JOURNAL OF MEMORY AND LANGUAGE 34, 311–334 (1995)

Monitoring the Time Course of Phonological Encoding

LINDA R. WHEELDON AND WILLEM J. M. LEVELT

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Three experiments examined the time course of phonological encoding in speech production. A new methodology is introduced in which subjects are required to monitor their internal speech production for prespecified target segments and syllables. Experiment 1 demonstrated that word initial target segments are monitored significantly faster than second syllable initial target segments. The addition of a concurrent articulation task (Experiment 1b) had a limited effect on performance, excluding the possibility that subjects are monitoring a subvocal articulation of the carrier word. Moreover, no relationship was observed between the pattern of monitoring latencies and the timing of the targets in subjects' overt speech. Subjects are not, therefore, monitoring an internal phonetic representation of the carrier word. Experiment 2 used the production monitoring task to replicate the syllable monitoring effect observed in speech perception experiments: responses to targets were faster when they corresponded to the initial syllable of the carrier word than when they did not. We conclude that subjects are monitoring their internal generation of a syllabified phonological representation. Experiment 3 provides more detailed evidence concerning the time course of the generation of this representation by comparing monitoring latencies to targets within, as well as between, syllables. Some amendments to current models of phonological encoding are suggested in light of these results. © 1995 Academic Press, Inc.

Most current models of speech production propose that articulation is preceded by the generation of an abstract representation of the form of the target word or utterance (Dell, 1986, 1988; Garrett, 1975; Levelt, 1989; Levelt & Wheeldon, 1994; Shattuck-Hufnagel, 1979, 1983, 1987). Indeed, the existence of an internal abstract speech code is a basic assumption within psycholinguistic theory. In speech recognition, a store of abstract lexical representations has been postulated to accommodate the fact that no two tokens of a given word form are acoustically identical (Lahiri & Marslen-Wilson, 1991). Similarly, in speech production the same word, in different contexts, can be spoken with very different segmental and syllable structure, intonation, duration, and amplitude. This generalization has long been captured in linguistic theory by the proposal that every linguistic item has a unique phonological representation which encodes only that information which can distinguish among words in the language. Phonological representations are categorical in nature and consist of discrete timeless segments. In contrast, a phonetic realization of a linguistic item is a context-dependent, quantitative representation, realized in time and space. All the possible phonetic realizations of an item in a given context can be derived by rule. For example, aspiration of stop consonants in some languages is a distinctive phonological feature that can change word meaning (e.g., /kal/, tomorrow and /k^{h}al/ drain, in Bengali) and would therefore be specified in the abstract phonological representation. The same feature in English never distinguishes words; it

The authors thank Ger Desserjer and Jouke Riepkema for their help with the running and analysis of the experiments reported, and Vincent Evers for carrying out speech lab measurements. Many thanks also to Antje Meyer, Pienie Zwitserlood, and an anonymous reviewer for their comments on an earlier version of this article. Address correspondence and reprint requests to Linda Wheeldon, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

0749-596X/95 $6.00

occurs predictably only for stressed syllable initial voiceless stops and would, therefore, be specified only in the surface phonetic representation.

Speech error data provide evidence that an abstract phonological code is computed during speech production. When segments of a word exchange, they usually accommodate themselves to their new environment so that they remain phonetically well formed. For example, an error which displaced an unstressed /p/ to a stressed syllable initial position would also cause it to acquire the appropriate aspiration. The error must therefore have occurred at a level of representation more abstract than and generated prior to the computation of phonetic form. Models of speech production almost universally locate segmental speech errors within the processes that generate a phonological representation of the target utterance (Dell, 1986, 1988; Levelt 1989, 1992; Shattuck-Hufnagel, 1992). These models explain such errors in much the same way. They propose that accessing the phonological representation of a word releases two kinds of information: (1) a frame which specifies the structure of the word terminating in slots for phonemes and (2) the phonemes to fill these slots. The phonemes are then assigned to their positions in the frame and segmental speech errors occur due to errors in this assignment process (Dell, 1986, 1988; Levelt, 1989, 1992; Shattuck-Hufnagel, 1979, 1983, 1987). Despite general agreement about this basic mechanism, models differ markedly, both in the time course of the phoneme to frame assignment process and in the types of word frame they postulate (cf. Dell & O'Seaghdha, 1992; Levelt, 1992; Meyer, 1992).

In Dell's (1986) model, the phonological representation of a word is not stored in linear order. Instead, each phoneme is marked for its syllable position (whether it occurs in onset, nucleus, or coda position). The model assumes that the constituent phonemes of a syllable become activated in parallel and can be assigned to their syllable positions in any order. This contrasts with the proposal that phonemes are associated in a strictly serial manner proceeding left-to-right from the beginning to the end of the word (Levelt, 1992; Meyer, 1990, 1991; Meyer & Schriefers, 1991; Shattuck-Hufnagel, 1983). According to such models, a word's phonological representation encodes the order of its phonemes, which are released in a linear fashion for assignment to the frame.

Models also differ in whether the frame they postulate is syllabified. There are aspects of phonological speech errors that can best be explained with reference to syllables. The syllable position constraint refers to the finding that interacting phonemes in speech errors usually have the same syllable position. For example, onsets exchange with onsets (e.g., *teep a cape*—keep a tape), vowels with vowels (e.g., *fash* and *tickle*—fish and tackle), and offsets with offsets (e.g., *arg* of the *fuwt*—art of the fugue. All examples from Fromkin, 1971). In order to account for the syllable position constraint, it has been claimed that phonemes marked for syllable position are associated to slots in a syllabified word frame (Dell, 1986; Levelt, 1989; Shattuck-Hufnagel, 1983, 1987). Alternatively, Shattuck-Hufnagel (1987, 1992) claimed that speech error data are also consistent with the association of phonemes to unsyllabified word frames. Approximately 80% of segmental exchange errors involve word-onset segments, and Shattuck-Hufnagel (1992) proposed that speech error data can be adequately accounted for by a frame detailing only word position and lexical stress.

More recently the classic structure/content distinction has itself come under attack. Dell, Juliano, and Govindjee (1993) presented a parallel distributed processing model which generates single words and produces phonologically constrained errors without an explicit separation of structured frames and phoneme fillers. Instead, constraints emerge from the model's sensitivity to phoneme similarity and sequential biases. However, this model produces only nonmovement errors in which a phoneme (or a sequence of phonemes) is replaced by another, inserted into a word or deleted from a word. Movement errors, especially complete segment exchange errors (see examples above), provide the strongest evidence for the structure/content distinction, and this model has no mechanisms that could, in theory, produce exchange errors.

The models we have discussed have been designed, almost exclusively, to account for speech error data. However, while speech errors have proved a rich source of evidence for isolating salient units of representation, they can tell us very little about the time course of their generation. And in particular they provide no evidence that would allow us to choose between the alternative accounts of phonological encoding processes described above (see Meyer, 1992, for a review of the limitations of error data). The aim of the work we report here was to devise an experimental methodology with which to track the time course of the generation of a phonological representation (henceforth phonological encoding) and to provide new data to constrain models of this process.

The methodology we introduce is a production version of the phoneme and syllable monitoring task frequently used in the field of speech perception (e.g., Frauenfelder, Segui, & Dijkstra, 1990; Mehler, Dommergues, Frauenfelder, & Segui, 1981). It requires subjects to monitor their own internal generation of words. The motivating idea behind this research is that phoneme monitoring of internal speech production is based on the output from a phonological encoding process allowing the time course of this process to be traced.

MONITORING INTERNAL SPEECH

Most people have experienced "hearing" their internal speech during the processes of reading and writing, of short-term memory, and of some kinds of thought (Jackendoff, 1987). That we sometimes spontaneously correct our own speech errors before they are articulated (Levelt, 1983) indicates that we can monitor our own speech production at some prearticulatory stage. Further evidence for internal speech monitoring comes from experiments designed to elicit slips of the tongue (Baars, Motley & MacKay, 1975). Subjects are much less likely to make a slip when it forms a rude word (e.g., *tool kits* → *kool tits*) than when it does not (Motley, 1980). Furthermore, when subjects make neutral errors like *kool kits* instead of the potentially rude error these errors are accompanied by an increased galvanic skin response (Motley, Camden, & Baars, 1982) suggesting that the rude slip was initially generated but detected and corrected prior to articulation.

While the existence of some kind of monitor is little contested, there are different theories concerning its characteristics. Some researchers assume that speakers have direct access to the workings of the production system (Laver, 1980). However, speakers can monitor their output for errors at many levels, such as social acceptability, grammaticality, lexical and phonological structure (Levelt, 1983, 1989). This would require a separate monitor working at each level of processing in the production system. This is an unattractive proposal, first because it requires many additional devices and second because the sheer speed and parallel nature of production processes precludes their being under central control.

