JOURNAL OF MEMORY AND LANGUAGE **37,** 356–381 (1997)
ARTICLE NO. ML972517

Prosodic Units in Speech Production

Linda Wheeldon

University of Birmingham, United Kingdom

and

Aditi Lahiri

University of Konstanz, Germany

Four experiments demonstrate effects of prosodic structure on speech production latencies. Experiments 1 to 3 exploit a modified version of the Sternberg et al. (1978, 1980) prepared speech production paradigm to look for evidence of the generation of prosodic structure during the final stages of sentence production. Experiment 1 provides evidence that prepared sentence production latency is a function of the number of phonological words that a sentence comprises when syntactic structure, number of lexical items, and number of syllables are held constant. Experiment 2 demonstrated that production latencies in Experiment 1 were indeed determined by prosodic structure rather than the number of content words that a sentence comprised. The phonological word effect was replicated in Experiment 3 using utterances with a different intonation pattern and phrasal structure. Finally, in Experiment 4, an on-line version of the sentence production task provides evidence for the phonological word as the preferred unit of articulation during the on-line production of continuous speech. Our findings are consistent with the hypothesis that the phonological word is a unit of processing during the phonological encoding of connected speech.

In order to produce fluent speech, pronounceable rhythmic articulatory gestures must be constructed from discrete lexical/phonological representations. Current models of speech production postulate similar mechanisms for sentence production. They assume that stored lexical representations are activated and assigned to positions in the evolving syntactic representation of the utterance followed by the retrieval of their phonological form (Dell, 1986, 1988; Levelt, 1989, 1992; Roelofs, 1992; Shattuck-Hufnagel, 1992). However, until recently no suggestions were made as to how this ordered string of discrete lexical representations is transformed into a rhythmic continuous utterance. The issue is not trivial. It cannot be done by simply concatenating the stored phonological representations of lexical items. This is because in connected speech the canonical sound-form of a word can undergo transformations which change its segmental content and can even be restructured in such a way that lexical and syllable boundaries do not coincide. For example, the sentence comprising the lexical items given in (1a) may, in fluent conversational speech, be articulated as the utterance given (1b).

1a. Get me a beer, if the beer is cold
1b. gɛʕ mi ə bɪə ɪf ðə bɪə rɪz koʊld

As can be seen, a number of changes have occurred to the phonological form of the original words. First, there are a number of changes in the segmental content of the words. The final /t/ of *get* has become a glottal stop and

This work was carried out during the authors' overlapping time at the Max-Planck Institute for Psycholinguistics in the Netherlands. The authors wish to thank Ger Desserjer for running the experiments. Thanks are also due to Fernanda Ferreira and two anonymous reviewers for their comments on the manuscript.

Address reprint requests and correspondence to Linda Wheeldon, School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom.

0749-596X/97 $25.00

a word final /r/ has been added to the second *beer* which does not occur (in certain dialects of British English) when the word *beer* is pronounced in isolation. Second, the syllable structure of some words has also been altered. The final segment of the second *beer* has resyllabified to become the onset of the following syllable /bɪə–rɪz/. Such phenomena are very common in fluent speech and cannot be explained in terms of low level articulatory accommodation because the same change need not occur to the first *beer*, despite identical segmental contexts (i.e., in both cases *beer* is followed by the same vowel).

The purpose of this restructuring of lexical form must be to prepare for articulation by producing strings of fluently pronounceable syllables (Levelt & Wheeldon, 1994). An adequate theory of connected speech production must provide an account for these and similar phenomena. Because much of the restructuring occurs across lexical boundaries and is not conditioned by the immediate segmental environment alone, any reasonable account of these phenomena must refer to structures larger than lexical items. Until recently, the only supra-lexical structures referred to in most current models of speech production were syntactic structures. However, developments in phonological theory (summarized below) suggest that syntactic structures alone cannot provide an account for many phenomena that occur during connected speech production. Instead, it has been proposed that these phenomena arise during the construction of the rhythmic (or prosodic) structure of an utterance (Inkelas & Zec, 1990; Nespor & Vogel, 1986; Selkirk, 1986). The claim is that, following the generation of the syntactic structure of an utterance, a nested hierarchy of prosodic units is generated and it is these prosodic units which guide the generation of the phonological form of the utterance.

The goal of this article is interdisciplinary in nature. The aim is to test whether prosodic constituents motivated by linguistic analyses have consequences for sentence production. In what follows, we will first discuss current linguistic views with respect to prosodic constituents built on surface syntactic structure. We will then review current psycholinguistic theories of sentence production. Finally, we will motivate our use of both prepared and on-line speech production paradigms to investigate the role of prosodic units in speech production.

PROSODIC CONSTITUENTS IN SENTENCES

Under the standard theory, the prosodic hierarchy consists of the following constituents: Phonological word [ω] → (Clitic Group) → Phonological Phrase [φ] → Intonational Phrase [IP] → Utterance [U]. These prosodic constituents may serve as domains of phonological rules. The phonological word is considered to be the smallest prosodic unit and it is generally assumed to be *at least* as large as a lexical word. The next in line is the clitic group. According to some scholars the clitic group can be subsumed under the phonological word (cf. Selkirk, 1986) and is therefore placed within parentheses. We will also argue that at least for Dutch, the phonological word includes clitics as well. The phonological phrase is well motivated for several languages and is the domain of many phonological rules. The intonational phrase, unlike the phonological phrase, is subject to semantic well formedness and is not just based on surface syntactic structure (Selkirk, 1986). The utterance is the largest prosodic constituent which may contain more than one intonational phrase. Usually the utterance corresponds to a grammatical sentence, but it may span more than one sentence, as has been argued by Nespor and Vogel (1986), Selkirk (1980), Odden (1980), and others.

Prosodic constituents are derived from syntactic constituents but are not necessarily isomorphic to them. The distinction between syntactic and prosodic structure can be seen in example (2) below. The broad syntactic phrasing is given in (2a) while the prosodic grouping is given in (2b). The syntactic bracketing is far richer than the prosodic bracketing and is, moreover, quite different from what the phonology requires.

2a. [[[The man]$_{NP}$ [[I]$_{NP}$ [[talked to]$_{V}$ [in the school]$_{PP}$]$_{VP}$]$_{S}$]$_{NP}$ [is ill]$_{VP}$]$_{S}$

2b. [[[[The man]$_\omega$ [I talked to]$_\omega$]$_\varphi$ [[in the school]$_\omega$]$_\varphi$]$_{IP}$ [[[is ill]$_\omega$]$_\varphi$]$_{IP}$]$_U$

The strongest motivation for prosodic constituents is that there exists a class of phonological rules which do not refer directly to syntactic structures but may refer to phrases of several syntactic constituents or to a string which corresponds to *no* syntactic constituent (Nespor and Vogel, 1986).

The phonological word is the subject matter of our paper and we will discuss it in some detail. We define a phonological word as the head of the minimal prosodic constituent above the foot, to which clitic-like words (usually unstressed function words) can attach. All full lexical words are phonological words which must be minimally one foot.[1] In English, non-phrase-final cliticization is usually rightward and Selkirk argues that clitics are attached to a lexical phonological word but do not become part of one: [clitic[Lex]$_\omega$]$_\varphi$. Examples would be structures like, *for Timothy,* and, *can pile.* There can be, however, encliticization as in sentences like, *John and Mary need him,* where the final pronoun is cliticized to the preceding phonological word as [[[need]$_\omega$ 'm]$_\omega$]$_\varphi$. Here the clitic attaches and becomes part of the phonological word (Selkirk, 1995, p. 447).

Our experiments were conducted in Dutch and the cliticizations concerned are always leftward (Booij and Lieber, 1993; Lahiri, Jongman, & Sereno, 1990; Gussenhoven, 1989; Berendsen, 1986). In Dutch, cliticization induces phonological word formation and has the following structure: [[[Lex]$_\omega$ clitic]$_\omega$]$_\varphi$. The examples in (3) illustrate this cliticization process.

