Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 4, 892–905

0278-7393/10/$12.00 DOI: 10.1037/a0019446

Does Word Length Affect Speech Onset Latencies When Producing Single Words?

Markus F. Damian and Jeffrey S. Bowers
University of Bristol

Hans Stadthagen-Gonzalez
Bangor University

Katharina Spalek
Humboldt-Universität zu Berlin

Most models of spoken production predict that shorter utterances should be initiated faster than longer ones. However, whether word-length effects in single word production exist is at present controversial. A series of experiments did not find evidence for such an effect. First, an experimental manipulation of word length in picture naming showed no latency differences. Second, Dutch and English speakers named 2 sets of either objects or words (monosyllabic names in Dutch and disyllabic names in English or vice versa). A length effect, which should manifest itself as an interaction between object set and response language, emerged in word naming but not in picture naming. Third, distractors consisting of the final syllable of disyllabic object names speeded up responses, but at the same time, no word-length effect was found. These results suggest that before the response is initiated, an entire word has been phonologically encoded, but only its initial syllable is placed in an articulatory buffer.

Keywords: spoken production, psycholinguistics, word length

Spoken language is inherently sequential. Apart from overt speech itself being serially ordered, most current models of language production converge on the assumption that the abstract assembly of speech sounds involves a sequential element; that is, the phonological content of speech (consisting of, e.g., phonemes, or syllables) is generated in the order in which it will be articulated (see, e.g., Levelt & Wheeldon, 1994; Meyer, 1990; Meyer & Schriefers, 1991; Roelofs, 2004; Sevald & Dell, 1994; van Turennout, Hagoort, & Brown, 1997, for evidence supporting the sequentiality of phonological encoding). For instance, Wheeldon and Levelt (1995) presented English words to Dutch participants with good knowledge of English and asked them to silently generate the Dutch translation and to monitor the internally generated speech for a particular target phoneme. By varying the position of the segment to be monitored, Wheeldon and Levelt demonstrated that reaction times steadily increased for later targets, suggesting that the internal code became incrementally available (see also Özdemir, Roelofs, & Levelt, 2007; Wheeldon & Morgan, 2002, for similar findings).

A more controversial issue concerns the degree to which speakers plan ahead at the phonological level before they initiate a response. *Minimalist* theories of planning (e.g., Dell, Juliano, & Govindjee, 1993; Jordan, 1990; MacKay, 1987) stipulate that only a very small degree of advance planning is carried out. Theories of this type assume that as soon as the initial portion of an utterance (i.e., its first segment) becomes available, articulation starts. By contrast, *nonminimalist* accounts (e.g., Levelt, Roelofs, & Meyer, 1999; Wheeldon & Lahiri, 1997) assume that the portion of the utterance that is buffered before it is articulated is considerably larger. The unit of advance planning assumed by these accounts is the *phonological word*, which is defined minimally as a stressed foot and maximally as a single lexical word plus associated unstressed function words, such as auxiliaries, determiners, conjunctions, and prepositions (cf. Wheeldon, 2000). Wheeldon and Lahiri (1997) stated that "when it is possible to do so, speakers preferentially initiate articulation following the phonological encoding of the initial phonological word of an utterance" (p. 377). Of importance, this account specifies a lower boundary on the degree of advance planning but leaves open the possibility that speakers may opt to encode more than a single phonological word (see Alario, Costa, & Caramazza, 2002a, 2002b; and Levelt, 2002, for a clarifying debate on how the scope of advance planning should be defined).

If phonological encoding indeed proceeds incrementally and if speakers encode at least an entire word before they start the corresponding articulation, then a word-length effect is predicted in single word production tasks, such as object naming: All else being equal, longer words should take longer to prepare than shorter words. Indeed, a number of previous studies have investigated potential effects of word length in spoken production. However, as we outline later, the existing results are complex, and their

---

Markus F. Damian and Jeffrey S. Bowers, Department of Experimental Psychology, University of Bristol, Bristol, England; Hans Stadthagen-Gonzalez, Economic and Social Research Council Centre for Research on Bilingualism in Theory and Practice, Bangor University, Bangor, Wales; Katharina Spalek, Department of German Language and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany.

This research was supported by Grant BB/C508477/1 from the Biotechnology and Biological Sciences Research Council to Markus F. Damian.

Correspondence concerning this article should be addressed to Markus F. Damian, University of Bristol, Department of Experimental Psychology, 12a Priory Road, Bristol BS8 1TU, United Kingdom. E-mail: m.damian@bristol.ac.uk

interpretation is made difficult by possible confounding factors. The current research takes another look at this issue.

Review of Past Research

One approach to characterizing the linguistic variables that affect speech production is to conduct multiple regression analyses on the results of picture-naming tasks when potentially relevant variables (e.g., frequency, name agreement, etc.) are included as predictors. In most analyses of this type, word length does not emerge as a significant predictor (e.g., Alario et al., 2004; Barry, Morrison, & Ellis, 1997). However, this null finding does not constitute unambiguous evidence that the length of an utterance does not affect spoken responses: Word length is strongly confounded with a number of other relevant variables, most notably with age of acquisition (AoA; short words tend to be learned earlier in life than longer ones) and frequency (short words tend to occur more often than longer ones). For instance, Alario et al. (2004) reported correlations of .292 between AoA and number of phonemes and correlations of .251 between AoA and number of syllables. The reported correlation between frequency and number of phonemes was −.219, and the correlation between frequency and number of syllables was −.251. In such collinear regressor sets, the unique contribution of each regressor is likely to be an underestimate of its true contribution, and as a result word length may not emerge in the results, despite genuinely contributing to naming latencies.

An alternative approach is to investigate the effects of word length by manipulating it as an experimental factor. Here, initial reports indicated positive results. Klapp, Anderson, and Berrian (1973; Experiment 4) compared naming times of objects with mono- and disyllabic labels and found that the former were named slightly faster (14 ms) than the latter (see also Eriksen, Pollack, & Montague, 1970, for a similar effect in the naming of two-digit numbers). However, as subsequently pointed out by Bachoud-Lévi, Dupoux, Cohen, and Mehler (1998), these early studies did not control for various confounding factors, such as frequency, ease of conceptual access, and so forth; hence, the results should be interpreted with caution. Bachoud-Lévi et al. reported results from a series of studies that controlled for such potential confounds. In five experiments, English or French speakers served as participants. The stimuli consisted either of pictures or of arbitrary symbols (e.g., *&&*) that participants were trained to respond to with words of varying lengths. In none of the experiments did word length exert a significant effect on response latencies. The authors interpreted their pattern of results as suggesting that speakers initiated their response before the entire word had been phonologically encoded, hence rendering latencies for short and long responses largely equivalent.

This conclusion has been challenged by more recent research. Meyer, Roelofs, and Levelt (2003) investigated word-length effects in picture-naming experiments in which stimuli with mono- and disyllabic labels were either randomly intermixed or were produced in separate (*pure*) blocks. The motivation for doing so stems from reports of *list composition effects* in related domains (e.g., Lupker, Brown, & Colombo, 1997), which are thought to reflect decision processes that are not intrinsic to the language system. According to this view, when items of varying difficulty are intermixed, a response criterion is set to an intermediate position, which acts to minimize underlying differences between conditions. On the other hand, if the two types of items are presented separately, the response criterion can be adjusted accordingly, and any real differences are more easily uncovered. Meyer et al. found no word-length effect in the intermixed condition; crucially, however, such an effect emerged in pure blocks. Meyer et al. argued that Bachoud-Lévi et al.'s (1998) use of intermixed conditions obscured a real word-length effect in object naming. In an additional study (Experiment 4) by Meyer et al., speakers produced the names of pairs of objects (e.g., *pear* and *scissors*); the name of the first object was either mono- or disyllabic, and eye movements before and during speech were monitored. Gaze durations for the left objects were shorter for monosyllabic than for disyllabic object names, which was taken as further evidence for the claim that the preparation of shorter words takes less time than the preparation of longer words. Meyer et al. concluded that phonological encoding proceeds from left to right, and contrary to the earlier view advanced by Bachoud-Lévi et al., a response is typically not initiated until the entire word has been retrieved.