An alternative account of internal speech claims that a prearticulatory speech output code is monitored during speech production by means of an internal loop to the speech comprehension system (Levelt, 1983, 1989, Monsell, 1987). No detailed model of the workings of such an internal loop has been proposed, and we know of no data which could guide a choice between these alternatives. Perhaps, for reasons of parsimony, the internal loop account is to be preferred as it posits one mechanism which allows internal speech to be monitored for many different kinds of information. The data we present below are consistent with both a production system monitor working on the output from phonological encoding and a comprehension monitor working on phonological output via an internal loop. However, these two possibilities do have different consequences for the interpretation of our results. We will therefore defer further consideration of this issue until the General Discussion.

Our aim was to use the internal monitoring task to track the time course of phonological encoding. In order to do this we first need to demonstrate that our methodology taps processing at exactly this level of representation. A number of different form representations have been proposed in the literature as intervening between the selection of a word we want to say and its articulation. Responses in our internal monitoring task could, in principle, be based on the generation of any one of these representations. We will therefore briefly review the kinds of form representation proposed by current models of speech production processes before discussing ways to determine which representation underlies performance in our task.

Form Representations in Speech Production Models

All models of speech production assume that the encoding of a word's sound form follows the retrieval of its meaning. Following Garrett (1975) and Kempen and Huijbers (1983), Levelt (1989) postulated that following the selection of a word's semantic/syntactic attributes, contact is made with a long-term store of word-forms or lexemes. These lexical form representations include information about a word's phonological segments, number of syllables (but not syllable structure) and stress pattern. As discussed above, most models of speech production claim that when a word is accessed, its frame and its phonemes are made available separately. The phonemes are then assigned to their positions in the frame (Dell, 1986, 1988; Levelt, 1989, 1992; Shattuck-Hufnagel, 1979, 1983, 1987). Our interest here is in the nature of the speech codes postulated. Both the input to the assignment process and the output from it are phonological representations: the input consisting of the word's constituent phonemes and the output consisting of a syllabified phonological representation.

Most theories of speech production only model the prearticulatory form representation at a phoneme or feature level, assuming implicitly (Dell, 1986, 1988; Shattuck-Hufnagel, 1979) or explicitly (MacKay, 1982) that this level of representation directly activates articulatory routines. A similar assumption was made in early phonological theory. Chomsky and Halle (1968) assumed that phonetic implementation was dictated purely by the physiology of articulation and was, therefore, automatic and universal. However, more recent work in phonetics has shown that a phonological code cannot form the input to the articulatory process. Some phonetic context effects have been shown to differ across languages (Keating, 1988), suggesting that they must form part of a language dependent phonetic representation. Levelt and Wheeldon (1994) discuss this gap in production modelling and propose a mechanism which takes as input an abstract phonological representation and translates it, syllable by syllable, into a context dependent phonetic representation which forms the input to articulatory processes.

In summary, four different levels of form representation have been proposed in the speech production literature, as intervening between the selection of the intended word and its articulation: (1) a stored lexical form, or lexeme; (2) the lexeme's discrete phonemes, which are made available independently of its prosodic frame; (3) the output from a procedure which assigns phonemes to their position in the frame—a syllabified phonological representation; (4) a quantitative phonetic representation detailed enough to guide articulation.

IDENTIFYING FORM REPRESENTATIONS

But what kind or kinds of prearticulatory code do we generate in tasks involving an internal speech code? Internal speech has often been observed to have a faster rate of production than overt speech (Anderson, 1982; Landauer, 1962; Weber & Castleman, 1970) and this has been interpreted as evidence that the internal speech code is more abstract than the code underlying overt speech. However, in some situations the experience of internal speech is accompanied by activity in the appropriate articulatory muscles (Sokolov, 1972). Electromyographic monitoring of the articulatory muscles during reading has produced evidence of muscle movement in subjects reading complex prose passages (Hardijk & Petrinovitch, 1970). Internal speech can, therefore, involve phonetic and articulatory as well as phonological levels of representations. How then are we to determine which level of representation underlies performance in a given task?

Segmental speech errors have been found to occur when subjects internally recite tongue twisters. Although fewer errors are detected, they are of the same kind and have the same relative frequency of occurrence as overt speech errors (MacKay, 1982, 1987; Dell & Repka, 1992). Therefore, applying the same argument that has been applied to overt speech errors, the internal speech code cannot consist only of activated lexical representations but involves the same phonological encoding processes that govern overt speech errors.

Investigations of short-term memory processes provide us with a tool for distinguishing between phonological and phonetic form representations. There is a wealth of evidence that a speech code plays an important role in memory span tasks which require short term retention of a sequence of verbal items—usually digits or letters (Baddeley, 1983; Baddeley, Thomson, & Buchanan, 1975; Vallar & Baddeley, 1984). Performance in such tasks is adversely affected by the existence of phonological similarities between the to-be-remembered items (Baddeley, 1966; Conrad, 1964; Conrad & Hull, 1964; Ellis, 1980; Wickelgren, 1965) and by an increase in the spoken duration of the items (Baddeley et al., 1975; Ellis & Hennelley, 1980), even though the items are presented visually. Conrad's (1970) demonstration that congenitally deaf children also show phonological confusions in remembering sequences provided evidence that the speech code is not acoustic in nature. If subjects are required to concurrently and repeatedly articulate a prespecified string (henceforth, articulatory suppression) while performing a short-term memory task, there is a dramatic decrease in performance and no evidence of an effect of word duration (Baddeley, Lewis & Vallar, 1984; Vallar & Baddeley, 1984). These effects were explained by arguing that subjects generate a phonological representation of the visual input and retain it in short-term memory by a process of subvocal rehearsal. Articulatory suppression, it was argued, interfered with the maintenance of a phonological representation.

However, that the *spoken* duration of the items affects performance suggests that the to-be-remembered-items are encoded at a phonetic level. When both the number of syllables and the number of phonemes in a word are held constant, short-term memory span is inversely related to the spoken duration of the vowels in the words (Baddeley, Thomson, & Buchanan, 1975; Cowan, Day, Saults, Keller, Johnson, & Flores, 1992, Experiment 1, but see Caplan, Rochon, & Walters, 1992). It has also been shown that memory span in children increases with rate of speech (Hitch, Halliday, & Littler, 1989). Any relationship between task performance and the spoken duration of the experimental words is evidence that, at least, a quantitative phonetic representation has been generated. Articulatory suppression, therefore, interferes mainly with phonetic and articulatory encoding processes.

Phonological processes, on the other hand, can survive articulatory suppression. While the duration effect disappears under conditions of articulatory suppression, the phonological similarity effect can survive (Longoni, Richardson, & Aiello, 1993). Besner and Davelaar (1982) showed that with visual presentation, nonwords that are pseudohomophones (e.g., focks) are better recalled under conditions of articulatory suppression than nonpseudohomophone nonwords but that the duration effect disappeared. Moreover, experiments designed to investigate the role of phonological encoding in reading have shown that articulatory suppression has little or no effect on speed or accuracy in homophony or rhyme judging tasks: tasks which require the generation of some representation of the sound-form of the words to be judged (Baddeley & Lewis, 1981; Besner, Davis, & Daniels, 1981).

We would suggest that the results discussed above are consistent with the interpretation that articulatory suppression interferes with the phonetic encoding of written language while leaving phonological encoding processes relatively intact. Articulatory suppression could therefore be used as a tool for isolating phonological encoding processes within our monitoring task. We would also suggest that the level of form representation generated is dependent on the nature of the task at hand. In particular, when the task requires the retention of an unstructured list of items, a phonetic representation is more likely to be computed. The generation of a phonetic representation involves the translation of discrete phonological segments into a graded representation which maps the transition from one segment to another, and it is possible that the more detailed ordering information available in a phonetic representation is an aid to retention. However, when memory requirements are minimal—as they are in our task—it seems that a more abstract phonological code is computed.

MONITORING PHONOLOGICAL ENCODING

We report three experiments that required subjects to monitor their internal production of single words for a target sound. In Experiment 1, we demonstrated that a phonological rather than a phonetic or articulatory code underlies performance in this task. Articulatory suppression is shown to have a limited impact on performance and no relationship was observed between the pattern of monitoring latencies and the timing of the targets in subjects' overt speech. Experiment 2 demonstrates that the code subjects are monitoring is syllabified. These experiments allow us to conclude that subjects are basing their responses on the output from the process that assigns phonemes to their slots in a prosodic frame. Finally, Experiment 3 provides more detailed data concerning the left-to-right time course of this assignment process. Together, these experiments provide new evidence that phonological encoding involves the left-to-right assignment of phonemes to a syllabified structure and place new constraints on current models of phonological encoding.

GENERAL METHOD

The translation task. All of the experiments we report share the same basic methodology, which takes advantage of the average Dutch person's English language skills. The subjects' task was to silently generate the Dutch translation of an auditorily presented English word (spoken by a native English speaker) and to monitor their production for a prespecified target segment in the Dutch translation word. The English word therefore serves the same function as the picture stimulus more traditionally used in word production experiments (Levelt et al, 1991; Wheeldon & Monsell, 1992, 1994). It activates the form representation of the target Dutch word either directly or via a semantic/

conceptual representation (see De Groot, 1992, for a discussion of possible routes). Whether activation of the Dutch lexeme is semantically mediated is not relevant to our claims, provided phonological encoding processes remain unaffected. It was, therefore, important to ensure that the translation pairs were never cognate forms (*e.g.*, *house* → *huis*) so that subjects could not generate the Dutch target words on the basis of nonlexical translation rules. Because of potentially large between-subject differences in the translation speed for particular words, all comparisons are made both within word and within subject. The translation task has the advantage over picture naming of allowing a wider range of words to be used (i.e., not only picturable concrete nouns). This was essential, as there were heavy selection constraints on the materials for the experiments we report.