3a. Ik [[[zoek]$_\omega$ het]$_\omega$]$_\varphi$ water
(I seek the water)

3b. Ik [[[drink]$_\omega$ de]$_\omega$]$_\varphi$ wijn
(I drink the wine)

3c. Ik [[[trap]$_\omega$ te]$_\omega$]$_\varphi$ hard
(I kick too hard)

Phonological cliticization mostly involves function words such as auxiliaries, pronouns, and conjunctions. If that is so, one might ask why we need to refer to constituents such as phonological words instead of using the distinction between content words vs function words. The reason is that a function word need not be a clitic but can also be a full phonological word. Under certain circumstances, a function word can bear stress and can therefore be minimally a foot and thereby a phonological word. Under focus, function words are always phonological words and at the end of phrases function words are often full phonological words. In the following example, the focused (4a) and phrase final (4b) constituents are italicized.

4a. I *can* eat it.
4b. Wherever she *is,* she will be admired.

Under no circumstances can it be the case that the above function words will be cliticized to an adjoining word: they must be full prosodic words. Thus, a content/function word distinction is insufficient to capture the prosodic groupings of a syntactic string.

MODELS OF SENTENCE PRODUCTION

Most psycholinguists interested in sentence production have studied the processes involved in grammatical encoding. These processes include the selection of lexical concepts and the generation of a syntactic structure appropriate for conveying the speaker's intended meaning or ''message'' (see Bock & Levelt, 1994, for a review). Following Garrett (1980), most models divide these processes into two stages, the functional and the positional. During functional processing, appropriate lexical concepts are retrieved from the mental lexicon and are assigned grammatical roles such as subject or object. During positional processing, the surface order of lexical items is determined. A hierarchical syntactic structure for the sentence is generated setting the positions of the lexical items and their grammatical inflections. The output from gram-

[1] As we mentioned above, this prosodic constituent is also called the clitic group in the literature (Hayes, 1989; Nespor & Vogel, 1986). There is, however, not enough motivation to claim another prosodic constituent in between the phonological word and the phonological phrase. We will assume, therefore, that clitics are incorporated into a neighboring phonological word (following Booij & Lieber, 1993; Gussenhoven, 1989; Lahiri, Jongman, & Sereno, 1990; Selkirk, 1989, 1995).

matical encoding is, therefore, a completed surface syntactic structure. Before the sentence can be articulated, however, this representation must be given phonological form. This process is known as phonological encoding.

One theory of phonological encoding has been proposed in which prosodic units are given an explicit role. In Levelt's (1989, 1992) model, the main input to phonological encoding is the surface syntactic structure with its associated lexical concepts. As the surface structure becomes available, the lexical concepts trigger access of their form representations. These representations release two separate kinds of information about a word's sound-form; its rhythmic structure (i.e., number of syllables and stress pattern) and its segmental content. Phonological encoding then involves the assignment of a word's segments to positions in a frame that specifies its rhythmic structure (see Dell, 1986, 1988; Roelofs, 1992; Shattuck-Hufnagel, 1992, for other slot-filler models of phonological encoding).

Levelt (1989, 1992) argues that the unit of phonological encoding is the phonological word. He postulates a prosody generator that takes as input the rhythmic information about the selected words (as well as surface syntactic information) and combines them into phonological word frames. The phonological segments for each word are made available separately and then associated to the newly constructed phonological word frames in a left to right manner. For example, in the utterance, *I gave it to him*, the four lexical items resyllabify to form one phonological word [ai-gei-vɪ-tɪm]$_{\omega}$ with one main stress.

Levelt (1989) makes one further claim concerning phonological words—that they are the minimal unit of articulation. As the segments for each syllable are associated to their prosodic frame they are used to retrieve stored, syllable-sized, articulatory routines (Crompton, 1982; Levelt & Wheeldon, 1994). The phonetic plan specifies, syllable by syllable, the articulatory gestures and their segmental and prosodic parameters as well as the global rate of articulation. When the articulatory routines for the entire phonological word have been retrieved, the phonetic plan is passed on to the articulator and executed. Thus, during the production of connected speech, a whole phonological word is constructed before articulation commences.

Few experimental studies provide empirical data directly relevant to the production of prosody. One study which sought evidence for effects of prosodic structure in speech production used duration measurements as the independent variable. Ferreira (1994) tested whether rhythmic structure in spoken sentences (i.e., duration of words and pauses) is best explained in terms of syntactic or prosodic phrasal structure. As her tool, she used the phenomenon of phrase final lengthening. This refers to the finding that a word and its following pause tend to have longer durations at the end of a syntactic phrase than in any other phrasal position (Cooper & Paccia-Cooper, 1980). A series of experiments demonstrated that word and pause durations were predicted more successfully by a hierarchical prosodic representation than a syntactic representation which was found to be neither necessary nor sufficient to account for the data. These data support the postulation of a level of prosodic structure intervening between the generation of syntax and phonology. However, duration studies are limited in what they can tell us about the processes underlying the *generation* of prosodic structure. If prosodic units are indeed constructed during speech production processes, then it must also be possible to demonstrate effects of this computation on speech production *latencies*. There are two findings in the literature that are at least suggestive of such effects. The first comes from the prepared speech production paradigm (Ferreira, 1991; Sternberg, Monsell, Knoll, & Wright; Sternberg, Wright, Knoll, & Monsell, 1978, 1980) and the second from on-line word production tasks investigating the syllable latency effect (Eriksen, Pollack, & Montague, 1970; Klapp, 1974; Klapp, Anderson, & Berrian, 1973). The experiments we will describe exploit both prepared and on-line speech production paradigms and we motivate our use of each of these paradigms below.

THE EXPERIMENTAL PARADIGMS

The prepared speech paradigm. Sternberg et al. (1978) provide data suggestive of an effect of prosodic structure on speech production latencies. They investigated the planning or motor programming of rapid movement sequences in speech. They asked subjects to produce prepared random lists of one to six letters or digits and examined the effect of the number of elements in a sequence on the temporal patterns of its production. There were two main findings. First, latency in beginning to speak was found to increase linearly with list length. Second, they found that the durations of these rapid utterances were related to the number of words they contained in a concave upward rather than a linear manner. In other words, on average, items in longer sequences were produced at slower rates.

One possible explanation is that part of the latency includes the time to retrieve information concerning the entire sequence. Sternberg et al. claimed that during the preparation interval subjects prepare an articulatory motor program for their utterance which specifies its elements and their order. On detection of the signal to respond, execution of this program is accomplished through a cycle of three processes. First, the program for the initial element of the sequence is retrieved, second, its contents are unpacked, and third, the appropriate articulatory commands are initiated. According to the model, the retrieval process (**R**) is sensitive to the number of items in the buffer (**n**) but not their properties; the more elements a program contains the longer it takes for any one element to be selected. Conversely, the duration of the unpacking process (**U**) is sensitive to the complexity of the unit to be unpacked. Thus, production latency (**L**) for a list is determined by the time it takes to retrieve and produce the first item in that list and therefore has a linear relationship to list length, $\mathbf{L} = \mathbf{Rn} + \mathbf{U}$. The production duration of a list is modeled as the sum of the production latencies for each item in the list. By claiming that the buffer is nondecreasing, **n** remains the same for the retrieval of each unit in the list. The duration effects can therefore be modeled by the quadratic function $\mathbf{Rn}^2 + \mathbf{Un}$ plus some constant for the intersect. Thus both latency and duration effects can be elegantly modeled by the same processes.

When interpreted in the light of this model further experiments of the Sternberg et al. provide clues about what the elements of retrieval and articulation might be. They are certainly not stored lexical representations as exactly the same pattern of effects is observed for nonwords. The units are not syllables as the slope of the function is the same for lists of matched bisyllabic words as for monosyllabic words (e.g., baby-rumble-market, bay-rum-mark), although a significant 4-ms increase in the intersection is observed which is attributed to the unpacking of the first unit. Most interestingly, the elements are also not syntactic words as the addition of unstressed words such as *and* (e.g., bay and rum and mark) did not alter the slope of the latency function. Sternberg et al. concluded that the unit of the buffer is the ''stress group'' or a unit of speech associated with a primary stress. This stress group is a prosodic unit that is built on syllables and feet and may therefore correspond to the phonological word.