Santiago, MacKay, Palma, and Rho (2000) also reported a word-length effect in picture naming; surprisingly, they found this effect with a mixed presentation rather than the pure presentation of the length manipulation advocated by Meyer et al. (2003; see Roelofs, 2002b, and Santiago, MacKay, & Palma, 2002, for a subsequent discussion of whether this difference should be interpreted as resulting from the number of syllables and the structural complexity or merely from the number of segments). Speakers named objects with either mono- or disyllabic labels, and Santiago et al. reported a substantial latency advantage for the former compared with the latter. In addition, Meyer, Belke, Häcker, and Mortensen (2007) carried out a study in which the word length of the object name was varied from one to three syllables, in pure presentation blocks. Picture-naming latencies were significantly faster for mono- than for trisyllabic picture names, supporting the claim that word length affects object naming.

Overall, the reviewed studies provide some support for the notion that word length constrains object-naming latencies, particularly when short and long responses are generated in separate experimental blocks (as in Meyer et al., 2003, 2007).

Difficulties Arising in Matching Stimuli Across Conditions

As pointed out by Bachoud-Lévi et al. (1998), in studies in which different stimuli are compared across conditions, it is imperative to control for possible differences between the sets other than those of interest. Hence, the validity of studies of the type summarized earlier depends on whether stimuli were matched in all aspects other than word length. This, however, turns out to be very difficult to accomplish in studies with pictorial stimuli. At minimum, stimuli need to be matched across the word-length conditions on frequency of occurrence, AoA, and name agreement, all of which are well-established linguistic predictors of object-naming latencies (e.g., Alario et al., 2004). Additionally, because of the well-documented variability with which spoken responses trigger digital voice keys (Kessler, Treiman, & Mullenix, 2002; Rastle & Davis, 2002), stimuli need to be matched on word-initial phonetic characteristics. Finally, the possibility must be excluded that latencies are confounded with differences in ease of object recognition and/or access to conceptual representations. As we show, all currently existing studies fall short on one or more of these criteria.

Meyer et al. (2003) found slower naming latencies for pictures with disyllabic than with monosyllabic labels. Pictures with mono- and disyllabic labels were selected on the basis of a pretest that assessed their name agreement, and they were matched on frequency of occurrence. AoA values were not reported. However, for the 13 out of 16 monosyllabic labels and the 12 out of 16 disyllabic labels for which values were available either from Stadthagen-Gonzalez and Davis (2006) and/or from Bird, Franklin, and Howard (2001), the results were similar (3.00 and 2.94, respectively, on a scale from 1 to 7). Furthermore, stimuli were pairwise matched on the initial segment. To control for nonlinguistic variables, Meyer et al. carried out a control experiment in which participants categorized the stimuli in a task that did not require name retrieval. An object/nonobject decision task was used in which critical objects were intermixed with pseudo-objects (i.e., entities that could be real objects but are not), and participants performed a corresponding manual response. No significant differences between pictures with mono- and disyllabic labels were found, either in pure or in mixed blocks; hence, the authors concluded that the significant difference emerging in object naming could be attributed to the linguistic variable word length rather than visual or conceptual processes.

A potential problem with this conclusion is that the object/nonobject decision task reported in Meyer et al. (2003) revealed a surprisingly fast average latency of 452 ms. By contrast, our own attempts to use this task have consistently shown substantially slower latencies of roughly 600 ms, which generally are in agreement with previous studies that have used this task (e.g., Gerlach, Law, Gade, & Paulson, 1999; Griffin & Bock, 1998; Holmes & Ellis, 2006; Kroll & Potter, 1984). This raises the possibility that participants in Meyer et al.'s control task may have based their decision on something other than genuine object recognition (e.g., it is possible that the objects could be distinguished from the nonobjects on the basis of surface characteristic of the drawings). If so, the conclusion that the reported latency difference in the picture-naming task can be attributed to word length could be compromised. To investigate this possibility, we retested their original pictures on a different widely used control task: word–picture matching.[1] In this task, a printed word is presented first, followed by the target picture, and participants manually decide whether the two match (also used by Bachoud-Lévi et al., 1998, Experiment 1; Meyer et al., 2007; Özdemir, Roelofs, & Meyer, 2007, and others). In a recent study, Stadthagen-Gonzalez, Damian, Pérez, Bowers, and Marín (2009) systematically investigated the properties of this task by collecting response times on a large set of stimulus pictures and conducting regression analyses that included conceptual (i.e., image agreement, familiarity) and lexical (i.e., frequency, AoA, and word length) predictors. Responses on match trials were sensitive to conceptual but not to lexical predictors, suggesting that participants based their responses on conceptual codes but did not retrieve the picture name. Response latencies in the mismatch condition were not sensitive to any of the predictors. This pattern suggests that *match* responses in this task capture all aspects of picture processing that do not include lexical access. Using this task, we collected name–picture latencies on Meyer et al.'s items, to assess whether they were matched adequately on all variables other than word length.[2] The results showed substantially faster responses for objects with monosyllabic labels (592 ms) than for those with disyllabic labels (626 ms): $F_1(1, 37) = 5.38$, $MSE =$ 19,885, $p = .026$; $F_2(1, 62) = 5.60$, $MSE = 27{,}305$, $p = .021$. Error rates in the two conditions were comparable (9.0% vs. 8.4%; F_1 and $F_2 < 1$). The difference in latencies suggests that the two stimulus sets were not perfectly matched on nonlinguistic variables: In a control task that evidently does not involve lexical access (as shown by Stadthagen-Gonzalez et al., 2009), objects with monosyllabic labels were processed faster than those with bisyllabic labels. Hence, the latency effect reported by Meyer et al. may be attributable to this difference rather than to the hypothesized word-length effect.

Santiago et al. (2000) reported latency differences between mono- and disyllabic stimuli in a picture-naming study. Their stimuli were matched on the Kučera and Francis (1967) frequency count, and our analysis confirmed that this was also the case on the larger and more reliable CELEX (Baayen, Piepenbrook, & Gulikers, 1995) and British National Corpus of spoken English (BNC; Burnard, 1995) norms. Stimuli were furthermore pairwise matched on initial phoneme, and a word–picture matching control task assessed visual/conceptual components. However, Santiago et al. did not control for name agreement, which is probably the most powerful of all contributors to naming latencies (e.g., Alario et al., 2004). Furthermore, stimuli were not matched on AoA. AoA values were available to us from Bird et al. (2001) and/or from Stadthagen-Gonzalez and Davis (2006) for 24 out of 28 monosyllabic labels and for 16 out of 28 disyllabic labels. The values indicated significantly earlier AoA for monosyllabic (2.87) than for disyllabic (3.41) picture labels ($p < .05$; values refer to the scale introduced by Gilhooly & Logie, 1980, with seven age bands beginning with ages 0–2 and increasing 2 years at a time up to ages 11–12 years, with the highest band referring to age 13 or older). This difference could conceivably have affected picture-naming latencies, as judged from previous studies that explicitly manipulated AoA. For instance, Barry, Hirsh, Johnston, and Wil-

---

[1] We thank Antje Meyer for providing us with the original stimuli.

[2] Thirty-eight participants took part in this task. The original pictures from Meyer et al.'s (2003) experiments were used, consisting of 16 objects with monosyllabic names and 16 with disyllabic Dutch names. In their English translation equivalents, 12 out of the 16 pictures with monosyllabic Dutch names were also monosyllabic in English, and 4 were disyllabic. Out of the 16 pictures with disyllabic Dutch names, 8 were monosyllabic in English, 7 were disyllabic, and 1 was trisyllabic. These targets were used for the match responses (i.e., the word preceding object presentation matched the object's name). We intermixed 32 unrelated pictures as filler items for the mismatch responses. The experiment consisted of two experimental blocks with 32 trials each (16 match and 16 mismatch responses in each block); half of the participants judged critical items with monosyllabic names first and those with disyllabic names second and vice versa for the other half. On each trial, a word was presented for 1,000 ms, followed by the picture, which remained on the screen for 1,500 ms. Participants indicated their response by pressing one of the two *shift* keys on the computer keyboard. Assignment of response hand to condition (match or mismatch) was counterbalanced across participants. Latencies on error trials were eliminated from the analysis (8.7%). There were no further outliers, defined as latencies larger than 1,500 ms or smaller than 200 ms.

liams (2001) varied pictures with names acquired early (1.95) and late (3.36) and found an effect of 92 ms. Dent, Johnston, and Humphreys (2008) varied pictures with names acquired early (2.55) and late (3.41) and found an effect of about 50 ms. Hence, it is not implausible to suggest that uncontrolled differences in AoA in Santiago et al.'s study may also have contributed to the results.