It is of course possible that subjects could generate nonspeech codes in order to complete the task. They may choose to construct a visual or a graphemic representation of the translation word and monitor this representation for letters. It is, however, unlikely that subjects would base their responses on a visual image of translation words. A number of experiments have shown that visualising letters takes two to three times as long as internally saying their names (Weber & Bach, 1969; Weber & Castleman, 1970), and even purely visual categorization tasks involving letters appear to be mediated by internal verbal rehearsal (Weber, Kelley, & Little, 1972). Nevertheless, subjects may generate a more abstract graphemic representation of the target word. There is little doubt that a close relationship holds between graphemic and phonemic codes for words. Phonological representations have been shown to affect the processing of visually presented words (Schneider & Healy, 1993; Van Orden, 1987) and, conversely, orthographic processes can affect the monitoring of auditorily presented words (Jakimik, Cole & Rudnicky, 1985; Seidenberg & Tanenhaus, 1979). Dutch has a fairly shallow orthography with regular spelling-to-sound correspondences which makes it difficult to test directly for graphemic effects. Instead, the experiments were designed to encourage subjects to generate speech rather than graphemic representations. Both the English words and the target phonemes were presented auditorily, and filler words were included which required subjects to monitor a devoiced word final /t/ in words written with a word final "d" (e.g., *avond*—evening, pronounced *avont*). The fact that the experiments to be reported yield effects of word stress (Experiment 1) and syllabification (Experiment 3) is evidence that subjects were indeed basing their responses on a phonological code.

Procedure. All three experiments consisted of the following stages. At the start of an experiment, subjects were given a list of English/Dutch word pairs to remember. When they were confident that they knew the pairings, subjects heard all English words once and produced the Dutch target words aloud. This procedure allowed subjects to practice the pairings and allowed the experimenter to check subjects' pronunciation. Naming latency and error rate for each translation pair were recorded. Subjects' speech was also tape recorded and digitized in order to allow measurements to be made of the duration intervening between target segments. During the experiment proper, on each trial, subjects were presented with an auditory description of the sound for which they had to monitor (see below for details), followed by an auditorily presented English word. Their task was to press a button if the Dutch translation they generated contained the target sound. Responses were measured from the onset of the English word. Exact descriptions of trial events will be given in the method section for each experiment.

Apparatus. The English stimulus words were presented using a Sony DTC-1000 ES DAT-recorder. Subjects' overt translations of the Dutch target words were recorded by a Sony DTC-55 ES DAT-recorder. An analogue voice-key registered voice onset and offset times during the overt translation phase. The experiment was controlled by a Hermac PC.

Data analysis. The analyses we report are based on data from correct response trials following some exclusions intended to reduce the noise in the data. Data from trials which immediately followed an error trial were removed, as errors can perturb subjects' responses on following trial. As the task was new, we were unable to set fixed outlier boundaries. Instead all data points beyond two standard deviations from the mean were counted as outliers and were also removed. Missing values were substituted by a weighted mean based on subject and item statistics calculated following Winer (1971, p. 488). The percentage of data lost due to these criteria will be reported for each experiment. Subjects with exceptionally long response times or high error rates were replaced. Similarly, problematic items were also removed. These exclusions and the statistics on which they were based will also be reported for each experiment. Both reaction time and percentage error rate are analyzed and we report separate analyses of variance treating subjects (F_1) and words (F_2) as random factors.

As the monitoring task requires some degree of skill, it is possible that subjects could develop strategies during the course of the experiment which might mask or exaggerate any effects due to phonological encoding processes. All experiments were designed such that the order of presentation of matched blocks of trials was rotated across subjects. This allowed us to detect any significant changes in subjects' performance block-for-block as the experiment progressed. To test for effects of practice, analyses of variance were performed including block position as a factor. Analyses of all three experiments yielded main effects of practice: Reaction times (but not error rates) decreased significantly as the experiment progressed. Importantly, however, we observed only one significant interaction of block position with monitoring latencies in the experimental conditions. Closer inspection of the data showed that this interaction was due to differences in the size rather than in the direction of the effect across blocks. Only this interaction with practice will be reported in the results sections below.

EXPERIMENT 1

In this experiment, the basic comparison was between the time taken to monitor bisyllabic words for first and second syllable-initial segments. This experiment had two aims. Our first aim was to support the claim that phonological encoding occurs from left to right and to provide new data detailing the time course of this process. Second, we wanted to show that subjects' responses are based on the production of an abstract phonological code rather than a phonetic/articulatory representation of the target word. The experiment therefore had two parts. Experiment 1b is a replication of Experiment 1a with the addition of an articulatory suppression task. The evidence we reviewed in the introduction suggests that articulatory suppression interferes with phonetic encoding processes but not with phonological encoding processes. If the internal monitoring task requires the construction of a phonetic code, then articulatory suppression should dramatically affect task performance. In addition subjects' overt production of the Dutch target words were tape recorded, in order to test for a relationship between monitoring times and actual phonetic durations.

Method

Vocabulary. Both word initial and second syllable initial target segments were monitored in all experimental words, so that the main comparison was within word. To avoid target-specific detection effects, each target segment was also monitored in both word positions. To be sure that any effect we observed was due to target position rather than target prominence, target segments in both word positions were matched for stress. For example, the target word *ca-deau* (gift) contains the targets unstressed /k/ and stressed /d/. Therefore, an unstressed /k/ and a stressed /d/ were monitored for in reversed positions in another word, *de-ken* (blanket). The experimental vocabulary consists of 16 such word pairs containing a total of 13 different target consonants (Appendix 1). 18 bisyllabic filler words were also selected.

Design. Each subject heard each experimental word three times: once to respond to the initial segment; once to respond to the second-syllable-initial segment and once when no response was required. The experiment consisted of 24 lists of between 5 and 12 words. The first three lists were practice lists containing only filler words. Experimental lists contained a maximum of four and a minimum of two experimental words to which a response was required. Within a list, experimental and filler words did not total more than 50% of the list items. For the remaining words no response was required. Experimental words never appeared in first position.

Three groups of seven lists were constructed with an experimental response word occurring only once in each group. In one group, a given word was monitored for its initial sound, in another group it was monitored for its second syllable initial sound, and in the remaining group, that word appeared as a no-response filler. Each filler item also appeared once in each group. The order of appearance of these groups was rotated across subjects so that each word was encountered first in each of the three conditions an equal number of times. Within these constraints the three orders of the lists were recorded with the condition that subjects never responded to the same target in two consecutive lists.

Procedure. Subjects first learned the translation pairs and completed the overt translation block. During the experiment proper all the English words were presented auditorily in lists of between 5 and 12 words. Subjects received three practice monitoring lists. Before a list subjects heard a description of the sound to which they should react. To avoid any effects of similarity between target description and particular target positions, each target was described, as follows, by a native Dutch speaker: "*Reageer nu op het geluid /l/ zo als in lepel, spelen, verhaal*" ("react now to the sound /l/ as in ladder, follow, barrel").[1] They were instructed to press a button whenever the word they generated contained the prespecified target segment. Subjects' reaction times were measured from the onset of the English word. Subjects were only told the maximum possible list length. They did not know when a given list would end. In Experiment 1a the sequence of events for each list was as follows. After the description of the target sound there was a 3-s pause followed by a high tone (600 Hz) for 50 ms to signal the start of the list. Three seconds later a low 50-ms tone (300 Hz) signalled the presentation of the first English word—the onset of which occurred 300 ms from tone offset. 4650 ms after the onset of the word, the low tone was repeated and the next word was presented 300 ms later. This sequence continued until the end of the list. The interval from word onset to the next word onset was exactly 5 s. There was a 10-s interval between lists. Subjects were given a longer break after the first 12 lists.

For Experiment 1b, the following changes were made to the procedure. Subjects were required to count from one onwards (in Dutch) starting prior to the presentation of the English cue-word and ending after they had made their response. A pretest of three subjects showed that subjects could articulate loudly enough to be recorded and still hear the input English

---

[1] It is difficult to pronounce or perceive stop consonants in isolation. Therefore, in the interest of clarity all target consonants were pronounced with a vowel like letter names.

word. Speed of articulation has been shown to be critical to obtain articulatory suppression effects (Besner et al., 1981). Subjects were, therefore, trained to articulate at a rate of approximately 200 numbers per minute (three numbers per second) by practicing in time with recorded beeps. Rates between 170 and 240 per minute were sufficient to yield the articulatory suppression effects in the STM and reading tasks discussed above. Following the description of the target sound there was a three second silence. Subject then heard a 300 Hz tone for 50 ms which signalled them to start counting. The onset of the English word occurred 1 s from tone onset. Three seconds from word onset, a second tone indicated to the subject to stop counting. There was then a 3-s pause before the next tone for onset of articulation. The time interval between English word onsets, was therefore 7 s.