The above findings suggest that prepared speech production latencies are sensitive to the prosodic structure of the utterance as a whole rather than to the number of lexical items to be produced. This task may, therefore, be used to determine whether prosodic structure is generated prior to articulation and which prosodic units are most salient when the sentence must be produced. However, the Sternberg et al. data cannot provide us with an answer to these questions. The fact that lists were used causes a number of problems.[2] Lists have no syntactic structure and a very flat prosodic structure. List intonation can consist of a series of concatenated intonational phrases or individual smaller phrases with a final phrase fall (Nespor and Vogel, 1986). Because lists have little prosodic structure it is impossible to tell what the smallest relevant unit might be. The Sternberg et al. stress group could be either a small or large linguistic con-

[2] Monsell (1986) reports a series of experiments comparing the production of lists and sentence materials (e.g., ''Barbara, Trixi, Arthur, Reuben, Dean'' and ''Barbara tricks a rather rueful Dean.''). The results for both types of materials are reported as being almost identical. However, the example sentence given is somewhat unusual and may also have elicited a list intonation prosodic structure.

stituent because each linguistic constituent can have its own major stress unit. Moreover, most of the results come from six highly trained subjects who received weeks of practice. We therefore do not know how the strategies that they developed relate to normal language production processes. Many of these problems were solved by Ferreira (1991) who elicited prepared sentences from a larger group of comparatively untrained subjects. She demonstrated that both the addition of a phonological word and greater syntactic complexity can increase production latencies in this task.

So how does prepared speech production relate to normal language production processes? In the prepared speech task an utterance must be constructed and held in memory for a period of time. This is actually not a rare state of affairs. During conversation, the rules of turn-taking may require that we hold onto a prepared utterance until it is possible to gain the floor. Moreover, it seems intuitively plausible that if the utterance is represented by a structured set of units then the number of units active in memory should in some way determine the time needed to prepare the first of those units for output. Ferreira (1991) argued that following the signal to respond, subjects translated their semantic/syntactic representation of an utterance into a phonological/phonetic one and that the more syntactic nodes a sentence contained the longer this process required. Our aim, however, was to demonstrate an effect of phonological structure on sentence production when syntactic complexity and number of lexical items was held constant. Ferreira's sentences were long (approx 8–14 syntactic words) so it is unlikely that subjects could hold a phonological representation of the whole utterance in short term memory (STM). In our experiments the sentences to be produced were matched for syntactic complexity and comprised no more than four lexical items. Sentences of this length could easily be held in phonological STM. In this situation, we would suggest that, prior to the signal to respond, subjects have constructed a complete surface syntactic structure and generated the phonological representation for the entire utterance. Following the signal to respond, subjects must retrieve and articulate the first output unit of this representation. We hoped to use this task to examine effects of phonological structure in isolation.

On-line sentence production. What the prepared speech production paradigm cannot tell us is how prosodic structure, if generated, affects on-line sentence production processes. An incremental model of sentence production makes clear predictions. Following Kempen and Hoenkamp (1987), Levelt (1989) proposes that processing at all levels occurs in an incremental fashion with a processor being triggered by any piece of characteristic input from the processors that feed into it. Thus, even though some processing must have occurred at a particular level before processing at the next can begin, processing at all levels can run in parallel but on different pieces of the utterance to be produced. Such a system requires that processing can occur from left to right in an utterance with minimal look ahead. Therefore, what a processor is doing with a particular fragment of an utterance should not be dependent on information available in later fragments of the utterance. For example, constructing the initial prosodic units of a sentence should not be dependent on how the sentence will end.

According to Levelt, the phonological word is the minimum unit of articulation, therefore, all other things being equal, sentence production latencies will be determined by the time required to generate the first phonological word of a sentence. In support of this claim Levelt (1989) cites the ''syllable latency effect,'' which refers to the finding that the time taken to initiate production of a visually presented word increases with the number of syllables it contains (Eriksen et al., 1970). This effect has been replicated using a digit reading task (e.g., 27 took longer than 26 (Klapp, 1974)) and in picture naming tasks (Klapp et al., 1973), suggesting that the effect is located in production rather than perceptual processes. As discussed above, according to Levelt (1989) the articulator waits for a whole phonological word before executing the first syllable's motor program.

However, the experiments reviewed above do not allow us to distinguish between phonological and lexical words as the minimal unit required before articulation may commence. Moreover, some theorists have argued that some aspects of prosodic structure cannot be produced incrementally but require more advanced planning. For example, there is evidence that rate of declination is dependent on utterance length (Cooper & Sorensen, 1981; De Pijper, 1983). We will return to these issues in the final experiment we report.

The present research. The experiments described below were designed to find evidence for the generation of phonological words during sentence production in Dutch. Experiments 1 to 3 exploit the prepared speech production paradigm. The results confirm that the phonological word is the unit that governs production latencies in the prepared speech paradigm. In contrast, the experimental methodology used in the final experiment tests online sentence production and provides evidence that the phonological word is the preferred unit of output during sentence production. Taken together, these experiments provide evidence that during the production of connected speech, discrete representations for words are retrieved and transformed on-line into prosodic units which form the interface between grammatical encoding and articulation.

EXPERIMENT 1

The aim of this experiment was to find an effect on prepared sentence production of the number of phonological words the sentence comprises when number of syllables, lexical words, and syntactic structure are held constant. In contrast to the procedure of Sternberg et al. (1978, 1980) a large number of relatively untrained subjects was tested and a more natural question–answer task was used to elicit the experimental sentences. In contrast to Ferreira (1991), we increased the number of phonological words in an utterance without adding extra lexical items.

As we discussed in the introduction, a phonological word for Dutch can be defined minimally as a stressed foot, to which unstressed syllables can be adjoined leftward (Booij & Lieber, 1993; Lahiri et al., 1990; Gussenhoven, 1989; Berendsen, 1986). Similar to English, the phonological word forms the domain of syllabification in Dutch. For example, consider the two example sentences below where phonological word structure is given by the brackets and syllables are separated by hyphens.

5a. [Ik-heb-een]$_\omega$ [laars]$_\omega$ [aan]$_\omega$
I have a boot on

5b. [Ik-heb-een]$_\omega$ [laar-ste]$_\omega$ [koop]$_\omega$
I have a boot to sell

In sentence (5a) the final word *aan* (on), comprises a heavy syllable which attracts stress and forms its own phonological word. In sentence (5b), however the adverb *te* (to) is usually destressed and cliticizes to the preceding noun *laars* (boot) to form a single phonological word (Lahiri et al., 1990). Within this phonological word, the final /s/ of *laars* resyllabifies to form the onset of the second syllable /ste/ (following the Maximal onset principle, Selkirk, 1984). In sentence (5a), however, the phonological word boundary intervening between *laars* and *aan* prevents a similar resyllabification to *laar–saan.*

Examples of the sentences produced in this experiment are given in Fig. 1. The nonclitic and clitic sentence types are matched for surface syntactic structure (given by phrase marker), number of lexical words, and number of syllables. They differ, however, in their number of phonological words (given by brackets). In the clitic sentences the words *het*, *de,* and *te* cliticize leftward to the verb becoming a single prosodic word which cuts across syntactic and phrasal structure. In contrast, in the nonclitic sentences, *Jans, vers,* and *heel* attract stress and form independent phonological words.