Finally, Meyer et al. (2007) reported longer naming latencies for pictures with trisyllabic than monosyllabic names.[3] Latencies on a word–picture matching control task showed slightly longer latencies for monosyllabic than for trisyllabic labels; hence, picture recognition is unlikely to have caused an artificial length effect. Name agreement values for the two sets, collected in a separate norming study, were very similar. Meyer et al. reported a marginally significant difference in frequency counts between mono- and trisyllabic picture labels in the COBUILD database (16.4 vs. 10.1 per million, respectively). When we analyzed the stimuli on alternative frequency counts, we found that the monosyllabic labels were significantly more frequent than the trisyllabic labels in the BNC (Burnard, 1995; 12.9 vs. 5.8 per million, $p < .001$) and in the Kučera and Francis (1967) norms (14.9 vs. 6.2 per million, $p = .014$) and that they were marginally more frequent in the spoken portion of CELEX (6.3 vs. 3.1 per million, $p = .072$). These frequency differences are admittedly rather subtle compared with those in studies (e.g., Jescheniak & Levelt, 1994) that explicitly assessed the effects of frequency in picture naming. Nevertheless, because frequency constitutes one of the main contributors to picture-naming latencies, the naming latency differences reported by Meyer et al. may have been partially caused by this variable rather than by word length.

AoA values were not reported but may have also contributed to the results: For the 12 out of 18 monosyllabic items and the 9 out of 18 trisyllabic items for which such norms were available (again from Bird et al., 2001, and/or from Stadthagen-Gonzalez & Davis, 2006), average AoAs were 2.59 and 3.24, respectively. Concerning articulatory variables, mono- and trisyllabic picture names were selected to have equal numbers of objects starting with vowels, plosives, and fricatives. However, recent research by Kessler et al. (2002) and Rastle and Davis (2002) has underscored the necessity of matching stimuli pairwise on the exact initial phoneme (and, ideally, even on full phonemic onset) if the aim is to equalize word-initial phonetic characteristics; hence, on this criterion the stimuli were less than optimally matched.

The outlined studies underscore the difficulties researchers face in comparing sets of linguistic stimuli across experimental conditions. Because investigations of word length have involved comparisons between two or more sets of stimuli, less-than-perfect matching between stimulus sets that vary on the variable of interest may have obscured a true underlying effect of word length, or it may have created an artifactual latency difference that is then erroneously attributed to word length. Of course, none of our additional analyses on stimuli used in previous studies demonstrated that the conclusions derived from these studies were incorrect. We merely point out that it is rather difficult to devise studies comparing linguistic stimuli across conditions that are entirely free from methodological concerns. The underlying problem afflicts all psycholinguistic studies in which different stimuli are selected for contrasting conditions (cf. Bowers, Davis, & Hanley, 2005; Cutler, 1981; Lewis, 2006). However, in studies of spoken production, the problem is particularly prominent because the visual and conceptual components of object processing, which are typically of no interest when investigating naming, contribute a large portion of variance to latencies and therefore need to be carefully controlled. Additionally, when dealing with pictorial stimuli, the pool of possible candidates to choose from is severely restricted. As a result, the status of word length as a variable contributing to response latencies in spoken production tasks remains to be clarified.

In our first experiment, we take another look at the issue by comparing naming latencies for pictures with mono- and disyllabic labels while controlling as tightly as possible for other potential differences between the two conditions. Following the design of Meyer et al. (2003), items were presented in pure blocks (i.e., short and long words were tested in separate experimental blocks), which should maximize our chances of obtaining an effect.

Experiment 1

Method

Participants. Twenty-four undergraduate students at the University of Bristol were paid a small fee to take part in this experiment. All were native speakers of English, had normal or corrected-to-normal vision, and had no history of language disorders.

Materials. Target line drawings were selected from Cycowicz, Friedman, Rothstein, and Snodgrass (1997). Name agreement and AoA were assessed by collecting ratings. In a printed booklet with pages in a different random order for each rater, 136 line drawings of objects with one- or two-syllable names in English were presented to 18 participants. They were asked to write down the name of each object next to the drawing and to give an estimate (in years) of the age at which they acquired each word. Name agreement was coded as the percentage of participants who gave the expected name for each drawing.

From these results, two sets of 30 line drawings were selected. Set A consisted of drawings with monosyllabic names, and Set B consisted of drawings with disyllabic names (see Appendix A). The experimental items in these two conditions were pairwise matched on either full phonetic onset or on initial vowel. The items had an average AoA estimate of 3.7 years for monosyllabic items and 3.8 years for disyllabic items ($F < 1$). Average name agreement was 97.8% for monosyllabic items and 98.9% for disyllabic items ($F = 1.57$, $p = .22$). The two item sets were also tightly matched on word frequency according to both the spoken portion of CELEX (Baayen et al., 1995; 10.9 vs. 7.9 per million; $F < 1$) and the BNC norms (Burnard, 1995; 19.6 vs. 22.1 per million; $F < 1$). Finally, the two sets of pictures were matched on name–picture matching latencies according to the norms compiled by Stadthagen-Gonzalez et al. (2009), with average match latencies of 506 ms for monosyllabic items and 502 ms for disyllabic items

---

[3] An additional condition with disyllabic labels was included. This condition showed somewhat (18 ms) faster RTs when compared with the monosyllabic condition. However, this condition was unmatched to the other conditions on articulatory characteristics; hence, the difference cannot be unambiguously ascribed to word length.

($F < 1$; latencies on nomatch responses were also very similar: 542 ms and 537 ms, respectively; $F < 1$).[4]

Design. The experimental design included word length (mono- vs. disyllabic) as a within-subjects and between-items variable.

Apparatus. Stimuli were presented with DMDX 3.0 (Forster & Forster, 2003) from an IBM-compatible computer on a 17-in. monitor. Spoken responses were measured with a headset (Sennheiser mb40) with attached microphone, which was connected to the computer, and DMDX determined the onset of each vocal response to the nearest millisecond. Pictures were standardized to a size of approximately 7 × 7 cm and were presented centrally as black line drawings on a white background.

Procedure. Participants were tested individually in a quiet room. Items were presented in pure sets (i.e., all items in a particular experimental block had the same length). Half the participants were presented with the monosyllabic items first, and the other half were presented with the disyllabic items first. Participants were presented with three consecutive blocks of each word-length condition, with items within each block presented in a different randomized order. At the beginning of each experimental half (i.e., the monosyllabic or disyllabic set of words), participants were familiarized with the experimental materials by studying a booklet containing all the pictures in the set and their corresponding expected names. They were instructed to name, as fast and accurately as possible, each picture that would subsequently appear on the screen and to use only the names they had previously studied. Once the participants finished the first half of testing, they were invited to rest briefly and were then tested on the second set of pictures, following the same procedure. The entire experimental session lasted approximately 15 min.

On each trial, a fixation cross was presented centrally for 800 ms, followed by a blank interval of 350 ms. Then the target appeared and remained on the screen for 1,800 ms, followed by an intertrial interval of 1,000 ms.

Results

Trials with hesitations, stutters, and responses that did not correspond to the expected picture name were classified as errors and were excluded from the analysis of latencies (1.1% in the monosyllabic and 0.8% in the disyllabic sets). These error rates were too small to allow for a meaningful analysis. Latencies larger than 1,500 ms and smaller than 200 ms were considered outliers and were also removed (0.3%).

The results showed a mean response latency of 676 ms for the monosyllabic condition and 666 ms for the disyllabic condition. Analyses of variance (ANOVAs) conducted on the data, with word length as a within-subjects but between-items variable, showed no significant difference (F_1 and $F_2 < 1$). To test for the possibility that an effect of word length may interact with the position of a testing block, we additionally analyzed the data with ANOVAs in which block order was included as a within-subjects and within-items variable. There was no significant Block Order × Word Length interaction (F_1 and $F_2 < 1$).