Subjects. Twenty-four subjects were run in Experiment 1a (16 women and 8 men) and 24 in Experiment 1b (14 women and 10 men). They were native speakers of Dutch and were members of the Max-Planck subject pool. They were paid for their participation.

Results

Exclusion of data. Data from four subjects were replaced due to error rates of over 10% (two each from Experiment 1a and Experiment 1b). Data points were excluded from the analyses according to the principles described under General Methods resulting in the loss of 4.2% of data points in Experiment 1a (1.8% trials following an error, 2.4% outliers) and 5.2% of data points in Experiment 1b (2.6% trials following an error, 2.6% outliers).

Experiment 1a (no articulatory suppression). The mean monitoring latencies for targets in first and second syllable position of initial and final stress words are shown in Table 1. Targets in both word types were responded to faster when they occurred in word initial position than when they occurred in second syllable initial position.

TABLE 1

Monitoring Latencies (in ms) and Percent Error Rate (in Parentheses) for First and Second Syllable Target Phonemes of Initial and Final Stress Words in Experiments 1a and 1b

	Target position		
	1st syl onset	2nd syl onset	Difference
Experiment 1a (no articulatory suppression)			
Initial stress	1118 (2.1)	1268 (1.3)	150 (0.8)
Final stress	1169 (3.1)	1264 (3.6)	95 (0.5)
Mean	1143 (2.6)	1266 (2.5)	123 (0.1)
Experiment 1b (with articulatory suppression)			
Initial stress	1171 (3.6)	1278 (6.8)	107 (3.2)
Final stress	1239 (5.2)	1276 (7.3)	37 (2.1)
Mean	1205 (4.4)	1277 (7.0)	72 (2.6)

An ANOVA was performed on mean correct monitoring latencies, which included the variables target position (1st or 2nd syllable) and word-stress (initial or final). The main effect of target position was significant, $F_1(1,23) = 109.5, p < .001, F_2(1,15) = 25.5, p < .001$.[2] Mean latency to targets in initial-stress words was faster than latencies to targets in final stress words but there was no main effect of word-stress, $F_1(1,23) = 3.5, p > .05, F_2 < 1$. Although the effect

---

[2] Four of the carrier words in Experiment 1 had their initial target phoneme also appearing in word final position (*tapijt, toekomst, tekort,* and *dekbed*—underlyingly at least). It is therefore possible that faster monitoring for word initial segments can be attributed to some kind of benefit due to this repetition. The analyses of Experiments 1a and 1b were therefore repeated excluding these words and their partner words. Despite the large reduction in the data set, the pattern of results was similar to that reported above. Importantly, the effect of target position survived for both experiments. In Experiment 1a, mean monitoring latencies for word onset and word medial targets were 1156 and 1264 ms, respectively. The 108-ms difference in monitoring latencies was significant $F_1(1,23) = 78.3, p < .001, F_2(1,11) = 18.7, p < .01$. In Experiment 1b, mean monitoring latencies for word onset and word medial targets were 1216 and 1277 ms, respectively. The 61-ms difference in monitoring latencies was also significant $F_1(1,23) = 31.9, p < .001, F_2(1,11) = 5.2, p < .05$. These results, together with the replication of the effect in Experiment 3 with no repeated segments, allows us to rule out this alternative explanation.

of target position was slightly larger for initial-stress words than final-stress words the interaction of target position and word stress was only significant by subjects, $F_1(1,23) = 5.0, p < .05, F_2(1,15) = 3.0$. A similar analysis of percentage error rates yielded no significant effects.

An ANOVA was run including position of occurrence in the experiment as a factor in order to look for any effect of practice on monitoring performance. Data were analyzed over each seven-list block of the experiment. Mean monitoring latencies decreased significantly over the three blocks, $F_1(2,14) = 9.6, p < .01, F_2(2,22) = 11.7, p < .001$. There was a significant interaction of block with target position, $F_1(2,14) = 3.8, p < .05, F_2(2,22) = 6.4, p < .01$. Closer inspection of the data showed that this was due to differences in the size rather than in the direction of the effect across blocks (185, 80, 116 ms respectively). The analysis of error rates yielded no main effect of block and no significant interactions.

Experiment 1b (with articulatory suppression). The pattern of results is very similar to that of Experiment 1a (see Table 1). Responses to word-initial targets were again significantly faster than to second syllable initial targets, $F_1(1,23) = 51.6, p < .001, F_2(1,15) = 11.7, p < .01$. Mean latency to targets in initial-stress words was again faster than latencies to targets in final stress words and this difference was significant by subjects, $F_1(1,23) = 6.8\, p < .05, F_2 < 1$. In this experiment the interaction of target position and word-stress also reached significance by words, $F_1(1,23) = 12.0$, $p < .01$, $F_2(1,15) = 4.8$, $p < .05$. The difference in reaction time between first and second syllable initial targets was more pronounced for words stressed on the initial syllable than words stressed on the final syllable. An analysis of percent error rates yielded only a significant effect of target position over subjects, $F_1(1,23) = 6.5, p < .05, F_2(1,15) = 4.5, p = .051$.

Experiment 1a and 1b comparison. Grand mean naming latency and percentage error rates were 1205 ms (2.5%) in Experiment 1a compared to 1241 ms (5.7%) in Experiment 1b. The addition of the concurrent articulation task therefore increased naming latencies by 36 ms and error rates by 3.2%. ANOVAs were performed including the variable Experiment (1a and 1b). The increase in naming latency was significant by words, $F_1(1,46) = 1.2, F_2(1,15) = 5.6, p < .05$. The increase in error rates was significant both by subjects and by words, $F_1(1,46) = 5.7, p < .05, F_2(1,15) = 22.0, p < .001$. The monitoring latency analysis also yielded a significant interaction of Experiment and Target position, $F_1(1,46) = 10.6, p < .01, F_2(1,15) = 7.9, p < .05$. The 123 ms difference between target position in Experiment 1a was significantly larger than the 72 ms difference in Experiment 1b.

Spoken word duration. The results are consistent with the interpretation that subjects are monitoring the left-to-right production of some kind of speech code. The spoken duration of a word has been shown to affect digit span in STM memory tasks that involve the production of a phonetic speech code. If subjects are generating a phonetic representation in order to complete this task, then the difference between monitoring latencies for word initial segments and second syllable initial segments should be a function of the spoken duration of the intervening syllable. In order to test for such a relationship, subjects' overt productions of the experimental words from the initial phase of the experiment were digitized and measurements were made of the time lapse from the onset of the first syllable to the onset of the second syllable.[3] Mean first syllable durations were Experiment 1a, 167 ms; and Experiment 1b, 158 ms. Spoken durations were significantly longer than the difference in monitoring times between target segments (Experi-

---

[3] The measurements for Part 1a were made by the first author and Ger Desserjer; the measurements for Part 1b by Vincent Evers. All measurement were made using the same criterion. Although the word tokens measured were different, the correlation between the measurements for each item in Part 1a and Part 1b was $r = .99, p < .001$.

ment 1a, $t(23) = -10.5, p < .001$; Experiment 1b, $t(23) = -8.5, p < .001$). Despite significant correlations between the number of phonemes in the first syllable (long vowels were counted as filling two positions) and its spoken duration (Experiment 1a, $r = .7, p < .001$; Experiment 1b, $r = .7$, $p < .001$), number of phonemes showed no correlation with difference in monitoring latency (both Experiments 1a and 1b, $r = .2$, $p > .1$). Importantly, simple regressions of first syllable duration with the difference in monitoring latencies showed no evidence of any relationship (Experiment 1a, $r = .0$, $p = .9$; Experiment 1b, $r = .1, p = .5$).

Discussion

This experiment yielded an effect of target position that is both large and reliable. Word onset targets were responded to faster than second syllable onset targets. This was true for targets occurring in both stress initial, and stress final carrier words. This finding is consistent with the proposal that the encoding of the first syllable of a word is initiated before the encoding of the following syllable.

The results of this experiment provide two pieces of evidence which demonstrate that subjects are not basing their monitoring responses on a phonetic or articulatory code. First, the addition of an articulatory suppression task in Experiment 1b had only a limited effect on monitoring performance. The increases in naming latency and error rate were small and do not suggest any major disruption of task completion, and the overall pattern of results was very similar to that of Experiment 1a. This result contrasts with the effect of articulatory suppression on performance in short-term memory tasks. We attribute these effects to increased cognitive difficulty due to the addition of a secondary task. Second, the difference in monitoring times between targets bears no relation to the spoken duration of the first syllable of the target word. While we grant that strong conclusions should not be drawn on the basis of the absence of a correlation between two measures, our finding does contrast with the finding of a relationship between spoken word duration and short-term memory span.