Predictions

If the latency effect is a function of the number of phonological words, then the latency in producing the clitic sentences should be shorter than the latency in producing the nonclitic sentences. Note, however, that the initial phonological word in the clitic sentences, e.g., [Ik zoek het]ω, has one more syllable than the initial phonological

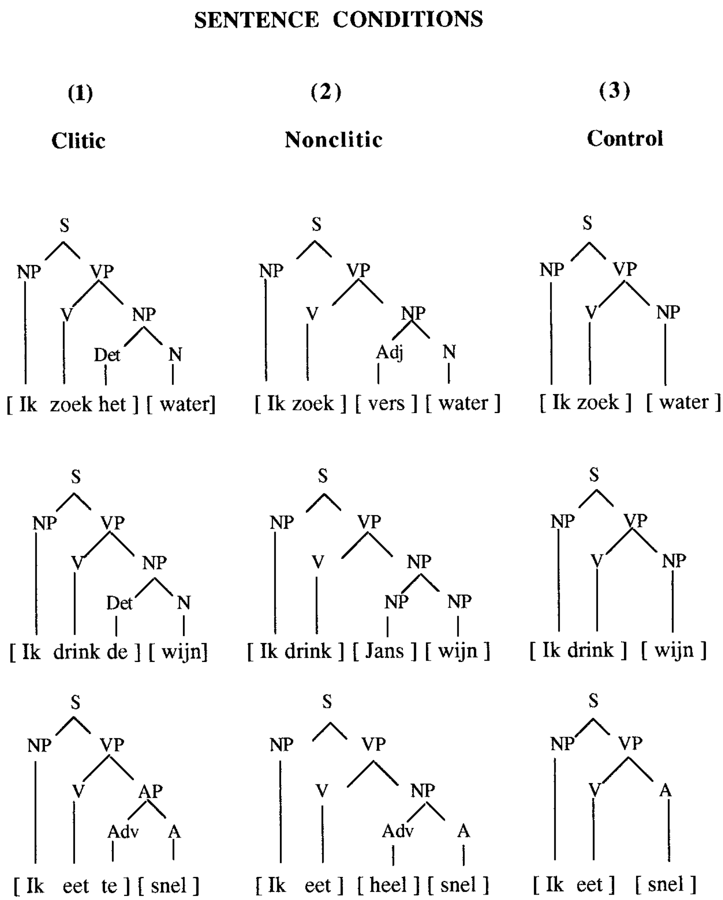

FIG. 1. The syntactic and prosodic word structure of the sentence types used in Experiments 1 and 4. Syntactic structure is given by the phrase marker above the sentences and the constituent prosodic words of each sentence are in brackets.

word in the nonclitic sentences, e.g., [Ik zoek]ω. It is possible therefore that any difference in latency due to number of phonological words may be reduced by an opposite effect due to the complexity of the initial phonological word to be retrieved (Sternberg et al., 1978). We therefore also included control sentences which are matched to the clitic sentences for number of phonological words but, like the nonclitic sentences have only two syllables in the initial phonological word. Any effect of the length of the initial phonological word should be observed in a latency difference between the clitic and control sentences.

Method

Sentences in all of the experiments we report were elicited from subjects using a question–answer procedure. In this procedure, subjects first saw a noun phrase or adjective phrase and then heard a question referring to that phrase. Their task was to construct a sentence in answer to the question using the words they had seen. Examples of the stimuli used to elicit the experimental sentences in Experiment 1 are given in Table 1.

Materials. The experimental materials consisted of 12 monosyllabic verbs (see Appendix 1 for full listing). Each verb was associated with either a noun or an adjective.[3] In the

[3] The nouns and adjectives were chosen such that they comprised only one stressed foot. This means that the number of feet varies with the number of phonological words across conditions. However, a similar experiment (to be reported in Lahiri and Wheeldon, in preparation)

TABLE 1

EXAMPLES OF THE EXPERIMENTAL SENTENCES IN THE THREE CONDITIONS OF EXPERIMENT 1

Visual display	Auditory question	Response
	Clitic sentences	
het water (*the water*)	Wat zoek je? (*what do you seek?*)	Ik zoek het water (*I seek the water*)
	Nonclitic sentences	
vers water (*fresh water*)	Wat zoek je? (*what do you seek?*)	Ik zoek vers water (*I seek fresh water*)
	Control sentences	
water (*water*)	Wat zoek je? (*what do you seek?*)	Ik zoek water (*I seek water*)

Note. The expected response is preceded by the visual display and auditory question used to elicit it.

clitic sentences the nouns occurred with their definite article. Dutch has two genders which are marked by the definite article: four of the nouns used were *de* words (e.g., *de wijn,* the wine) and four were *het* words (e.g., *het water,* the food). The four adjectives occurred with the adverb *te* (e.g., *te snel,* too fast). In the experimental sentences, *de, het,* and *te* all cliticize to the preceding verb to form one phonological word. In the nonclitic condition *de*-nouns occurred with *Jans* (John's), *het*-nouns with *vers* (fresh), and adjectives with *heel* (very), all of which attract stress and are produced as separate phonological words. Finally, in the control condition nouns and adjectives occurred in isolation.

Design. Each of the 12 verbs occurred in the three sentence types, resulting in 36 possible sentences: 12 in each condition. After receiving instructions and completing a practice set of sentences, each subject produced the 36 experimental sentences four times each. The experiment consisted of six blocks of 24 trials. Each sentence occurred twice within the first three blocks and twice within the second three blocks of the experiment. No sentence occurred twice within the same block and each block contained equal numbers of sentences from each condition. The order of presentation of the six blocks was rotated across subjects.

Apparatus. The Dutch questions were presented using a Sony DTC-1000 ES DAT-recorder. Subjects' responses were recorded by a Sony DTC-55 ES DAT-recorder. An analog voice-key registered voice onset and offset times during sentence production. The experiment was controlled by a Hermac PC.

Procedure. Subjects were tested individually in a sound-proof booth. They were seated in front of a window through which they could see a computer screen and wore headphones through which they heard the experimental questions. Before beginning the experimental blocks, subjects received instructions and completed a set of practice trials. Subjects were told that they would see words on the screen and then hear a question which referred to the words they had read. Their task was to prepare a full sentence response to the question using the words they had seen. They were told that they would have approximately 4 s to prepare their response, followed by a signal to respond. They were asked to prepare their responses as fully as possible and to produce

tested the delayed production of compounds (e.g., ooglid, *eyelid*) which comprised two feet and morphologically simple words (e.g., orgel, *organ*) which comprised one foot. No significant difference in production latencies was observed, suggesting that the number of feet is irrelevant in this task.

their sentence as quickly as possible after hearing the response signal. They were also asked to speak naturally, putting stress on the last word of their response. All subjects then completed six practice trials during which they first saw a practice trial and heard a recorded example response. They completed the same trial immediately after. Subjects were allowed short breaks between blocks.

Events on each experimental trial were as follows. A fixation cross appeared centered on the screen for 500 ms. Five hundred milliseconds after the offset of the fixation cross a two word phrase appeared centered on the screen for 500 ms. Following another 500-ms pause subjects heard a short question. This question was followed by a series of three beeps; the first occurring 2 s after the offset of the question and the second occurring 1 s later. In order to prevent subjects anticipating the third and last beep, it occurred at one of four possible latencies measured from the offset of the second beep: 750, 1000, 1250, or 1500 ms. Each verb in each condition occurred once at each of the four latencies. There was a two second interval between trials. Subjects' response latencies were measured from the onset of the third beep to their voice onset using a voice key. The total duration of their utterances was also measured and subjects' responses were recorded. An experimental session lasted approximately 1 h.

It was important to ensure that subjects actually produced the sentences with the intended prosodic structure. In particular, we needed to know that sentence stress was correctly placed (in this case on the final word). However, stress has no single physical correlate. It can be realized by either an increase in the duration or amplitude of a syllable or a change in pitch or in any combination of the three. Thus it is impossible to provide any reliable acoustic measure of degree of stress for our stimuli. However, since stress is a perceptual variable, during the experiment the experimenter listened to each production of a sentence to check that it was produced with the stress and intonation pattern required. This was a reasonably simple task for our sentences. However, if the experimenter was uncertain about any particular production, the recorded version of the sentence was double checked at the end of the experiment.

Subjects. Eighteen subjects were tested. They were all native Dutch speakers who were members of the Max Planck subject pool. They were paid for their participation.

Results

Data preparation. The analyses we report are based on data from correct response trials, following some exclusions intended to reduce the noise in the data. All data points beyond two standard deviations from the mean were counted as outliers and were removed. Incorrect responses were also removed from the latency data. This resulted in the loss of only 3.2% of the data. A response was marked as an error when the subject produced a sentence that differed from the intended sentence in either lexical content or syntactic structure or when the subject produced the intended sentence with any disfluency. Correct responses which were produced before the final beep were also excluded.

Responses were also marked as an error when subjects' productions deviated from the intended prosodic structure. In the most common deviation, subjects assigned sentence stress to a nonfinal word. This occurred rarely but most often in the nonclitic sentences where the penultimate word may also receive sentence stress. It was important to remove any such responses to allow the strongest comparison between the nonclitic sentences and the clitic sentences in which the penultimate word cannot attract sentence stress.