We additionally assessed the statistical power of our study. If we hypothesize a word-length effect of 30 ms on the basis of previous findings, such as those by Meyer et al. (2003), given an observed standard deviation of 40 ms for the difference between mono- and disyllabic conditions, the effect size d is .75. The resulting power value of .94 makes it unlikely that the statistical power of our study was insufficient to detect an existing effect.

Discussion

In Experiment 1, pictures with mono- and disyllabic names were presented in pure blocks, as recommended by Meyer et al. (2003), and the results showed no significant difference in naming latencies. At the same time, the two stimulus sets were matched as tightly as possible on potentially confounding variables. The null finding challenges the claim that word length constrains naming latencies in spoken production.

Nevertheless, it is always possible that some unknown variable confounded with word length compromised our conclusion. In an attempt to address this concern, we introduced in the next experiment a procedure that allowed perfect matching of nonlinguistic variables in a picture-naming task; that is, we assessed the impact of word length while comparing pictures to themselves.

In Experiment 2, objects were presented and named across two different languages, English and Dutch. Half of the pictures (Picture Set 1) had monosyllabic labels in English (*broom*) and disyllabic labels in Dutch (*bezem*); the other pictures (Picture Set 2) had disyllabic labels in English (*candle*) and monosyllabic labels in Dutch (*kaars*). All pictures were named by English and Dutch speakers in their respective native languages. If a length effect exists, then it should contribute to latencies in a systematic manner and emerge as an interaction between language of the speakers and object set: English speakers should have a relative advantage in naming those pictures that are monosyllabic in English, whereas the Dutch speakers should have a relative advantage at naming those pictures that are monosyllabic in Dutch. Of importance, independent of whether main effects of stimulus set, language, or both are obtained, an underlying word-length effect should emerge as an interaction between the two variables. By contrast, the absence of such an interaction would suggest that latencies are not constrained by word length. The key advantage of this approach was that the same pictures are compared across different word lengths, eliminating any possible visual or semantic confounds with length—each picture acts as its own control.

It should be noted that a similar approach of using a cross-language (or cross-population) comparison to control for potential differences between items has been previously used (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Costa, Caramazza, & Sebastian-Galles, 2000). For instance, Costa et al. (2000) asked Spanish–Catalan bilinguals to name pictures whose names either were or were not cognates across the two languages. Latencies were faster for cognate than for noncognate responses (*cognate facilitation effect*). By contrast, when Spanish monolinguals named the same pictures, no differences emerged between cognate and noncognate responses, excluding the possibility that the effect found in bilinguals was due to uncontrolled differences in the materials. Our own experiment, by contrast, takes a slightly different approach: There are no target and control languages, and we are not interested in differences between stimulus sets in either

---

[4] These norms can be accessed online at http://language.psy.bris.ac.uk/name-picture_verification.html

language. Rather, the interaction between response language and stimulus set constitutes the critical source of evidence.

To validate this approach, we additionally asked participants to name the stimuli as printed words. In word naming, length effects are well documented (Balota & Chumbley, 1985; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Jared & Seidenberg, 1991; Weekes, 1997). To our knowledge, there is yet no consensus on the exact origin of the word-length effect in reading aloud, with various authors attributing it to word recognition (e.g., Whitney, 2001), pronunciation (Kawamoto, Kello, Jones, & Bame, 1998), or the operation of a serial print-to-sound assembly route (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Rastle et al., 2009). In the current context, the exact underlying mechanism is of minor interest to us; the word-length effect in reading aloud is used as an empirical benchmark against which to pit potential word-length effects found in picture naming. Specifically, in our experiment, we predict an interaction between response language and stimulus set in word naming. The central question then is whether a similar interaction emerges in object naming. If so, then word length can be safely assumed to contribute systematically to object-naming latencies. If not, then a three-way interaction between stimulus set, language, and modality should be observed, with the interaction between language and stimulus set restricted to written word naming.

Experiment 2

Method

Participants. Thirty-two participants were paid a small fee to take part in the experiment. Sixteen undergraduate students at the University of Bristol, all native speakers of English, contributed to the English portion; 16 students at the Radboud Universiteit Nijmegen, the Netherlands, all native speakers of Dutch, contributed to the Dutch portion. All participants had normal or corrected-to-normal vision and no history of language disorders. None of the English participants took part in the experiment reported earlier.

Materials. Two sets of 12 stimuli were selected as targets (see Appendix B). Stimuli in Set 1 had monosyllabic names in English and disyllabic names in Dutch; stimuli in Set 2 had disyllabic names in English and monosyllabic names in Dutch. For each stimulus, the English and the Dutch labels were matched on their initial phonemes.[5] We collected ratings of name agreement and AoA from 20 additional participants in each language in the same manner as described in Experiment 1. The results showed that stimuli were matched across the two languages on both measures (Fs < 1).

Length characteristics in terms of letters and phonemes are displayed in Table 1. An ANOVA conducted on length in letters, with the variables stimulus set and language showed no effect of set or language (Fs < 1) but showed a highly significant interaction, $F(1, 44) = 52.16$, $MSE = 36.75$, $p < .001$. The same was found for word length in phonemes, with neither a set nor a language effect ($F < 1$) but with an interaction, $F(1, 44) = 83.84$, $MSE = 40.33$, $p < .001$. Stimuli in Set 1 had an average frequency of occurrence of 17.8 per million in the English portion of the CELEX database and had a frequency of 16.6 in the Dutch portion (ANOVA on log-transformed frequencies: $F = 2.47$, $p = .145$). Set 2 had an average frequency of 5.8 in English and an average of 6.1 in Dutch ($F < 1$).

Table 1
Experiment 2: Average Length Characteristics of Stimuli, in Numbers of Letters and Phonemes, Separately for Response Language and Stimulus Set

Length characteristic and set	English	Dutch	Difference
Length (letters)			
Set 1	4.3 (0.65)	6.0 (1.04)	1.7*
Set 2	6.2 (0.83)	4.3 (0.78)	1.9*
Difference	1.9*	1.7*	
Length (phonemes)			
Set 1	3.4 (0.51)	5.4 (0.90)	2.0*
Set 2	5.3 (0.65)	3.7 (0.65)	1.6*
Difference	1.9*	1.7*	

Note. Standard deviations are in parentheses.
* $p < .001$.

Design. The experimental design included response language (Dutch vs. English) as a between-subjects but within-items variable, response modality (word vs. picture naming) as a within-subjects and within-items variable, and stimulus set (Set 1 vs. Set 2) as a within-subjects but between-items variable.

Apparatus. The apparatus was the same as the one described in Experiment 1. Words were presented centrally in lowercase bold 18-point Arial font, in black against a white background. Pictures were standardized to a size of approximately 7 × 7 cm and were presented centrally as black line drawings on a white background.

Procedure. Participants were tested individually in a quiet room. At the beginning of the experiment, participants were familiarized with the experimental materials by studying a booklet containing all the pictures in the set and their corresponding expected names. They were instructed to name, as quickly and accurately as possible, each picture that subsequently appeared on the screen and to use only the names they had previously studied. Each experimental session consisted of two blocks, one in which printed words were named and one in which pictures were named. The order of the two blocks was rotated across participants. Each experimental block consisted of six experimental subblocks, and within each subblock12 stimuli were presented and named once. Subblocks 1 through 3 contained monosyllabic stimuli, and Subblocks 4 through 6 contained disyllabic stimuli, or vice versa. Again, the order was rotated across participants. Hence, each participant named a total of 144 experimental trials (72 picture and 72 word-naming trials). The entire experimental session lasted approximately 15 min.

On each trial, a fixation cross was presented centrally for 800 ms, followed by a blank interval of 350 ms. Then the target appeared and remained on the screen for 1,800 ms, followed by an intertrial interval of 1,000 ms.

---

[5] For some of the stimuli, our attempt to match them on the initial phoneme was merely an approximation; for example, the initial sounds of the English word *whale* and the Dutch equivalent *walvis* are not entirely identical. The remaining differences were, however, small enough to make it unlikely that systematic differences in voice key triggering times were introduced.