There was also a tendency for initial stress words to show a larger effect of target position than final stress words, and this difference was significant in Experiment 1b. Closer inspection of the monitoring latencies in Table 1 suggests that this difference is due to relatively longer monitoring latencies for unstressed word initial segments. There are a number of possible explanations for this effect. It may be due to slower phonological encoding for words with infrequent stress patterns (Dutch has predominantly initial syllable stress). Alternatively it may be a monitoring effect. If monitoring is accomplished via an internal link to the comprehension system (see introduction) unstressed phonemes may simply be harder to detect. We will return to this issue under General Discussion. Nevertheless, this finding does suggest that stress is represented at the level of representation being monitored and therefore that a phonological rather than graphemic code is being generated.

Despite the overall similarity of the results of Experiments 1a and 1b, there was one significant effect of articulatory suppression on monitoring latencies: Experiment 1a yielded a larger effect of target position than Experiment 1b. Inspection of Table 1 suggests that this is due to a larger effect of articulatory suppression on monitoring latencies for word initial target segments than word medial target segments. Speech error data demonstrate that phonemes in word onset position are more vulnerable than phonemes in other word positions in that they are most likely to participate in errors (Shattuck-Hufnagel, 1992). It is therefore possible that increased difficulty due to the addition of the articulatory suppression task could cause the most disruption to word onset encoding.

We conclude that subjects are responding on the basis of the unfolding phonological representation of the target word. As reviewed in the introduction, there are two phonological codes postulated which could underlie performance in the translation task. Subjects could be responding to the initial release of phonemes following the access of a lexeme or to a syllabified phonological representation generated by the process that assigns phonemes to their syllabified prosodic frame (cf. Levelt, 1992; Levelt & Wheeldon, 1994). The next experiment was designed to test whether the phonological code that subjects are monitoring is syllabified.

EXPERIMENT 2

This experiment is a production variant of the classic syllable monitoring task in speech perception. Mehler, Dommergues, Frauenfelder, and Segui (1981) found that when subjects monitored for syllable targets (e.g., */pa/* and */pal/*) in spoken French carrier words (e.g., *pa-lace* and *pal-mier*) their responses were faster when the string of target segments corresponded to the first syllable of the target word (e.g., */pa/* in *pa-lace*, */pal/* in *pal-mier*), than when it did not (e.g., */pal/* in *pa-lace*, */pa/* in *pal-mier*). This finding has been replicated in other romance languages (Morias, Content, Cary, Mehler, & Segui, 1989; Sanchez-Casas, 1988). However, attempts to replicate the syllable match effect in English have failed (Cutler, Mehler, Norris, & Segui, 1986). Moreover, English speakers showed no effect of syllable match when monitoring French materials, while French speakers do show the effect on matched English materials. Cutler et al. (1986) introduced the idea of language-specific segmentation strategies to account for their results. They claimed that speakers make use of salient cues in the acoustic signal in order to decide at which point to access the lexicon. Romance languages are characterized by clear and unambiguous syllabification and speakers therefore adopt a syllable segmentation strategy. English, on the other hand, has many words with "ambisyllabic" consonants which are treated as part of the first as well as the second syllable, such as *ba*[*l*]*ance*, where the */l/* is an intervocalic consonant which follows a short stressed vowel. Such segments drastically reduce the efficiency of the syllable segmentation strategy in English.

More recent work using Dutch has provided evidence that processing strategies are not solely based on the availability of consistent segmentation cues in the acoustic input but can also be based on the phonological structure of the language. Zwitserlood, Schriefers, Lahiri, and van Donselaar (1993) exploited some salient differences between Dutch and English. Dutch, like English, has ambisyllabic segments. However, ambisyllabicity in Dutch is solely determined by the phonological quantity of the preceding vowel. Only syllables ending in long vowels are legitimate open syllables in Dutch (e.g., *boe-te*, penalty). Syllables containing short vowels must always be closed by a consonant and (according to the maximalization of onset principle) a single intervocalic consonant must therefore close the first syllable as well as open the second (e.g., *bo*[*t*]*e*, bones) (Van der Hulst, 1984). Moreover, unlike English, there are no cues available in the acoustic signal to determine whether a consonant is ambisyllabic. In English, ambisyllabic segments can become aspirated in contrast to their syllable-final counterparts (Gussenhoven, 1986). This does not occur in Dutch, and Jongman and Sereno (1992) have further shown that ambisyllabic Dutch consonants do not differ in duration from their monosyllabic counterparts. Nevertheless, Zwitserlood et al. (1993) obtained a syllable match effect: Dutch subjects were faster in monitoring for */buk/* than */bu/* in the carrier word *bu*[*k*]*en*. Dutch listeners are, therefore, sensitive to the syllable structure of their language and this sensitivity is based on the phonological rather than the phonetic structure of their language. Thus, if the phonological code our subjects are monitoring is syllabified, then we should be able to replicate the effect of syllable match in their monitoring of their own internal speech.

Method

Vocabulary. The experimental vocabulary consisted of 16 pairs of bisyllabic words (see Appendix 2). The words within a pair shared initial phonemes but differed in syllabification: For example the word pair *mager-maagden* (thin-virgins) have the same first three phonemes */maag/* but the first syllable of *mager* is *ma,* and the first syllable of *maagden* is *maag.* There were, therefore, two syllable targets associated with each word-pair, each of which matched the first syllable of one, but not the other word of a pair. All words had clear syllable boundaries. As we wanted to isolate the effect of syllable structure on monitoring latencies from effects of morphological structure, the first syllable of the target words was never a morpheme of the target word.

The 32 experimental words were divided into four groups of eight words. Each group contained only one member of a word-pair. Each group comprised four words with an open first syllable (e.g., *bei-tel,* chisel) and four words with a closed first syllable (e.g., *beit-sen,* to stain). In order to ensure that subjects encoded the whole of the target word, not just the first syllable, each syllable target also occurred in a medial or final position of a filler word (e.g., *ontbijt,* breakfast; *bijt* and *beit* have the same pronunciation). In order to ensure that subjects monitored for the whole syllable target, rather than just the initial phoneme, another filler word was selected for each syllable target, which shared its onset phoneme (e.g., *bedelaar,* tramp). Finally, for each target two words were selected which had no phoneme overlap with the target. Where possible each filler word served more than one purpose: in total 18 filler words were selected.

Design. Target-type (open or closed syllable) was crossed with carrier-word type (open or closed first syllable). Four word sets were constructed. In each set, all 32 experimental words occurred once combined with syllable targets such that 16 words required a "yes" response (eight with syllable overlap and eight without), and 16 words required a "no" response (on filler trials). Each group also contained 16 filler words, eight with a "yes" response and eight with a "no" response. These word sets were then divided in two experimental blocks, each one with 16 experimental words and eight filler words. The experiment therefore consisted of eight blocks of 24 trials. The order of presentation of blocks was rotated across subjects two blocks at a time. Six subjects received one of the four possible orders. Within the experiment each experimental syllable-target occurred six times: three times with a "yes" response and three times with a "no" response. Each experimental word occurred four times: two times with a "yes" response (one with, and one without, syllable match), and two times with a "no" response.

Procedure. After the translation practice phase, subjects completed a practice block of 24 filler words in which all experimental syllable targets occurred once. All syllable targets were produced in isolation by a native Dutch speaker and were presented auditorily. In the practice block and the following eight experimental blocks, events on each trial were as follows. Subjects first heard a warning beep (50 ms). 550 ms after the onset of this beep, the subjects heard the target syllable, and 1450 ms after the target onset they heard the stimulus word. Subjects then had 5 s in which to make their response and prepare for the onset of the next trial. One trial therefore took 7 s. After each block, subjects received feedback about their performance; their mean reaction time and number of errors were displayed on the screen. There were 24 subjects, 18 women and 6 men.

Results

Exclusion of data. Data from four subjects were replaced. Two had very slow mean monitoring latencies (1802 and 1786 ms, respectively, compared to a mean of 1376 ms—the next slowest subject was 1621 ms). Two were replaced due to high error rates (16 and 22% respectively). Data from one word were also excluded from the analysis due its extremely high error rate (*plateau*, 29%). In order to retain a balanced design, data from the other member of the pair were also excluded. Data from trials following an error trial were removed (5.3% of the data) as well as all data beyond two standard deviations from the mean (1.6% of the data).

An ANOVA was performed including the variables target type (open or closed syllable) and word type (open or closed first syllable). Mean monitoring latencies (and percent error rates) for the open and closed syllable targets were 1313 ms (5.8%) and 1305 ms (8.6%) respectively. The main effect of target type on monitoring latencies was nonsignificant, F_1 and $F_2 < 1$, as was the effect on percent error rates, $F_1(1,20) = 3.3$, $p > .05$, $F_2(1,28) = 2.8$.

Mean monitoring latencies and error rates for open and closed first syllable words were 1318 ms, 8.3%; and 1299 ms, 6.1%, respectively. The main effect of word type was again nonsignificant for monitoring latencies, $F_1(1,20) = 2.1$, $F_2(1,28) = 1.0$, and percent error rates, $F_1(1,20) = 2.1$, $F_2 < 1$.