Missing values were substituted by a weighted mean based on subject and item statistics calculated following Winer (1971, pp. 488). Separate analyses were conducted with means calculated by averaging over subjects (F_1) and over items (F_2). Mean production latencies and percentage error rate in each condition are given in Table 2 as a function of preparation latency. Latencies in the nonclitic condition were 14 ms longer than in both the clitic and control conditions which do not differ. This difference, though small, was very reliable. An ANOVA was performed on nam-

TABLE 2

MEAN PRODUCTION LATENCIES IN MS AND PERCENTAGE ERROR RATES FOR THE THREE SENTENCE TYPES OF EXPERIMENT 1 AS A FUNCTION OF PREPARATION TIME

Latency (% error)	Clitic	Nonclitic	Control	Mean
Phonological words	2	3	2	
Lexical words	4	4	3	
Syllables	5	5	4	
Preparation time				
750	400 (1.4)	425 (3.7)	400 (2.8)	408 (2.6)
1000	381 (5.6)	395 (3.2)	379 (3.2)	385 (4.0)
1250	376 (1.4)	380 (2.8)	368 (1.9)	375 (2.0)
1500	363 (0.0)	375 (1.9)	372 (3.2)	370 (1.7)
Mean	380 (2.1)	394 (2.9)	380 (2.8)	

Note. The number of phonological words, lexical words, and syllables in each sentence type are also shown.

ing latencies including the variables sentence type (1–3) and preparation time (PT) (1–4). The main effect of sentence type was significant, $F_1(2,30) = 14.4$, $p < .001$, $F_2(2,22) = 5.6$, $p < .01$. The main effect of preparation time was also significant by subjects and marginally significant by items, $F_1(3,45) = 24.9$, $p < .001$, $F_2(3,33) = 2.8$, $p = .054$. Newman–Keuls pairwise comparisons showed that mean production latency at PT750 was significantly slower than at PT1250 and PT1500 ($p < .05$). The differences between the nonclitic sentences and the other conditions tended to be larger at the shorter preparation times, however, the interaction of sentence type and preparation time was not significant, $F_1(6,90) = 2.1$, $F_2 < 1$.

During the course of the experiment each subject produced each sentence four times. It is possible, therefore, that practice or repetition may have influenced performance as the experiment progressed. In order to test for effects of practice or repetition, an ANOVA was conducted which included the variable experiment half. A main effect of experiment half was observed in the subjects analysis, $F_1(1,17) = 3.7$, $p < .001$, $F_2 < .01$, due to a 53-ms decrease in production latencies in the second half of the experiment. Importantly, however, the pattern of results was very similar in both halves of the experiment and there was no interaction of experiment half and sentence type, $F_1(2,34) = 1.5$, $F_2(2,10) = 1.9$. No other interactions with this variable were significant. Percentage error rates were small and a similar ANOVA on the error data yielded no significant effects.

Utterance durations are given in Table 3 and show a quite different pattern of results. Not surprisingly, the nonclitic condition has the longest duration but importantly the clitic and control conditions also differ in duration by 91 ms. The main effect of sentence type was again significant, $F_1(2,30) = 516.3$, $p < .001$, $F_2(2,22) = 229.6$, $p < .001$. Newman–Keuls pairwise comparisons showed all means to differ significantly from each other ($p < .001$ by subjects and by items).

TABLE 3

MEAN PRODUCTION DURATIONS FOR THE THREE SENTENCE TYPES OF EXPERIMENT 1 AS A FUNCTION OF PREPARATION TIME

Duration	Clitic	Nonclitic	Control	Mean
Phonological words	2	3	2	
Lexical words	4	4	3	
Syllables	5	5	4	
Preparation time				
750	648	754	561	654
1000	648	771	562	660
1250	659	768	567	665
1500	667	763	571	667
Mean	656	764	565	

Note. The number of phonological words, lexical words, and syllables in each sentence type are also shown.

The effect of preparation time was significant by subjects only, $F_1(3,45) = 4.1$, $p < .05$, $F_2 < 1$. This was due to significantly longer mean naming durations at PT1500 and PT1250 than at PT750 ($p < .05$). Subjects, therefore, showed a slight tendency to speak more slowly at the longer preparation times. The interaction of preparation time and sentence type was nonsignificant, $F_1(6,90) = 1.3$, $F_2 < 1$.

Utterance durations decreased by 45 ms in the second half of the experiment but the main effect of experiment half was only significant in the subject analysis, $F_1(1,17) = 12.9$, $p < .01$, $F_2 < 1$. The subject analysis also yielded significant interaction of experiment half with sentence type, $F_1(2,34) = 8.1$, $p < .001$, $F_2(2,10) = 1.5$. Examination of the cell means showed that this was due to small differences in the size rather than in the direction of the effects across experiment half. The interaction of experiment half with preparation time was also significant by subjects, $F_1(3,51) = 5.1$, $p < .01$, $F_2 < 1$. In the first half of the experiment, durations were longest at PT1500. In the second half of the experiment, durations were longest at PT1250.

Discussion

The paradigm was successful in eliciting significant results from a large number of relatively untrained subjects. All subjects could easily produce the correct sentences and made very few stress errors. The experiment yielded significantly longer production latencies for sentences comprising three phonological words than for sentences comprising two phonological words, when those sentences were matched for syntactic structure, number of lexical words, and number of syllables. This result provides strong support for the phonological word as the output unit in the prepared speech production task. Moreover, the 14-ms effect is similar in size to the slope of the Sternberg et al. (1978) function and it seems probable that we are tapping into the same process that underlies their results.

In contrast, production latencies in the two conditions where sentences comprised two phonological words do not differ despite the fact that the control sentences (e.g., *Ik zoek water*) are simpler than the clitic sentences (e.g., *Ik zoek het water*) in a number of ways: they have a simpler syntactic structure, fewer lexical words, and a shorter initial phonological word. Thus, unlike Sternberg et al. (1978) we have no evidence of an effect on naming latency of the complexity of the first phonological word. However, as we have mentioned above, Sternberg et al. used a small group of highly trained subjects and it is possible that our methodology lacks the sensitivity to detect such small articulatory unpacking effects.

The observed latency results cannot be explained in terms of whole utterance duration, i.e., that sequences of longer duration take longer to initiate, as utterance durations showed a very different pattern of results. Crucially, there is a large and significant difference in utterance duration between the clitic and control conditions despite identical latency results.

Our effects were also robust with respect to practice and repetition. This suggests that the preparation time was sufficient to allow subjects to reach a fully prepared state and that subjects did not build task specific strategies as the experiment progressed.

These results are consistent with the claim that the phonological word is the prosodic unit that determines production latencies in the prepared speech production task. This finding supports Levelt's (1989) claim that the phonological word is a unit of phonological encoding.

However, an alternative explanation is that our sentence production latencies are determined by the number of content words our sentences contained. Content words are major syntactic class items (e.g., nouns, verbs, and adjectives) which are often referred to as open class items because they readily accept new members. Open class items carry most of the semantic information in a sentence and usually have stress. In contrast, function words (e.g., prepositions, pronouns, determiners and conjunctions) are referred to as closed class items because they have a fixed membership. Closed class items usually carry information relevant to the syntactic roles of the content words and do not attract stress.

Closed and open class items can exhibit different behavior in language production. Errors involving open class items occur more frequently in speech error corpora (Garrett, 1990), whereas closed class items seem to be more readily lost in aphasic speech (Saffran, Schwartz, & Martin, 1980). Garrett (1982) has accounted for these differences by suggesting that open and closed class items form functionally different vocabularies. He claims that closed class items should be seen as features of the syntactic frame generated during positional encoding, whereas open class items must undergo a process of association to the frame. It is possible therefore, that the prepared speech production task is tapping the process of assigning open class words to their position in a syntactic frame rather than the retrieval of the initial unit of a prosodic structure. The next experiment was designed to test this alternative explanation of the results of Experiment 1.

EXPERIMENT 2

The aim of Experiment 2 was to test whether sentence production latencies are determined by the number of phonological words or the number of content words a sentence contains. Materials were constructed which allowed a comparison between sentences comprising the same number of phonological words but different numbers of lexical words as well as a comparison between sentences comprising the same number of lexical words but different numbers of phonological words. As in Experiment 1 a question-answer technique was used to elicit sentences from subjects. Examples of the stimuli used to elicit experimental sentences in Experiment 2 are given in Table 4.