Results

Trials with hesitations, stutters, and responses that did not correspond to the expected name were classified as errors and excluded from the analysis of latencies (0.5%). These error rates were too small to allow for a meaningful analysis. Latencies larger than 1,500 ms and smaller than 200 ms were considered outliers and were also removed (2.5%). The results are shown in Figure 1 (left panels), separately for word (top panel) and picture (bottom panel) naming. The right panels of Figure 1 display contrasts and associated 95% confidence intervals, calculated with the method described in Masson and Loftus (2003).

ANOVAs conducted on response latencies showed a significant three-way Response Language × Stimulus Set × Modality interaction, $F_1(1, 30) = 6.10$, $MSE = 2{,}563$, $p = .019$; $F_2(1, 44) = 4.39$, $MSE = 1{,}895$, $p = .042$. Subsequently, the data were analyzed for word and picture naming separately.

Word naming. ANOVAs with response language and stimulus set as variables showed that the effect of response language was not significant in the analysis by participants, $F_1(1, 30) = 0.56$, $MSE = 3{,}710$, $p = .461$, but was significant by items, $F_2(1, 22) = 4.50$, $MSE = 2{,}656$, $p = .045$, with average latencies of 459 ms for Dutch speakers and 468 ms for English speakers. The effect of stimulus set was not significant, $F_1(1, 30) = 0.12$, $MSE = 64$, $p = .732$; $F_2(1, 22) = 0.19$, $MSE = 46$, $p = .672$, with very similar latencies for Sets 1 and 2 (466 ms and 468 ms, respectively). Of importance, a highly significant Language × Stimulus Set interaction was obtained, $F_1(1, 30) = 14.37$, $MSE = 7{,}662$, $p < .001$; $F_2(1, 22) = 9.75$, $MSE = 5{,}755$, $p = .005$, suggesting that naming latencies were affected by the length of the stimuli.

For Dutch speakers, the 20-ms effect of set neared significance by participants, $F_1(1, 15) = 4.14$, $MSE = 3{,}165$, $p = .060$, and was significant by items, $F_2(1, 22) = 5.30$, $MSE = 2{,}386$, $p = .031$. For English speakers, the 24-ms effect of stimulus set was significant, $F_1(1, 15) = 15.06$, $MSE = 4{,}561$, $p = .001$; $F_2(1, 22) = 8.79$, $MSE = 3{,}415$, $p = .007$.

Picture naming. Parallel ANOVAs conducted on picture-naming latencies showed that the effect of response language was not significant, $F_1(1, 30) < 0.01$, $MSE = 2$, $p = .986$; $F_2(1, 22) = 0.01$, $MSE = 6$, $p = .931$, with numerically identical latencies in Dutch and English (611 ms). The effect of stimulus set was

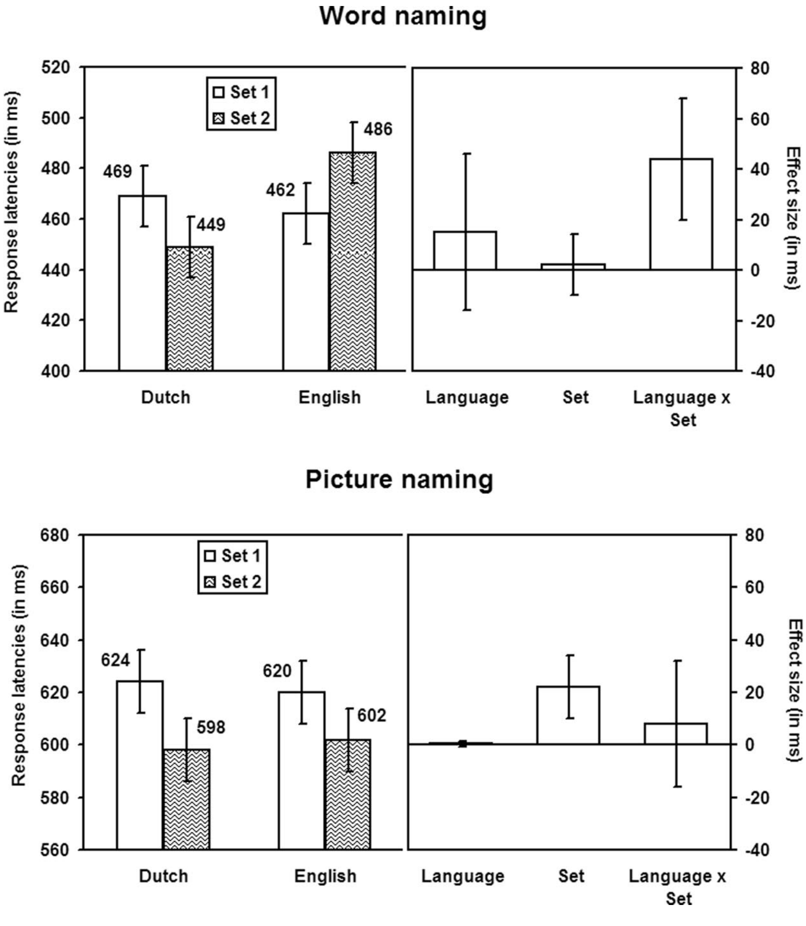

Figure 1. Left: Response latencies (in ms) for word naming (top) and picture naming (bottom), separated by stimulus set (Set 1: monosyllabic in English/disyllabic in Dutch; Set 2: disyllabic in English/monosyllabic in Dutch) and response language (Dutch vs. English). Right: Contrasts for each effect plotted with 95% confidence intervals, calculated with the method described in Masson and Loftus (2003). For convenience, each contrast is plotted as a positive value.

significant, $F_1(1, 30) = 13.75$, $MSE = 7{,}623$, $p < .001$; $F_2(1, 22) = 3.45$, $MSE = 5{,}723$, $p = .077$, with average latencies of 622 ms and 600 ms for Sets 1 and 2, respectively. Crucially, language and stimulus set did not statistically interact, $F_1(1, 30) = 0.46$, $MSE = 254$, $p = .504$; $F_2(1, 22) = 0.26$, $MSE = 205$, $p = .619$, suggesting that in picture naming, word length did not systematically affect latencies.

As in Experiment 1, we additionally analyzed the data with ANOVAs in which block order was included as a within-subjects and within-items variable. Block order did not interact with any of the other variables in the overall analysis or in the analyses conducted for word and picture naming separately (F_1 and $F_2 < 1$).

One aspect of the study that potentially complicates its interpretation arises from the fact that most Dutch speakers are relatively fluent in English. As outlined earlier, previous research has documented a *cognate facilitation effect* (Costa et al., 2000) in bilingual individuals, such that pictures with labels that are cognates in the two languages are named faster than those with noncognate names, all else being equal. Given that some of our items were rather similar across the two languages (*tulip–tulp*), it is conceivable that their cognate (or near cognate) status could have distorted the results and obscured an underlying word-length effect. To investigate this possibility, we calculated a phonological similarity score for each Dutch–English picture name pair: An index ranging from 0 to 1 was computed as the average of the fraction of shared phonemes in and out of position. The resulting scores ranged from a minimum of 0.20 to a maximum of 0.80. Five stimulus pairs showed a score of 0.50 or higher (*helmet–helm*, *mask–masker*", *tulip–tulp*, *king–koning*, and *rail–reling*). We subsequently reanalyzed our data with these five pairs excluded. The results still showed a significant Response Language × Stimulus Set interaction in word naming, $F_1(1, 30) = 9.16$, $MSE = 4{,}807$, $p = .005$; $F_2(1, 17) = 4.10$, $MSE = 2{,}840$, $p = .059$, but not in picture naming, $F_1(1, 30) = 0.02$, $MSE = 11$, $p = .901$; $F_2(1, 17) = 0.01$, $MSE = 8$, $p = .920$. The three-way Response Language × Stimulus Set × Modality interaction remained significant in the analysis by participants, $F_1(1, 30) = 4.77$, $MSE = 2{,}186$, $p = .037$, although it was no longer significant in the analysis by items, $F_2(1, 34) = 2.73$, $MSE = 1{,}254$, $p = .108$. Consequently, we consider it unlikely that cognate effects obscured a word-length effect in picture naming.