Importantly, however, latencies to monitor syllable targets were faster when the target matched the first syllable of the carrier word than when it did not. The interaction between target type and word type was highly significant, $F_1(1,20) = 121.8$, $p < .0001$, $F_2(1,28) = 22.0$, $p < .001$ (See Fig. 1). In a similar analysis of percent error rate the same interaction was marginally significant in the subjects analysis only, $F_1(1,20) = 4.1$, $p = .056$, $F_2(1,28) = 1.8$.

Discussion

This experiment yielded a textbook interaction of syllable target and carrier word syllabification. Subjects were much faster in monitoring for a target string when it matched the first syllable of a target word than when it did not, regardless of the length of the target string.

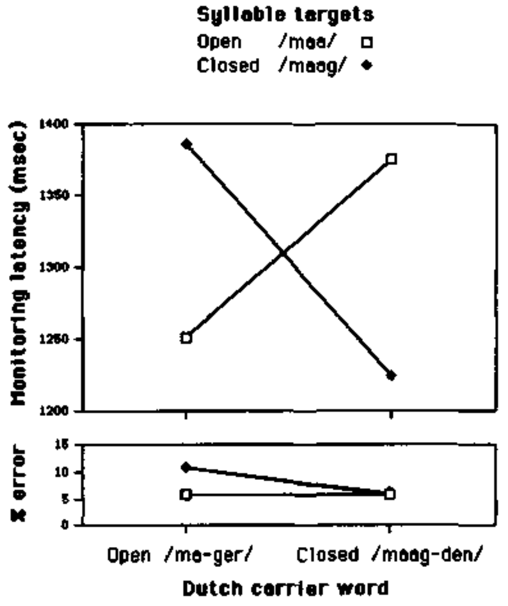

FIG. 1. Monitoring latencies and percent error rates in Experiment 2 are shown for the open and closed syllable targets in Dutch carrier words with open and closed first syllables.

The results of this experiment also support the claim that subjects are generating a phonological rather than a graphemic code. In Dutch, long vowels which occur in open syllables are written with one letter but are written with two letters when they occur in closed syllables (e.g., *ma*/*maag*). While the closed syllable targets did often differ in the graphemic match between open and closed syllable carrier words (e.g., /*maag*/ overlaps graphemically with *maag-den* but not with *ma-ger*), the graphemic overlap for the open syllable targets was the same in both open and closed syllable words (e.g., /*ma*/ in *ma-ger* and *maag-den*). Thus the effect of syllable overlap for these targets is best explained with reference to syllable structure.

Taken together, the results of Experiments 1 and 2 provide evidence that subjects are basing their responses on their generation of a syllabified phonological representation—the output from the process that assigns phonemes to their positions in a prosodic frame.

EXPERIMENT 3

The final experiment was designed to obtain more detailed data concerning the time course of the generation of a syllabified phonological representation. Experiment 1 demonstrated that onset phonemes are assigned to their syllable frames in a left-to-right manner. In this experiment we wanted to investigate the time course of segment-to-frame association both within and between a word's syllables. To do this, Dutch translation words with a CVC-CVC structure were used and all four consonants were monitored.

Method

Vocabulary. Twenty bisyllabic words were selected with the structure CVC-CVC (vowels could be either long or short) such that each of the four consonants within a word were different. The twenty words provided ten different target segments. Across words an attempt was made to match the occurrence of different segments in each of the four positions. Each segment occurred in each position at least once and no more than four times (see Appendix 3). Ten bisyllabic filler words were also included.

Design. Each subject responded to each word four times: one response to each segment position. In addition each subject encountered each word on four more trials where no response was required. The experiment consisted of 40 lists of between 4 and 12 words. These lists were combined into four blocks of ten. Within each block each of the 10 target segments were assigned to one list. The segments occurred in a different random order in each block. In each block all words occurred twice: once when a response was required and once when no response was required. Within a block, five experimental words were responded to in each of the position conditions. The order of blocks was again rotated across subjects.

Procedure. The procedure was the same as in Experiment 1a. There were 20 subjects, 16 women and 4 men.

Results

Exclusion of data. Three subjects were replaced due to error rates greater than 10%. 1.1% of the data were cases following an error and 2.3% were outliers. These data points were excluded from the analysis.

Analysis. Mean monitoring latencies and percent error rates for the four target phoneme positions are shown in Fig. 2. Monitoring latencies increase from left to right across target position. The effect of target position was significant, $F_1(3,48) = 24.1$, $p < .01$, $F_2(3,57) = 6.6$, $p < .01$. The size of the increase in monitoring latencies between the three pairs of consecutive phonemes is shown in Table 2.

Newman–Keuls pairwise comparisons showed that the increase from C1 to C2 was significant by subjects and approached significance by words, $Q_2(2,48) = 4.74$, $p < .01$, $Q_2(2,57) = 2.5$, $p > .05$, as was the increase from C2 to C3, $Q_2(2,48) = 4.83$, p

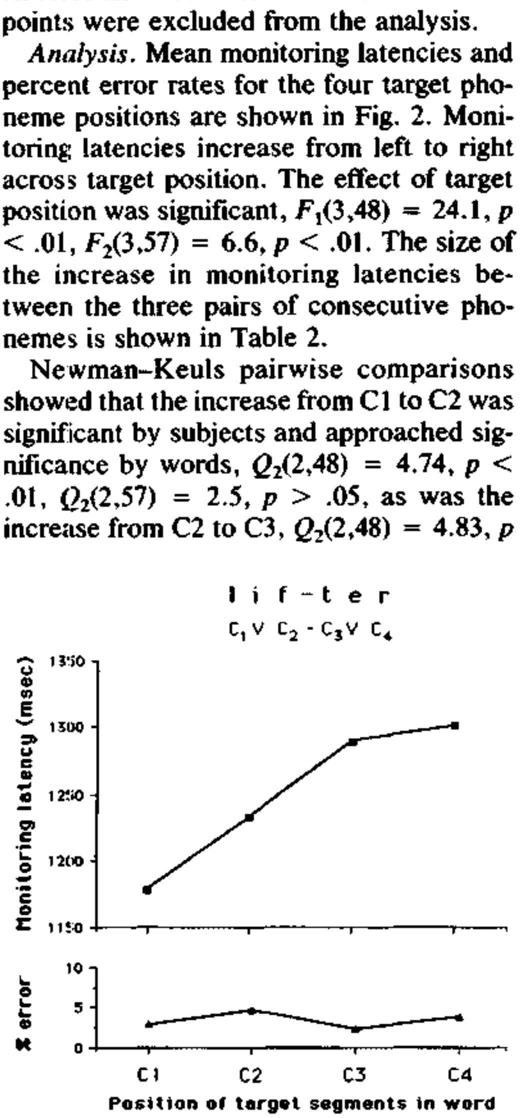

FIG. 2. Monitoring latency and percent error rates are shown for the four phoneme-target positions within the experimental words of Experiment 3.

TABLE 2

DIFFERENCE IN MONITORING LATENCIES (MS) AND SPOKEN DURATION BETWEEN THE THREE PAIRS OF CONSECUTIVE TARGET PHONEMES IN EXPERIMENT 3

	Difference scores		
	C1-C2	C2-C3	C3-C4
Monitoring latency	55	56	13
Spoken duration	140	69	190

$< .01$, $Q_2(2,57) = 2.5$, $p > .05$. The smaller increase from C3 to C4 did not approach significance, $Q_2(2,48) = 1.12$, $Q_2(2,57) = .6$. The increase from C1 to C4 was 111 msec. This difference is similar in size to the effect observed in Experiment 1a and was once again significant, $Q_3(2,48) = 9.57$, $p < .01$, $Q_3(2,57) = 5.0$, $p < .01$. A similar analysis of percent error rates yielded no significant effects.

Spoken word duration. Subjects' overt production of the translation words during the initial stage of the experiment was again recorded and digitized to allow measurements to be made. Table 2 shows the size of the increase in spoken duration between the three pairs of consecutive phonemes within a word.

Differences in monitoring latencies were again significantly shorter than spoken durations from C1 to C2, $t(19) = 2.9$, $p < .01$, and from C3 to C4, $t(19) = 4.5$, $p < .001$, but not from C2 to C3, $t < 1$. More importantly, the pattern of increases in spoken duration is clearly very different from the pattern of increases in monitoring latency. As in Experiment 1, although the spoken duration between the within syllable consonants was a function of the intervening vowel length (long or short: first syllable $r = .5$, $p < .05$; second syllable $r = .7$, $p < .001$), there was no correlation between the increase in monitoring latency and the increase in spoken word duration (C1 − C2, r = 0.1; C2 − C3, r = 0.0; C3 − C4, r = 0.1).

