, Christopher C. Kello, and an anonymous reviewer for valuable comments on an earlier version of this paper.

Appendix A. Picture names in Experiments 1–4

Monosyllabic targets: bijl (axe), boor (drill), eend (duck), hert (deer), kaars (candle), kar (cart), kraan (crane), muis (mouse), pauw (peacock), peer (pear), rits (zipper), schaats (skate), spook (ghost), taart (cake), vork (fork), wiel (wheel).

Disyllabic targets: bezem (broom), borstel (brush), cactus (cactus), eekhoorn (squirrel), hamer (hammer), kachel (heater), ketting (chain), masker (mask), pinguin (penguin), puzzel (puzzle), robot (robot), schommel (swing), spijker (nail), trommel (drum), vlinder (butterfly), wortel (carrot).

A.1. Right objects of Experiment 4

Monosyllabic items: bank (couch), bed (bed), bloem (flower), deur (door), knoop (knob), kruik (warm water bottle), neus (nose), pan (casserole), rok (skirt), rups (caterpillar), schaar (scissors), stoel (stool), tent (tent), vis (fish), wieg (crib), wolk (cloud).

Disyllabic items: anker (anchor), appel (apple), ballon (balloon), banaan (banana), ezel (donkey), gieter (watering can), kano (canoe), kanon (canon), kikker (frog), ladder (ladder), lepel (spoon), pistool (pistol), potlood (pencil), sleutel (key), tafel (table), trompet (trumpet).

References

- Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, 39, 331–346.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, 46, 178–198.
- Damian, M. (in press). Articulatory duration in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1–29.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123–147.
- Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanisms in perceptual encoding? *Journal of Experimental Psychology*, 84, 502–507.
- Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57–84.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82, B1–B14.
- Griffin, Z. M. (in press). Pushing the limits of incremental word production. *Psychology Bulletin and Review*.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.
- 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.
- Humphreys, G. W., Lamote, C., & Lloyd-Jones, T. J. (1995). An interactive activation approach to object processing: Effects of structural similarity, name frequency, and task in normality and pathology. *Memory*, 3, 535–586.
- Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 862–885.
- Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. V. Q. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 362–381.

- Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 719–750.
- Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General*, 129, 340–360.
- Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading, reconsidered. *Journal of Experimental Psychology*, 100, 368–374.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23, 39–66.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223–232.
- Levelt, W. J. M., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, 12, 433–452.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 570–590.
- MacKay, D. G. (1987). *The organisation of perception and action*. New York: Springer.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables. *Journal of Memory and Language*, 29, 524–545.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.
- Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming of picture viewing and picture naming. *Psychological Bulletin & Review*, 7, 314–319.
- 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, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66, B25–B33.
- Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 452–467.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97–131.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460–490.
- Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, 35, 854–876.
- Roelofs, A. (1997a). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.
- Roelofs, A. (1997b). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, 12, 659–696.
- Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb–particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 904–919.
- Roelofs, A. (2002a). Spoken language planning and the initiation of articulation. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, 55A, 465–483.
- Roelofs, A. (2002b). Syllable structure effects turn out to be word length effects: Comment on Santiago et al. (2000). *Language and Cognitive Processes*, 17, 1–13.
- Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skills*. Mahwah, NJ: Erlbaum.
- Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognitive Processes*, 15, 1–44.
- Santiago, J., MacKay, D. G., & Palma, A. (2002). Length effects turn out to be syllable structure effects: Response to Roelofs (2002). *Language and Cognitive Processes*, 17, 15–29.
- Schriefers, H., & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, 11, 17–50.
- Sevold, C. A., & Dell, G. S. (1994). The sequential cueing effect in speech production. *Cognition*, 53, 91–127.
- Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Erlbaum.
- Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 117–138.
- van Schagen, I., Tamsma, N., Bruggemann, F., Jackson, J. L., & Michon, J. A. (1982). *Namen en normen van plaatjes. [Names and norms for pictures]*. Heymans Bulletins Psychologische Instituten R.U. Groningen. HB-82-571-EX (internal report, Groningen University).
- van Turenhout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysical evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787–806.
- Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, 41, 101–175.
- Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37, 356–381.
- Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.