Discussion

When Dutch and English participants named the same set of pictures, there was no Response Language × Stimulus Set interaction. As in Experiment 1, this suggests that object-naming latencies are not constrained by the length of the utterance. Although the absence of an interaction in object naming constitutes a null finding, this result does not reflect a lack of statistical power in that the exact same manipulation worked quite well in word naming; indeed, the null effect for objects interacted with the significant length effect observed with words. The key advantage of the present study compared with previous work is that we perfectly matched for nonlinguistic factors in picture naming, given that each picture acted as its own control. The mixed length findings reported previously undoubtedly reflected difficulties in matching different pictures in all relevant ways. The absence of a Stimulus Set × Response Language interaction constitutes rather strong evidence that word length does not contribute to object-naming latencies.

The conclusion derived from Experiments 1 and 2 implies one of two theoretical possibilities. Contrary to what is conventionally assumed, speakers may assemble the phonological content of an utterance in a parallel rather than a serial process; hence, there is no cost for longer compared with shorter words. This seems unlikely given the substantial evidence for the sequentiality of phonological planning in speaking (e.g., Meyer, 1990; Meyer & Schriefers, 1991; Roelofs, 2004; Sevald & Dell, 1994; van Turennout et al., 1997; Wheeldon & Levelt, 1995). The second possibility is that speakers initiate articulation of an utterance before they have planned a single word: Perhaps only the first syllable of an utterance is planned before the response commences, hence rendering latencies for mono- and disyllabic stimuli very similar. Prima facie, this account appears incompatible with recent evidence showing that speakers plan at least one phonological word (e.g., Wheeldon & Lahiri, 1997) and most likely substantially more than one word (e.g., Costa & Caramazza, 2002; Damian & Dumay, 2007; Schnur, Costa, & Caramazza, 2006). For instance, Schnur et al. (2006) demonstrated that when pictorial stimuli were named with sentences such as *The orange girl walks*, distractors that were form related to the verb (e.g., *walnut*) speeded up responses relative to an unrelated condition. This indicates that prior to initiating the response, speakers had coactivated the phonological constituents of the entire utterance.

A possible solution was suggested by Meyer et al. (2003, pp. 144–145). According to the model advocated by Levelt et al. (1999), form encoding is divided into a phonological stage in which segments of an utterance are activated and sequentially assigned to syllable position and a phonetic phase in which articulatory programs corresponding to syllables are accessed and placed in an output buffer (see, e.g., Cholin, Levelt, & Schiller, 2006, for evidence supporting the notion of a syllable-based phonetic level). An outline of this view is presented in Figure 2. The general assumption in previous work has been that articulation is typically initiated when the content of at least one phonological word has been placed in the buffer. However, the authors consider the possibility that less than one word, for instance, merely the initial syllable of a response, is buffered prior to articulation. Hence, speakers may phonologically plan substantial chunks of the utterance, but a response is initiated as soon as the content of the first syllable is placed into the articulatory buffer. This scenario may be able to reconcile studies suggesting a substantial degree of phonological advance planning with the present findings showing the absence of a word-length effect.

Experiment 3 attempted to assess this possibility. The study's experimental design was very similar to the one used by Meyer et al. (2003) and adopted in our first experiment. However, in addition, we combined it with a manipulation taken from recent work by Roelofs (2002a). As in Roelofs's work, we aimed at assessing two different priming effects that potentially reside at distinct processing levels. In addition to trials on which participants named pictures with mono- and disyllabic names, trials were included on which the target pictures were presented and named in conjunction with auditorily presented distractors, consisting of the final syllable of the disyllabic picture names. By manipulating the match between distractor and disyllabic picture names, we predicted a

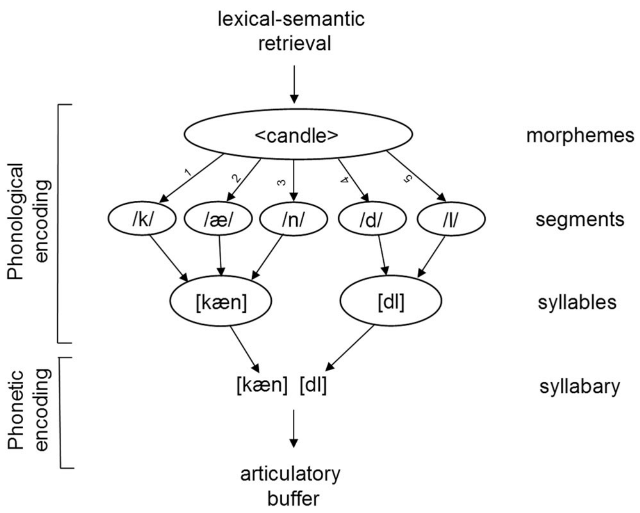

Figure 2. Memory representation of the word *candle* in the WEAVER (word-form encoding by activation and verification) model. For the sake of simplicity, a number of details, most prominently metrical structure encoding, have been omitted. Adapted from "The WEAVER model of word-form encoding in speech production," by A. Roelofs, 1997, *Cognition, 64,* p. 258. Copyright 1997 by Elsevier.

facilitatory effect, such that naming latencies in the related condition (e.g., *candle* presented with the distractor /dl/) should be faster than latencies in the unrelated condition (*candle* presented with /g∂*/). Facilitatory effects of form relatedness in picture–word interference tasks are typically attributed to the stage of phonological encoding (e.g., Starreveld, 2000). Specifically with regard to end-related distractors of the type used here (e.g., Damian & Dumay, 2007; Meyer & Schriefers, 1991; Roelofs, 2002a), the implication is that phonological encoding of the target name must have encompassed the overlapping portion. For us, the aim was to investigate whether such a facilitation effect from end-related distractors could be obtained in conjunction with a renewed failure to obtain an effect of word length. Such an outcome would then suggest that phonological encoding encompassed the entire word, yet only a single syllable was phonetically encoded by the time a speaker initiated articulation.

Experiment 3

Methods

Participants. Twenty undergraduate students from the University of Bristol were paid a small fee to take part in this experiment. All were native English speakers, had normal or corrected-to-normal vision, and had no history of language disorders. None had taken part in the first two experiments.

Materials. The pictures with mono- and disyllabic names from Experiment 1 were again used in this experiment. Each picture was presented and named three times: once by itself (i.e., without a distractor) and twice in conjunction with distractor syllables. Distractor syllables consisted of the final syllables of the disyllabic picture names. For disyllabic picture names, these were chosen such that in the related condition, distractors consisted of the final syllable of the corresponding picture name (average overlap was 50.7% of target phonemes; *SD* = 15.9); for the unrelated condition, pictures and distractor syllables were recombined such that there was no phonological overlap. For pictures with monosyllabic names, the distractors were unrelated under both instances of occurrence, such that any kind of form overlap was avoided; note that both distractor conditions for monosyllabic picture labels merely served as fillers in the context of this experiment, and because they do not contain meaningful information, results are not reported. See Appendix A for all combinations between picture names and distractors.

To record the auditory distractors, a female speaker pronounced the final syllable of the disyllabic stimuli. Sound files were digitized with a sampling frequency of 16 kHz and presented during the experiment over the Sennheiser headset at comfortable volume. The stimuli had an average length of 353 ms (*SD* = 62).

Procedure. Pictures with mono- and disyllabic labels were named in separate halves of the experiment; whether mono- or disyllabic items were presented first was counterbalanced across participants. Within each half, three blocks of 30 trials each were presented, with each of the 30 pictures named once within each block. Within a block, 10 pictures were presented without distractors, and 20 were presented with distractors (for pictures with disyllabic labels, 10 distractors were form related, and 10 were unrelated; for pictures with monosyllabic labels, all distractors were unrelated). Assignment of stimuli to conditions was rotated across blocks. A new random sequence of trials within a block was generated for each participant, and the order of the three blocks within each half was randomized for each participant. The entire experiment consisted of 180 trials and took approximately 30 min to administer.

Results

Trials with hesitations, stutters, and responses that did not correspond to the expected name were classified as errors and excluded from the analysis of latencies (1.7%). Latencies larger than 1,500 ms and smaller than 200 ms were considered outliers and also removed (0.1%).