Discussion

The monitoring latencies for the four target positions tested increased from the beginning to the end of the word. This experiment yielded a difference in monitoring latencies between the first and second syllable onsets of a similar size to that observed in Experiment 1a. In addition, the increase in monitoring latencies was marginally significant between targets occurring in the onset and offset positions of the first syllable, and between the offset of the first syllable and the onset of the second syllable. There was no significant difference in monitoring latencies between the onset and offset targets of the final syllable. As in Experiment 1, the difference between monitoring latencies for consecutive target phonemes showed no correlation with the spoken duration measured from phoneme onset to phoneme onset. Moreover, the pattern of increases for monitoring latencies for consecutive targets was markedly different to the pattern for the speech measurements. If our measurements are reliable, the speed of segment to frame association is at least twice as fast as the speed of articulation. The timing of monitoring responses is clearly not dependent on the generation of a representation that encodes the surface timing of the target word.

On the basis of these results, we conclude that assignment of phonemes to the *first syllable of a word occurs in a left-to-*right manner and that encoding of the first syllable is usually completed before encoding of the second syllable begins. However, this left-to-right pattern appears to break down for the assignment of phonemes to the second syllable. Clearly, we do not want to claim that the left-to-right assignment process operating on the first syllable of a word, operates in a parallel fashion on later syllables. A more plausible explanation is that the left-to-right assignment process somehow speeds up towards the end of the word, becoming too fast to be detected by our task.

Interestingly, the difference in monitoring latencies for the consecutive phoneme targets flanking the syllable boundary was the same size as the difference between the targets of the first syllable which were separated by a vowel. We know from Experiment 2 that subjects are monitoring a syllabified representation. The monitoring difference between the targets at the syllable boundaries is consistent with the existence of a marked syllable boundary or some process of syllabification which delays encoding of the second syllable.

GENERAL DISCUSSION

In this paper we have introduced a new methodology which yields data concerning the time course of phonological encoding during language production. Several aspects of the data allow us to conclude that the representation underlying the monitoring response is both phonological in nature and syllabified. The conclusion that the code is phonological is arrived at by a process of elimination. The limited impact of a concurrent articulation task on performance (Experiment 1b) excludes the possibility that subjects are monitoring a subvocal articulation of the Dutch translation word. The possibility that the code is nonmotoric but nevertheless phonetic in nature is rendered unlikely by the absolute lack of a relationship between the pattern of monitoring latencies and the timing of the targets in subjects' overt speech (Experiments 1a, 1b, and 3). Subjects are therefore monitoring an internal abstract speech code. Experiment 2 replicated the syllable monitoring effect observed in speech perception (Zwitserlood et al., 1993), demonstrating that the code subjects monitor is also syllabified. Further evidence of an on-line syllabification process is provided by the relatively large difference in monitoring latencies between the consecutive phoneme targets flanking the syllable boundary (Experiment 3). We therefore conclude that subjects are basing their responses not on the initial availability of a word's constituent phonemes, but on the output from the process that assigns these phonemes to a syllabified prosodic frame.

Let us turn now to the evidence concerning the time course of this assignment process within the syllables of a word. Experiment 1 demonstrates that the assignment of initial phonemes to the first syllable of a word precedes assignment of initial phonemes to its second syllable. This is true regardless of whether the Dutch carrier words were stressed on their initial or final syllable, although the difference in monitoring latencies were larger for stress initial carrier words (Experiment 1b). Experiment 3 provides information concerning the assignment process both within and between syllables. The phonemes of the first syllable of a word are assigned to their frame in a left-to-right manner, and the assignment is usually completed before encoding of the second syllable begins. The assignment of initial phonemes to the second syllable of a word does not occur significantly faster than assignment of final phonemes, suggesting that the speed of the assignment process increases towards the end of a word.

These findings place new constraints on the mechanisms that have been proposed to govern phonological encoding during speech production. What plausible account might be made of them? Levelt (1992) and Levelt and Wheeldon (1994) propose that a word's prosodic frame is made available syllable-by-syllable and that a word's phonemes are made available from left-to-right. Phonemes are assigned to their slots in the syllable frame as they become available. No claims are made about the relative time course of these processes. Two additional assumptions are required in order to accommodate this model to our experimental results. First, we must assume that the setting of a syllable boundary after the encoding of the first syllable delays the initiation of assignment of phonemes to the following syllable. Second, we must assume that the constituent phonemes of a word continue to be made available while this process occurs. Thus, when assignment of segments to the second syllable frame begins, all or most of its phonemes are already available and their left-to-right assignment to the frame can occur at greater speed.

An activation spreading account of our results is also possible. In Dell's (1988) update of his (1986) activation spreading model of phonological encoding, an activated morpheme node spreads activation (via syllable nodes) to its constituent phoneme nodes and also to a word-shape node, consisting of a string of consonant–vowel slots. These word-shape nodes spread activation to phoneme category nodes marked for prevocalic, vowel, or postvocalic position, which in turn send additional activation to phoneme nodes. The most highly activated phoneme nodes are then associated to the appropriate slots in the word-shape. Association of phonemes occurs in parallel. However, in multisyllabic words, constituent syllable nodes are activated serially. After the phonemes for the first syllable have been assigned to their slot in the syllable frame they are deactivated and the second syllable receives activation, becoming the current node. This model contains mechanisms which could, in principle, account for some aspects of our data. Phonemes are assigned to the word-initial syllable before assignment of phonemes to the following syllable is initiated, and time is required between syllable encoding to deactivate the first syllable and to allow the second syllable to become current. However, the model does not predict the left-to-right assignment of phonemes within the first syllable of a word. Two relatively simple modifications are required in order to bring this model into line with our results: first that activation of a syllable node triggers the ordered left-to-right activation of its constituent phoneme nodes, and second that the timing of the activation of phoneme nodes is independent of the speed of activation of syllable frames, allowing the phonemes of the second syllable to become activated before their frame is available.

However, interpretation of the time course data is complicated by questions concerning the nature of the monitoring system. We cannot simply assume that the monitor works directly on the output from phonological encoding processes. Alternatively, monitoring may be accomplished via an internal loop to the comprehension system (see introduction) and we must consider the possibility that some aspects of our results reflect a combination of production and comprehension processes.

The interaction of target position and word stress in Experiment 1 is a possible candidate. In this experiment, the difference in reaction time between first and second syllable initial targets was larger for words stressed on the initial syllable than words stressed on the final syllable. In speech perception experiments, phoneme monitoring latencies can be faster for targets in stressed syllables than targets in unstressed syllables (Cutler, 1976; Shields, McHugh, & Martin, 1974). Therefore, in our task, monitoring latencies for the incoming phonological code may be modified by the slow detection of unstressed phoneme targets. For carrier words with final stress, such a comprehension effect would slow the detection of the unstressed word-initial segments and speed the detection of the second syllable initial stressed phoneme relative to the same targets in the carrier words with initial stress. This would result in the smaller effect of target position that we observed for final stressed words compared to initial stressed words.

Perceptual effects may also have influenced the pattern of results observed in Experiment 3. It is possible, for example, that these results also reflect edge effects in scanning the incoming phonological representation. Edge effects could occur in our task, if first and final phonemes are perceived more clearly than word internal phonemes because they suffer less interference from adjacent phonemes. Such an edge effect would predict the larger difference in monitoring latencies between C1 and C2 than between C3 and C4 that we observed because monitoring latencies for the first and final phonemes (C1 and C4 respectively) would be speeded relative to latencies for the word internal phonemes (C2 and C3). However, a perceptual edge-effect explanation cannot predict the complete pattern of results observed in Experiment 3. It would predict longer monitoring latencies for C3 than for C4, which is not what we observed. It would also predict that the difference in monitoring latencies between C1 and C2 should be greater than the difference between C2 and C3 which is also not what we observed. At best, therefore, edge effects may modulate the left-to-right processing of the input representation. Future experimentation with words in connected inner speech production, could determine any contribution of perceptual edge effects to our results.

Finally, a comprehension explanation must also be considered for the speed-up in monitoring latency for the final phoneme of a carrier word. If the comprehension system reacts to the incoming phonological representation in the same way as it reacts to an acoustic speech signal, then the speed-up in processing for word-final phonemes (which would occur after the word's uniqueness point, see Frauenfelder et al., 1990) may be a comprehension effect. In other words, responses to word-final phonemes may be based on phonological information made available following lexical access in the comprehension lexicon. However, it is not obvious that post-access phoneme inference could overtake the detection of word-final phonemes in internal speech. It depends on the relative speed of generation of internal speech and the release of segmental information following word recognition. This ratio may well be different for overt speech and internal speech and is an issue for future investigation.

It is therefore possible that details of the time course data reflect comprehension processes, and further research is undoubtedly required in order to illuminate the workings of the internal speech monitor. Nevertheless, the main results of these experiments are best explained with reference to speech production processes. They provide converging evidence that the process of phonological encoding involves the left-to-right assignment of phonemes to a syllabified prosodic frame.