As in Experiment 1, all sentences were produced with main stress on the final word of the sentence. The sentences produced in the Clitic and the Nonclitic conditions were identical in structure to those produced in Experiment 1. In this experiment, however, two new sentence conditions were constructed. In the pronoun sentences the noun phrase consisted of the pronoun *het* (it). This pronoun is phonologically identical to the neutral Dutch article. However, in the pronoun sentences, *het,* is phrase final and receives stress thereby becoming a phonological word in its own right. The pronoun sentences thus comprise the same number of phonological words as the clitic sentences but have a different number of content words (e.g., *zoek*). In contrast, the control sentences, like the pronoun sentences, have only one content word but these sentences differ in the number of phonological words that they comprise. The question of interest is therefore, whether the latencies to pronoun sentences are similar to latencies to the clitic or to the control sentences.

Method

Vocabulary. The experimental materials consisted of nine of the monosyllabic verbs used in Experiment 1 (Appendix 2). Each of these verbs occurred in the four different sentence types elicited by the questions shown in Table 4 resulting in a total of 36 sentences, nine in each condition. Fewer verbs were included in order to keep the length of the experimental sessions under one hour.

Design and procedure. Each subject produced the 36 experimental sentences four times. The experiment consisted of eight blocks of trials. Each sentence occurred once in every two block set and the presentation order of the blocks was rotated across subjects.

Events on each trial were the same as in Experiment 1. Subjects' response latencies and durations were measured and their responses were again recorded onto tape. Subjects received the same instructions as in Experiment 1. They were again asked to speak naturally and to place stress on the last word of the sentence. The stress and intonation of each response was again checked by the Experimenter. They first saw a practice trial and heard a recorded example response. They completed the same trial immediately after. The procedure during the rest of the experiment was the same followed in Experiment 1. Twenty subjects from the Max Planck subject pool were tested.

Results

Data preparation. Data were excluded from the analysis following the same procedure as

TABLE 4

EXAMPLES OF THE EXPERIMENTAL SENTENCES IN THE FOUR CONDITIONS OF EXPERIMENT 2

Visual display	Auditory question	Response
	Clitic sentences	
het water (*the water*)	Wat zoek je? (*what do you seek?*)	[Ik zoek het] [water] (*I seek the water*)
	Nonclitic sentences	
vers water (*fresh water*)	Wat zoek je? (*what do you seek?*)	[Ik zoek] [vers] [water] (*I seek fresh water*)
	Pronoun sentences	
het (*it*)	Wat zoek je? (*what do you seek?*)	[Ik zoek] [het] (*I seek it*)
	Control sentences	
zoek (*seek*)	Wat doe je? (*what do you do?*)	[Ik zoek] (*I seek*)

Note. The expected response is preceded by the visual display and auditory question used to elicit it.

in Experiment 1. Data trimming resulted in the loss of 2.8% of the data. Missing values were again substituted by a weighted mean based on subject and item statistics. Mean production latencies and percentage error rate in each condition are given in Table 5 as a function of preparation time.

Analysis of variance yielded a significant main effect of sentence type, $F_1(3,57) = 13.6$, $p < .001$, $F_2(3,24) = 18.8$, $p < .001$. Newman–Keuls pairwise comparisons yielded a number of significant differences. As in Experiment 1, latencies for the nonclitic sentences were significantly slower (by 14 ms) than latencies for the clitic sentences ($p < .01$ by subjects, $p < .05$ by items) and the pronoun sentences ($p < .01$ by subjects, $p < .05$ by items). Latencies in the clitic and pronoun sentences did not differ. Production latencies for the control sentences were significantly faster

TABLE 5

MEAN PRODUCTION LATENCIES IN MS AND PERCENTAGE ERROR RATES FOR THE THREE SENTENCE TYPES OF EXPERIMENT 2 AS A FUNCTION OF PREPARATION TIME

Latency (% error)	Clitic	Nonclitic	Pronoun	Control	Mean
Phonological words	2	3	2	1	
Content words	2	3	1	1	
Syllables	5	5	3	2	
Preparation time					
750	418 (1.7)	431 (10.1)	426 (1.7)	391 (1.7)	417 (3.8)
1000	405 (1.1)	406 (4.0)	385 (1.7)	375 (3.3)	392 (2.5)
1250	378 (1.1)	401 (2.8)	390 (0.6)	371 (0.6)	385 (1.3)
1500	385 (1.7)	400 (4.4)	385 (3.3)	369 (1.1)	395 (2.6)
Mean	396 (1.4)	410 (5.3)	396 (1.8)	377 (1.7)	

Note. The number of phonological words, content words, and syllables in each sentence type are also shown.

than for all other conditions ($p < .01$ in both subject and item analyses). There was also a significant effect of preparation time, $F_1(3,57) = 13.8$, $p < .001$, $F_2(3,24) = 21.6$, $p < .001$. Similar to Experiment 1, Newman–Keuls pairwise comparisons showed that mean production latency at PT750 was significantly slower than at PT1250 and PT1500 ($p < .05$ by subjects and by items). The interaction of preparation time and sentence type was nonsignificant, $F_1(9,171) = 1.9$, $F_2 < 1$. In order to test for effects of practice or repetition, an ANOVA was conducted which included the variable experiment half. No significant effects involving this variable were observed.

The analysis of percentage error rates also yielded a significant effect of sentence type, $F_1(3,57) = 6.6$, $p < .01$, $F_2(3,24) = 10.3$, $p < .001$. The error rate in the nonclitic condition was significantly higher than in all other conditions ($p < .05$ by subjects and by items) which did not differ. This was due to an increased tendency in these subjects to destress the final word in these sentences. Despite a tendency for error rates to decrease as preparation time increased, the main effect of preparation time was not significant, $F_1(3,57) = 2.0$, $F_2(3,24) = 2.2$. There was, however, a significant interaction of sentence type and preparation time in the items analysis, $F_1(9,171) = 1.8$, $p > .05$, $F_2(9,72) = 2.3$, $p < .05$. This was due to a decrease in the difference between nonclitic sentences and the other sentences as preparation time increased. Not surprisingly, subjects made fewer stress errors on the nonclitic sentences at the longer preparation times. The analysis including the variable experiment half again yielded no significant effects.

Sentence durations are given in Table 6. There was a highly significant main effect of sentence type, $F_1(3,57) = 450$, $p < .001$, $F_2(3,24) = 636$, $p < .001$. Newman–Keuls pairwise comparisons showed that all conditions differed significantly from each other ($p < .001$ by subjects and by items). There was also a significant main effect of preparation time, $F_1(3,57) = 5.8$, $p < .01$, $F_2(3,24) = 3.9$, $p < .05$. This was due to a significantly longer mean naming duration at PT750 than at PT1250 ($p < .05$ in subjects and item analyses). In contrast to Experiment 1, subjects here showed a slight tendency to speak more slowly at the *shortest* preparation time. This difference most probably reflects the increased complexity of this experiment. The interaction of sentence type and preparation time was not significant, F_1 & $F_2 < 1$.

Utterance durations yielded a significant interaction of experiment half with preparation time, $F_1(3,57) = 4.1$, $p < .05$, $F_2(3,9) = 5.0$, $p < .05$. Similar to Experiment 1, in the first half of the experiment, durations were longest at PT1500. In the second half of the experiment, durations were longest at PT1000.

Discussion

This experiment replicated the effect observed in Experiment 1 for the first two conditions. Once again latencies for the nonclitic sentences were significantly longer (14 ms) than latencies for the clitic sentences. Importantly, however, clitic sentences like, *Ik zoek het water,* yielded identical naming latencies to pronoun sentences like, *Ik zoek het,* despite differences in the number of content words they contain. Both of these sentence types comprised two phonological words. Latencies for both the clitic and the pronoun sentences were significantly longer (20 ms) than for the control sentences like, *Ik zoek,* which comprised one phonological and one content word. Clearly, production latencies are a function of prosodic structure rather than a function of the number of content words a sentence contains.