Effect of word length. For the trials on which targets were presented without distractors, the results showed a mean latency of 664 ms for monosyllabic items and a mean latency of 658 ms for disyllabic items. ANOVAs showed that the difference in reaction times was not significant (F_1 and $F_2 < 1$), hence replicating the null finding regarding word length observed in the first experiment with the same stimuli. Errors constituted 1.2% of the trials in the monosyllabic and 1.0% in the disyllabic sets; the difference between these error rates was not significant (F_1 and $F_2 < 1$).

Effect of picture-distractor relatedness. Latencies for trials with end-related distractors were substantially faster (692 ms) than for those with unrelated distractors (730 ms); the difference of 38 ms was significant, $F_1(1, 19) = 30.05$, $MSE = 14{,}164$, $p < .001$; $F_2(1, 29) = 8.55$, $MSE = 21.862$, $p = .007$. Corresponding error rates were numerically identical (1.7%; F_1 and $F_2 < 1$).

Discussion

The results from the no-distractor conditions within Experiment 3 lend further support to the claim that response latencies in picture naming are not constrained by word length. The results of the distractor relatedness manipulation—a facilitatory effect from distractors related to the second syllable of a disyllabic response—additionally suggest that speakers, by the time they initiated a response, had phonologically encoded the second syllable of the disyllabic response words. In combination, the results support the possible scenario suggested by Meyer et al. (2003): At the level of phonological encoding, speakers clearly prepare the entire word before commencing a response, but at the subsequent phonetic level, a response is initiated as soon as the content of the first syllable is placed into the articulatory buffer.

General Discussion

Summary and Implications

The assumption of incremental phonological encoding, combined with the claim that encoding encompasses at least a single word, predicts a word-length effect in the spoken production of single words: All else being equal, longer words should take more time to prepare than shorter ones. As summarized in the introductory section, this prediction is difficult to verify, mainly because the crucial comparison is between sets of items for which it has to be assured that items are matched on all variables other than word length. In Experiment 1, we used a set of stimuli that were stringently matched on a range of variables, and we did not obtain evidence for a word-length effect. In Experiment 2, we approached the issue by asking speakers of Dutch and English to name the same pictures across the two languages. In this way, the pictures were perfectly matched nonlinguistically. Again, no evidence for an influence of word length was found. Finally, in Experiment 3, we asked how a null finding concerning word length could be reconciled with previous observations that the scope of phonological encoding is relatively far reaching. By combining an experimental manipulation of word length—as in Experiment 1—with the use of spoken distractors, which were related to the second syllable of disyllabic picture names, we showed that, indeed, a single experiment can document both a null finding of word length and facilitatory effects of distractors on the word-final syllable. Overall, the results imply that at least for our target language (discussed later) and for the length difference we tested (mono- vs. disyllabic stimuli), word length does not contribute to naming latencies. A practical consequence for future studies of single word production is that there is no need to be concerned about matching stimuli across conditions on this variable (although the need to control for other variables, such as AoA and frequency, which are highly correlated with word length, is worth noting).

The conclusion that word length does not contribute to naming latencies evidently contradicts Meyer et al.'s (2003) suggestion, based on the presence of such an effect in their results, that speakers plan an entire word before initiating a response (as argued in our introductory section, their positive finding quite likely arose from mismatched stimuli). At the same time, the central finding of Experiment 3—facilitation from form-related distractor words that are related to the second syllable of disyllabic responses—is at odds with minimalist accounts, such as the one by Bachoud-Lévi et al. (1998), according to which speakers can base their response on incomplete access to form representations of the response. In the next section, we discuss this constellation of findings with regard to current models of spoken production, and we argue that the absence of a word-length effect is generally compatible with observations of quite extensive phonological priming, as long as the two observations are attributable to two different processing levels.

A Detailed Account of the Findings

Our results raise important theoretical issues. Specifically, if the absence of a word-length effect suggests that speakers initiate a response as soon as the initial syllable of an utterance has been prepared (as suggested by, e.g., Bachoud-Lévi et al., 1998), how can this suggestion be reconciled with the finding that speakers phonologically plan substantial portions of an utterance (e.g., Schnur et al., 2006)? One possible solution is to assume, in accordance with Meyer et al. (2003), that effects of priming and of word length are attributable to different processing hierarchy levels. To evaluate this scenario, we discuss the theoretical framework outlined by Levelt et al. (1999; see Roelofs, 1997, for computational details) in some detail. According to the WEAVER (word-form encoding by activation and verification) model (see Figure 2), a target morpheme (or lexeme) initially activates all of its corresponding phonological segments in parallel, together with information about their order. Segments subsequently activate successive syllable nodes in a strictly sequential fashion. Access time of a particular syllable depends on its activation value, relative to those of competing syllables. Successive chunks of articulatory programs are then retrieved from a *mental syllabary* and stored in an articulatory buffer.

In the current version of the model, a response is initiated when the phonetic content of at least one phonological word (e.g., a single lexical word plus associated unstressed function words, such as auxiliaries, determiners, conjunctions, and prepositions) has been placed into the articulatory buffer. Indeed, this assumption is consistent with Meyer et al.'s (2003) claim that picture naming is subject to a word-length effect. However, given the present findings, it appears that speakers may be able to initiate their response on the basis of the phonetic content of just the initial syllable of a response.

How does the model account for the effects of form-related distractors in picture–word interference tasks? Somehow the spoken distractor needs to activate the initial syllable of the picture name before its phonetic content is entered into the response buffer. Given that in our Experiment 3, the spoken distractors consisted of the second syllable of the picture names, it must be that the distractors activated the morphemes of the target names, which in turn activated the initial segments and corresponding first syllable. For instance, in Figure 2, the distractor [dl] activates the morpheme , which in turn activates the segments /k/ and /æ/ and /n/, which in turn activate the syllable [kæn]. This in turn facilitates the selection of the articulatory score of [kaen] during the process of picture naming and speeds up the pronunciation of *candle*.

There are various ways in which the distractor [dl] could activate the morpheme . For example, the process could be lexically mediated, with [dl] activating a set of form-related lexical items in the phonological input lexicon (e.g., *fiddle, candle, spindle*), which in turn activates in the output lexicon. This possibility is compatible with evidence that perception of a spoken word entails the coactivation of end-related words (e.g., Allopenna, Magnuson, &

Tanenhaus, 1998; Bowers, Davis, Mattys, Damian, & Hanley, 2009; Shillcock, 1990). This mechanism would preserve the modular nature of speech production, as assumed by Levelt et al. (1999). Alternatively, the spoken distractor [dl] could directly activate the segments /d/ and /l/ (either because of connections between segments in the input and output phonological systems or because of the input and output segments being one and the same in a common phonological system; cf. Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Martin & Saffran, 2002). These segments would then need to activate the morpheme , which would in turn activate the initial segments /k/, /æ/, and /n/, thus facilitating the production of *candle*. This later scenario would require feedback connections between segments and morphemes in the speech production process (contrary to Levelt et al., 1999). It should be additionally noted that we are not theoretically committed to WEAVER and its assumptions, specifically to the claim that form-related priming emerges only through reduced access time of relevant syllable nodes. It may be that two independent criteria need to be met before articulation can start: At the phonological level, the complete form of at least one phonological word has been created, and at the articulatory level, at least one syllable has been selected. In either case, the critical point is that effects of phonological facilitation and of word length (or rather the absence thereof) are not incompatible with each other, under the assumption that they reside at separate and successive processing levels, one operating in parallel and the other sequentially.

Are There Effects of Word Length in Other Contexts?

Our finding of a null effect of word length in single word production must be qualified in that we tested only stimuli with mono- or disyllabic names. Perhaps word-length effects would emerge with a stronger contrast, by comparing monosyllabic responses to those with three (*parachute*) or four (*alligator*) syllables? Such an experiment would be valuable, but because pictures with names that are longer than two syllables are relatively rare, the constraints on stimulus selection make it very difficult to match them to stimuli with shorter names. The approach introduced in Experiment 2—rendering matching superfluous by comparing a picture to itself across two target languages—is even more difficult to extend to longer stimuli.