APPENDIX 1

Vocabulary for Experiment 1a and 1b: The English Stimulus, the Syllabified Dutch Target Word, and the Target Phonemes (Targets from Stressed Syllables Are Marked With ')

English word	Dutch word	Target phonemes	
visible	zicht-baar	'z	b
visit	be-zoek	b	'z
roofs	da-ken	'd	k
gift	ca-deau	k	'd
future	toe-komst	't	k
office	kan-toor	k	't
between	tu-ssen	't	s
lemon	ci-troen	s	't
chain	ke-ting	'k	t
shortage	te-kort	t	'k
softer	zach-ter	'z	t
unless	ten-zij	t	'z
body	li-chaam	'l	g
similar	ge-lijk	g	'l
stove	ka-chel	'k	g
dressed	ge-kleed	g	'k
together	sa-men	's	m
perhaps	mi-sschien	m	's
to talk	pre-ten	'p	t
carpet	ta-pijt	t	'p
accident	toe-val	't	v
departure	ver-trek	v	't
useful	nu-ttig	'n	t
stage	to-neel	t	'n
windows	ra-men	'r	m
swamp	moe-ras	m	'r
to glue	lij-men	'l	m
environment	mi-lieu	m	'l

chilli	sam-bal	's	b
definitely	be-slist	b	's
quilt	dek-bed	'd	b
covered	be-dekt	b	'd

APPENDIX 2

The Experimental Vocabulary of Experiment 2: English Words, Dutch Translation Words, and Target Syllables Are Given for Target Word Pairs Differing in Syllabification

English word	Dutch target	Syllable targets
eagle	a-rend	a / aar
nice	aar-dig	
chisel	bei-tel	bei / beit
to stain	beit-sen	
moose	e-land	e / eel
Callused	eel-tig	
thread	ga-ren	ga / gaar
orchards	gaar-den	
tentpeg	ha-ring	ha / haar
fireplaces	haar-den	
meager	ha-rig	ka / kaar
maps	kaar-ten	
hangover	ka-ter	ka / kaat
bounce	kaat-sen	
nonsense	la-rie	la / laar
boots	laar-zen	
thin	ma-ger	ma / maag
virgins	maag-den	
rogue	re-kel	re / reek
sequences	reek-sen	
crowd	scha-re	scha / schaar
scarcity	schaar-ste	
previous	vo-rig	vo / voor
over	voor-bij	
supper	di-ner	die / dien
services	dien-sten	
huge	e-norm	e / een
ducks	een-den	
table-land	pla-teau	pla / plaat
to put	plaat-sen	
troubles	so-res	so / soor
pedigrees	soor-ten	

APPENDIX 3

Experimental Vocabulary for Experiment 3: The English Stimulus, the Syllabified Dutch Target Word, and the Target Phonemes

English word	Dutch word	Target phonemes			
		CV	C−	CV	C
garbage	vuil-nis	v	l	n	s
quiet	rus-tig	r	s	t	g
numerous	tal-rijk	t	l	r	k
haircut	kap-sel	k	p	s	l
cyclist	fiet-ser	f	t	s	r
hitch hiker	lif-ter	l	f	t	r
backache	rug-pijn	r	g	p	n
waiter	kel-ner	k	l	n	r
outfit	Kos-tuum	k	s	t	m
paint brush	pen-seel	p	n	s	l
shrimp	gar-naal	g	r	n	l
sale	ver-koop	v	r	k	p
garden wall	tuin-muur	t	n	m	r
napkin	ser-vet	s	r	v	t
manufacture	maak-sel	m	k	s	l
cream cheese	room-kaas	r	m	k	s
to go wrong	mis-gaan	m	s	g	n
anonymous	naam-loos	n	m	l	s
expert	vak-man	v	k	m	n
adorable	poes-lief	p	s	l	f

REFERENCES

ANDERSON, R. D. (1982). Speech imagery is not always faster than visual imagery. *Memory and Cognition*, **10**, 371–380.

BAARS, B. J., MOTLEY, M. T., & MACKAY, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behaviour*, **14**, 382–391.

BADDELEY, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 362–365.

BADDELEY, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society London B*, **302**, 311–324.

BADDELEY, A. D., & LEWIS, V. (1981). Inner active processes in reading: The inner voice, the inner ear, and the inner eye. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, NJ: Erlbaum.

BADDELEY, A. D., LEWIS, V. J., & VALLAR, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, **36**, 233–252.

BADDELEY, A. D., THOMSON, N., & BUCHANAN, M. (1975). Word-length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, **14**, 575–589.

Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology*, 36, 701–711.

Besner, D., Davis, J., & Daniels, S. (1981). Reading for meaning: the effects of concurrent articulation. *Quarterly Journal of Experimental Psychology* 33a, 415–437.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology*, 45A, 177–192.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55, 75–84.

Conrad, R. (1970). Short-term memory processes in the deaf. *British Journal of Psychology*, 61, 179–195.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, 55, 429–432.

Cowan, N., Day, L., Saults, S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, 31, 1–17.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, 20, 55–60.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllables differing role in the segmentation of English and French. *Journal of Memory and Language*, 25, 385–400.

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

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

Dell, G. S., 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., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42, 287–314.

Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed) *Experimental Slips and Human Error: Exploring the Architecture of Volition*.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behaviour*, 19, 624–634.

Ellis, A. W., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative case of mental calculation in Welsh and English. *British Journal of Psychology*, 71, 43–52.

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

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.

Garrett, M. (1975). The analysis of sentence production. In G. Bower (Ed.) *Psychology of learning and motivation: Vol 9*. New York: Academic Press.

De Groot, A. M. B. (1992). *Bilingual lexical representation: A closer look at conceptual representations*. In R. Frost and L. Katz (Eds). Orthography, Phonology, Morphology, and Meaning. Elsevier Science Publishers B. V.

Gussenhoven, C. (1986). English plosive allophones and ambisyllabicity. *Grammar*, 10, 119–141.

Hardiik, C. D., & Petrinovitch, L. R. (1970). Subvocal speech and comprehension level as a function of the difficulty of the reading material. *Journal of Verbal Learning and Verbal Behavior*, 9, 647–652.

Hitch, G. J., Halliday, M. S., & Littler, J. E. (1989). Item identification time and rehearsal as predictors of memory span in children. *Quarterly Journal of Experimental Psychology*, 41A, 321–327.

Hulst, van der, H. (1984). *Syllable structure and stress in Dutch*. Dordrecht, The Netherlands: Foris.

Jackendoff, R. (1987). *Consciousness and the computational mind*. MIT Press.

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition, *Journal of Memory and Language*, 24, 165–178.

Jongman, A., & Sereno, J. A. (1992). Compensatory effects due to vowel length. *Proceedings of the International Congress of Phonetics*, 2, 294–297.

Keating (1988). Underspecification in phonetics. *Phonology*, 5, 275–292.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14, 185–209.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.

Launder, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In

V. A. Fromkin (Ed.) *Errors in linguistic performance*. New York: Academic Press.

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

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

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, **42**, 1–22.

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

Levelt, W. J. M., & Wheeldon, L. R. (1994). Do speakers have access to a mental syllabary? *Cognition*, **50**, 239–269.

Longoni, A. M., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory and Cognition*, **21**, 11–22.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behaviour. *Psychological Review*, **89**, 483–506.

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

Mehler, J., Dommergues, J., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behaviour*, **20**, 298–305.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, **42**, 181–211.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *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., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asychrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 1146–1160.

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

Morias, J., Content, A., Cary, L., Mehler, J., & Segui, J. (1989). Syllabic segmentation and literacy. *Language and Cognitive Processes*, 4, 57–67.

Motley, M. T. (1980). Verification of "Freudian slips" and semantic prearticulatory editing via laboratory induced spoonerisms. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.

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

Sanchez-Casas, R. M. (1988). *Access representations in visual word recognition*. Doctoral dissertation, Monash University, Australia.

Schneider, V. I., & Healy, A. F. (1993). Detecting phonemes and letters in text: Interactions between different types and levels of processes. *Memory and Cognition*, **21**, 739–751.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory* **12**, 315–328.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech*. New York: Springer.

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

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, **42**, 213–259.

Shields, J., McHugh, A., & Martin, J. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, **102**, 250–255.

Sokolov, A. N. (1972). *Inner speech and thought*. New York: Plenum Press.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, **1**, 121–141.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, **15**, 181–198.

Weber, R. J., & Bach, M. (1969). Visual and speech imagery. *British Journal of Psychology*, **60**, 199–202.

WEBER, R. J., & CASTLEMAN, J. (1970). The time it takes to imagine. *Perception & Psychophysics*, **8**, 165–168.

WEBER, R. J., KELLEY, J., & LITTLE, S. (1972). Is visual image sequencing under verbal control? *Journal of Experimental Psychology*, **96**, 354–362.

WHEELDON, L. R., & MONSELL, S. (1992). The locus of repetition priming of spoken word production. *Quarterly Journal of Experimental Psychology*, **44A**, 723–761.

WHEELDON, L. R., & MONSELL, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language*, **33**, 332–356.

WICKELGREN, W. A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, **38**, 583–588.

WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

ZWITSERLOOD, P., SCHRIEFERS, H., LAHIRI, L., & DONSELAAR, W. VAN (1993). The role of the syllable in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 260–271.

(Received August 24, 1993)

(Revision received April 21, 1994)