It is still possible, however, that our effect is due not to the retrieval of an abstract prosodic representation of the utterance but to the generation of a concrete phonetic representation prior to articulation. Phonetic encoding involves (among other things) the assignment of absolute stress levels to the syllables to be produced (Levelt, 1989; Levelt & Wheeldon, 1994). In Experiments 1 and 2, the clitic and nonclitic sentences were produced with a declarative intonation pattern in which primary stress is assigned, by default, to the accented syllable of the last word. However, in the nonclitic condition, the extra nonclitic syllable

TABLE 6

MEAN PRODUCTION DURATIONS IN MS FOR THE THREE SENTENCE TYPES OF EXPERIMENT 2 AS A FUNCTION OF PREPARATION TIME

Duration	Clitic	Nonclitic	Pronoun	Control	Mean
Phonological words	2	3	2	1	
Content words	2	3	1	1	
Syllables	5	5	3	2	
Preparation time					
750	808	902	436	343	622
1000	789	868	424	340	605
1250	765	867	411	339	595
1500	779	863	420	341	601
Mean	785	875	423	341	

Note. The number of phonological words, content words, and syllables in each sentence type are also shown.

(e.g., *Jans, heel, vers*) also attracts a degree of stress which would be calculated and assigned during the phonetic encoding of the utterance. Although phonological word formation is dependent on whether syllables can attract or lose stress, it should be independent of the absolute stress levels associated with syllables, which can change depending on where the primary stress falls. The aim of the next experiment was to test whether the results of Experiments 1 and 2 would generalize to sentences produced with a different primary stress and intonation pattern.

EXPERIMENT 3

Experiment 3 tested the production of sentences similar to those produced in Experiment 1. The materials used were essentially the same as Experiment 1 except that primary stress was placed on the first word rather than on the last word of the sentence. Examples of the experimental stimuli are given in Table 7. The assignment of primary stress to the first word of a sentence changes the stress levels assigned to the following words (Liberman & Prince, 1977). The crucial difference for our purposes is that the absolute stress differences between the three sentences become muted.

The assignment of primary stress to an utterance is not necessarily determined by structural information alone. Instead, it may be determined by semantic factors such as given versus new information and focus, which can depend on information available in previous utterances as well as aspects of the conversational situation such as the shared knowledge of the speaker and hearer. The sentences in Experiment 3 were produced with focus intonation on the first phonological word. In order to focus the first word of the sentence a proper name was displayed to the subjects (e.g., *Riet, Henk*, or *Bert*) and the question provided the rest of the information necessary to construct the sentence. As the proper name is the new information in the sentence to be produced it receives primary sentence stress. The prosodic structure of the response sentences is shown in (6) below.

6a. $[[[\text{Riet}]_\omega]_\varphi\ [[\text{zoekt het}]_\omega\ [\text{water}]_\omega]_\varphi]_{IP}$
6b. $[[[\text{Riet}]_\omega]_\varphi\ [[\text{zoekt}]_\omega\ [\text{vers}]_\omega\ [\text{water}]_\omega]_\varphi]_{IP}$
6c. $[[[\text{Riet}]_\omega]_\varphi\ [[\text{zoekt}]_\omega\ [\text{water}]_\omega]_\varphi]_{IP}$

Each sentence type now has an additional phonological word but, as in Experiments 1 and 2, sentence (6b) comprises one more phonological word than sentences (6a) and (6c). Moreover, each sentence is produced with a downward intonation contour with the main sentence stress on the first phonological word (i.e., *Riet*). This has the effect of changing the stress levels assigned to the following words such that the absolute stress differences between the three sentences become muted. If the effect observed in Experiment 1 is due to absolute difference in stressed syllables then

TABLE 7

EXAMPLES OF THE EXPERIMENTAL SENTENCES IN THE FOUR CONITIONS OF EXPERIMENT 3

Visual display	Auditory question	Response
	Clitic sentences	
Riet (*Riet*)	Wie zoekt het water (*Who seeks the water?*)	Riet zoekt het water (*Riet seeks the water*)
	Nonclitic sentences	
Riet (*Riet*)	Wie zoekt vers water (*Who seeks fresh water?*)	Riet zoekt vers water (*Riet seeks fresh water*)
	Control sentences	
Riet (*Riet*)	Wie zoekt water (*Who seeks water?*)	Riet zoekt water (*Riet seeks water*)

Note. The expected response is preceded by the visual display and auditory question used to elicit it.

the effect should be reduced in sentences like those in (6) above.

Method

Design. As in Experiment 1, the 12 verbs were produced in the three different phrase conditions. These sentences were produced with three different monosyllabic Dutch proper names (Reit, Joop, and Henk) and at three different preparation latencies.

Each subject produced the 36 experimental sentences three times each. Three sets of 36 trials were constructed such that each experimental sentence occurred once only. Within a set each of the three sentence types for each verb occurred with a different name and warning period. Assignment of names and warning periods were rotated across the three sets so that each sentence occurred once with each name and each warning period.

Procedure. Each block of 36 was divided into two blocks of 18. The order of presentation of the three pairblocks was rotated across subjects. Each subject thus received 6 blocks of 18 trials and six subjects were assigned to each of the three rotations. As in Experiment 1, subjects received instructions and a practice set of sentences before the experiment proper began. Events on each trial were the same as in Experiments 1 and 2 except that this time the third and last beep occurred at one of three possible latencies from the offset of the second beep: 800, 1100, and 1300 ms. The Experimenter again checked each response for deviations from the desired stress and intonation pattern. As in Experiment 1, 18 native Dutch speakers were tested. They were members of the Max Planck subject pool and were paid for their participation.

Results

Data were excluded and substituted according to the same criteria used in Experiments 1 and 2. This resulted in the loss of 6.6% of the data. Mean naming latencies and percent errors are given in Table 8. Reaction times in this experiment were somewhat faster than in the previous two experiments despite the increase in the length of the sentences in terms of number of phonological words. This can most likely be attributed to sentence initial word stress resulting in earlier triggering of the voice key.

Once again, the main effect of sentence type was significant $F_1(2,34) = 23.7$, $p < .001$, $F_2(2,22) = 7.2$, $p < .01$. As in Experiment 1, latencies in the nonclitic condition are longer than in both the clitic (16 ms) and the control conditions (22 ms). Newman–Keuls pairwise comparisons showed both of these differences to

TABLE 8

MEAN PRODUCTION LATENCIES IN MS AND PERCENTAGE ERROR RATES FOR THE THREE SENTENCE TYPES OF EXPERIMENT 3 AS A FUNCTION OF PREPARATION TIME

Latency (% error)	Clitic	Nonclitic	Control	Mean
Phonological words	3	4	3	
Lexical words	4	4	3	
Syllables	5	5	4	
Preparation time				
800	380 (1.9)	411 (11.8)	369 (3.4)	387 (5.7)
1100	366 (5.7)	368 (5.3)	362 (4.3)	365 (5.1)
1300	370 (4.8)	386 (12.1)	367 (4.8)	374 (7.2)
Mean	372 (4.1)	388 (9.7)	366 (4.2)	

Note. The number of phonological words, lexical words, and syllables in each sentence type are also shown.

be significant ($p < .01$ by subjects and by items). The clitic and control conditions differed by a nonsignificant 6 ms. There was again a main effect of preparation time, $F_1(2,34) = 5.0$, $p < .05$, $F_2(2,22) = 10.6$, $p < .01$. This experiment yielded a significantly faster mean production latency at PT1100 than at PT800 ($p < .05$ by subjects and items). The interaction of preparation time and sentence type was also significant, $F_1(4,84) = 6.3$, $p < .001$, $F_2(4,44) = 2.5$, $p = .053$. The effect of sentence type was greatly reduced at the PT1100 compared to the shorter and longer preparation times. The most likely explanation for this finding is that the reduced number of different preparation times allowed subjects to try to anticipate the final signal to respond. The preparation time of 1100 ms would be the easiest to anticipate as it is closest to the rhythm of the preceding warning beeps. This hypothesis is supported by the finding that naming latencies were fastest at PT1100.

As in Experiments 1 and 2 an analysis was conducted to test for effects of repetition or practice on the main effects observed. In this experiment subjects produced each sentence three times, once in each pairblock of the experiment. An ANOVA was therefore conducted including the variable pairblock with three levels (first, second, and third). The pattern of results was similar across the three pairblocks and this analysis yielded no significant effects.