Could it be the case that effects of word length in spoken production emerge in some, but not all, languages? Given that previous studies have suggested different basic strategies in speech perception across languages (e.g., Cutler, Mehler, Norris, & Segui, 1983; Sebastian-Galles, Dupoux, Segui, & Mehler, 1992), perhaps the phonological encoding mechanism similarly varies with response language. Indeed, in Bates et al.'s (2003) large-scale regression analysis of picture naming across many languages, word length, measured in number of syllables, emerged as a significant predictor for naming latencies in Spanish, Italian, and Hungarian but not in English, German, Bulgarian, and Chinese. With English and Dutch as the response languages in our experiment (both Germanic languages), our failure to find a word-length effect is generally compatible with the regression results. Thus, the possibility remains that equivalent experiments carried out with Spanish and Italian speakers would obtain results different from our own. This possibility clearly needs to be addressed in future research.

The work reported in this article was concerned with the role of word length in the spoken production of single words. Here, we believe that the current data provide rather strong evidence for the claim that speakers are generally able to initiate a response as soon as the initial syllable becomes available, hence rendering the length of the response irrelevant. This inference does not necessarily imply that speakers could not, depending on the context, choose larger planning units than a single syllable. The production of utterances that consist of more than a single word entails additional requirements, such as balancing the need for fluency with minimizing processing load caused by planning (e.g., Levelt & Meyer, 2000). In the growing literature on multiword utterances, an empirical observation has recently been documented that bears some resemblance to the topic of the present article, namely a so-called *reversed word-length effect*. Griffin (2003) asked speakers to name two objects, presented side by side, with two nouns, and simultaneously monitored eye movements associated with the response. In such situations, speakers typically move their eyes to the second object before initiating a response (*speech–gaze lag*). It was shown that utterances with a short noun in the first position (*wig–carrot*) exhibited a longer speech–gaze lag than those with a long noun in the first position (*windmill–carrot*). A reversed word-length effect was additionally observed on speech onset latencies. According to Griffin, this pattern may arise from the fact that pronouncing a short word takes less time than pronouncing a long one. Consequently, if speakers need to coordinate an utterance consisting of more than a single word, more planning is required prior to speech onset when the initial word is short than when it is long. An alternative account was recently proposed by Meyer et al. (2007), who reported a similar reversed word-length effect on speech–gaze lag but not (contrary to Griffin, 2003) on speech onset latencies. It was proposed that speakers start their utterance as soon as the initial syllable of the leftmost object is encoded. As a consequence, the speech–gaze lag should be longer when the first object name is short than when it is long. This account, if true, would be perfectly compatible with our finding of an absent word-length effect in the naming of single pictures; by contrast, Griffin's account would suggest that findings from studies on single word production, such as ours, reflect rather atypical circumstances (i.e., little risk of generating dysfluencies and, consequently, speakers engaging in minimal planning), whereas speech in a real-life context requires a more balanced trade-off between efficiency and fluency. No matter which account eventually turns out to be correct, it is clear that findings from single word production studies need to be put into perspective with regard to how connected speech is generated. So, in practice, in fluent speech, there may well be word-length effects, but for reasons other than the retrieval of single words.

References

Alario, F. X., Costa, A., & Caramazza, A. (2002a). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes, 17,* 299–319.

Alario, F. X., Costa, A., & Caramazza, A. (2002b). Hedging one's bets too much? A reply to Levelt (2002). *Language and Cognitive Processes, 17,* 673–682.

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36,* 140–155.

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

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

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.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language, 44,* 350–375.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A,* 560–585.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., . . . Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10,* 344–380.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33,* 73–79.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97,* B45–B54.

Bowers, J. S., Davis, C. J., Mattys, S. L., Damian, M. F., & Hanley, D. (2009). The activation of embedded words in spoken word identification is robust but constrained: Evidence from the picture–word interference paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1585–1597.

Burnard, L. (1995). *The user's reference guide for the British National Corpus.* Oxford, England: Oxford University Computing Services.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. C. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1430–1450.

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99,* 205–235.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

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.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1283–1296.

Cutler, A. (1981). Making up materials is a confounded nuisance, or—Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10,* 65–70.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature, 304,* 159–160.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65,* 171–237.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57,* 195–209.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17,* 149–195.

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

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 282–301.

Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology, 84,* 502–507.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124.

Gerlach, C., Law, I., Gade, A., & Paulson, O. B. (1999). Perceptual differentiation and category effects in normal object recognition—A PET study. *Brain, 122,* 2159–2170.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12,* 395–427.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10,* 603–609.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language, 38,* 313–338.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13,* 884–910.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 824–843.

Jordan, M. I. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 796–836): Hillsdale, NJ: Erlbaum.

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.

Kessler, B., Treiman, R., & Mullenix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47,* 145–171.

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.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levelt, W. J. M. (2002). Picture naming and word frequency: Comments on Alario, Costa and Caramazza, *Language and Cognitive Processes, 17*(3), 299–319. *Language and Cognitive Processes, 17,* 663–671.

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,* 122–142.

Lewis, M. B. (2006). Chasing psycholinguistic effects: A cautionary tale. *Visual Cognition, 13,* 1012–1026.

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 organization of perception and action: A theory for language and other cognitive skills.* New York, NY: Springer.

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

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220.

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., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language, 57,* 210–231.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48,* 131–147.

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.

Özdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition, 105,* 457–465.

Rastle, K., & Davis, M. H. (2002). On the complexities of naming. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 307–314.

Rastle, K., Havelka, J., Wydell, T. N., Coltheart, M., & Besner, D. (2009). The cross-script length effect: Further evidence challenging PDP models of reading aloud. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 238–246.

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

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

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition, 32,* 212–222.

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.

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.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research, 35,* 189–213.

Sebastian-Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31,* 18–32.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.

Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name–picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *Quarterly Journal of Experimental Psychology, 62,* 1581–1597.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38,* 598–605.

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

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

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50*(A), 439–456.

Wheeldon, L. (2000). Generating prosodic structure. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 249–274). Hove, England: Psychology Press.

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.

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

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8,* 221–243.

Appendix A

Stimuli Used in Experiments 1 and 3

Monosyllables	Disyllables	Distractor 1 (Experiment 3)	Distractor 2 (Experiment 3)
ant	anchor	/k∂*/	/sl/
axe	apple	/pl/	/bIt/
bat	barrel	/r∂l/	/dl/
bird	basket	/skIt/	/pl/
bow	bottle	/tl/	/d∂*/
bus	button	/tn/	/r∂l/
cake	camel	/ml/	/∂n/
comb	candle	/dl/	/g∂*/
cow	carrot	/r∂t/	/tIn/
dart	donkey	/kI/	/ð*/
flag	feather	/ð*/	/ml/
fan	finger	/g∂*/	/tn/
fork	flower	/∂*/	/tl/
harp	hammer	/m∂*/	/skIt/
hat	hammock	/m∂k/	/sl/
lamp	kettle	/tl/	/l∂*/
kite	ladder	/d∂*/	/gwIn/
leg	lion	/∂n/	/kI/
moon	monkey	/kI/	/d∂/
mouse	mountain	/tIn/	/d∂ʊ/
pen	panda	/d∂/	/kI/
pig	pencil	/sl/	/m∂k/
pear	penguin	/gwIn/	/r∂t/
pipe	pumpkin	/kIn/	/∂*/
rake	rabbit	/bIt/	/d∂*/
rope	ruler	/l∂*/	/kIn/
spoon	spider	/d∂*/	/bl/
tie	table	/bl/	/k∂*/
watch	whistle	/sl/	/m∂*/
worm	window	/d∂ʊ/	/tl/

Note. Items in the Distractor 1 column are form related to disyllabic picture labels.

Appendix B

Stimuli Used in Experiment 2

Set 1		Set 2	
English	Dutch	English	Dutch
bra	beha	candle	kaars
brush	borstel	collar	kraag
king	koning	dagger	dolk
mask	masker	dragon	draak
monk	monnik	fairy	fee
nail	nagel	helmet	helm
peach	perzik	peacock	pauw
rail	reling	rudder	roer
rain	regen	scissors	schaar
skull	schedel	scooter	step
swing	schommel	spider	spin
whale	walvis	tulip	tulp

Received October 16, 2008
Revision received February 1, 2010
Accepted February 9, 2010 ■