Percentage error rates also yielded a significant effect of sentence type, $F_1(2,34) = 10.9$, $p < .001$, $F_2(2,22) = 10.9$, $p < .001$, once again due to the higher error rate in the nonclitic condition due mainly to errors in stress. Some subjects still had a tendency to give some stress to the second proper name in the sentences ‘‘*Riet* drinkt *Jans* wijn.’’ Latencies from such trials were removed from the analysis. The effect of preparation time was nonsignificant, $F_1(2,34) = 2.2$, $F_2 < 1$. The interaction of preparation time and sentence type was significant in the by subjects analysis, $F_1(2,34) = 698.0$, $p < .001$, $F_2(2,22) = 121$, $p < .001$, due to the fact that the increase in error rate in the nonclitic condition was not observed at PT1100 ms. Percentage error rates yielded no significant main effect of pairblock and no significant interactions with this variable.

The pattern of results for the durations was also similar to that for Experiment 1 (see Table 9). The nonclitic condition had the longest duration. Durations for the clitic sentences were 81 ms longer than those for the control sentences. An ANOVA yielded a significant main effect of sentence type $F_1(2,34) = 698.0$, $p < .001$, $F_2(2,22) = 121$, $p < .001$. Newman–Keuls pairwise comparisons were performed and all means differed significantly ($p < .001$ by subjects and by items). As in Experiment 1, there was a main effect of preparation time, $F_1(2,34) = 9.1$, $p < .001$, $F_2(2,22) = 4.7$, $p < .05$. Newman–Keuls comparisons showed that mean naming duration at PT800 was significantly shorter than

TABLE 9

MEAN PRODUCTION DURATIONS IN MS FOR THE THREE SENTENCE TYPES OF EXPERIMENT 3 AS A FUNCTION OF PREPARATION TIME

Duration	Clitic	Nonclitic	Control	Mean
Phonological words	3	4	3	
Lexical words	4	4	3	
Syllables	5	5	4	
Preparation time				
800	862	1015	794	890
1100	888	1006	801	898
1300	897	1015	810	907
Mean	882	1012	802	

Note. The number of phonological words, lexical words, and syllables in each sentence type are also shown.

at PT1300 ($p < .05$ by subjects and items). As in Experiment 1, subjects showed a slight tendency to speak more slowly at longer preparation times. The interaction of preparation time and sentence type was significant by subjects, $F_1(4,68) = 4.3$, $p < .01$, $F_2(4,44) = 1.3$. As in Experiment 1, this was due to small differences in the size of effects. The pattern of effects across preparation times was the same. Production durations yielded no significant main effect of pairblock and no significant interaction with this variable.

Discussion

This experiment yielded a pattern of results similar to that of Experiment 1: naming latencies were significantly longer for the nonclitic sentences than for the clitic and control sentences which did not differ. The effect of the number of phonological words on sentence production latencies is therefore robust with respect to changes in the position of the primary stress, intonation pattern and absolute stress levels. This result is consistent with the claim that the prosodic representation generated in the prepared speech production task is an abstract phonological representation containing no concrete phonetic information. The effect observed in Experiment 1 was also undiminished in sentences comprising two phonological phrases. Therefore, the number of larger prosodic units a sentence contains does not modulate the effect of number of phonological words it comprises.

Taken together, the results of Experiments 1 to 3 suggest that in the prepared speech production task subjects generate a representation of a sentence which encodes neither syntactic class information nor concrete phonetic detail. These findings are consistent with the claim that subjects generate an abstract prosodic representation of the sentence to be produced.

While Experiments 1 to 3 provide evidence of the construction of a prosodic representation, the prepared speech production paradigm cannot tell us how prosodic structure affects sentence production processes when the time to prepare an utterance is limited. In normal conversational situations, the amount of time a speaker has to prepare an utterance can differ dramatically. A sentence may be held fully prepared while the speaker waits for their turn in a conversation. Alternatively, during a period of fluent speaking, planning will have to occur on-line with limited time and resources. In this situation, if speech is to remain fluent, it is likely that only the minimal production unit is prepared prior to articulation. According to Levelt (1989, 1992) the minimal unit of production is the phonological word. In other words, the articulator must wait until a whole phonological word has been delivered before beginning to output the first syllable. This hypothesis makes a clear prediction about on-line sentence production, namely, that (all other things being equal) latency to produce a sentence should be a function of the time required to construct the first phonological word rather than a function of the total number of phonological words it contains. If this is the case then sentences beginning with long initial phonological words should have longer production latencies than sentences beginning with short phonological words. Such an effect should be distinguished from the effect of initial word length demonstrated by Sternberg et al. (1978) in their prepared speech paradigm (which we failed to replicate in Experiment 1). That was a late effect due to the retrieval and articulation of a *prepared* representation. In Experiment 4, we are looking for an effect of phonological word length on its *on-line construction.*

EXPERIMENT 4

This experiment tested the production of the same sentences used in Experiment 1. The method was essentially the same as in Experiment 1, except that subjects were requested to begin sentence production as soon as they could on hearing the question. In order to measure the sentence construction process, latencies were measured from the onset of the verb in the question. Levelt's model predicts that production latencies should now be a function of the size of the initial phonological word in the utterance. In both the nonclitic and control sentence conditions, the initial phonological word comprises the pronoun and the verb (e.g., $[\text{Ik zoek}]_\omega$). In the clitic sentences the initial phonological word also contain the determiner (e.g., $[\text{Ik zoek het}]_\omega$). Thus, production latencies for the clitic sentences should now be longer than for the nonclitic sentences which should not differ.

Method

Vocabulary. The experimental vocabulary was the same as in Experiment 1. However, in order to prevent subjects from anticipating the noun phrase–verb pairings on presentation of the noun phrase, two additional filler verbs were chosen for each experimental noun phrase. These verbs were also monosyllabic and differed in sound form from the experimental verb. For example, the filler verbs for the sentence, *Ik zoek het water* (I seek the water), were, *kook* (boil) and *test* (test). Each of the 12 experimental nouns now occurred in the three noun phrase conditions in combination with three different verbs, giving a total of 108 different sentences.

Design. Each subject produced the 108 experimental sentences twice. The experiment consisted of six blocks of 36 trials. All experimental sentences occurred once within the first three blocks and were repeated within the next three blocks. Within a block each NP occurred three times, each time with a different verb in a different condition. One of each of the three NP + verb pairings was randomly assigned to each of the three blocks. In the second set of three blocks a different random assignment was used. The order of presentation of blocks 1 to 3 and of blocks 4 to 6 was rotated across subjects such that each sentence occurred in each block position an equal number of times in the first and second halves of the experiment. Six subjects were randomly assigned to each of the three presentation orders.

Procedure. After receiving their instructions, subjects heard an example set of trials with a taped example response. They then completed a practice set of sentences. During the experiment proper each subject produced the 108 experimental sentences twice. Events on each trial were as follows: a fixation cross appeared centered on the screen for 500 ms. Five hundred milliseconds after the offset of the fixation cross a two word phrase appeared centered on the screen for 500 ms. Following another 500-ms pause subjects heard a short question. As soon as the subjects could construct their answer they were to begin speaking. Sentence onset time was measured from the onset of the verb in the question. The voice key was activated by a pulse placed at verb onset in the auditory questions and triggered as usual by subject's voice onset. There was a 2-s pause between trials. Eighteen Dutch speakers were tested. None of the subjects had taken part in any of the previous experiments.

Results

Following the same criteria used in Experiment 1, 4.3% of the data were substituted. The

TABLE 10

Mean Production Latencies and Durations in ms and Percentage Error Rates for Sentences in the Three Experimental Conditions of Experiment 4

	Clitic	Nonclitic	Control
Phonological words	2	3	2
Lexical words	4	4	4
Syllables	5	5	4
Latency	715	697	691
% Error	5.3	7.4	3.0
Duration	738	851	646

resulting mean production latencies, production durations, and percentage error rate in each condition are given in Table 10. The results show a quite different pattern than the results of Experiment 1.

Latencies in the clitic condition are now longer than latencies in both the nonclitic and control conditions which show only a small difference. The main effect of sentence type was significant, $F_1(2,30) = 3.8$, $p < .05$, $F_2(2,22) = 8.6$, $p < .01$. Naming latencies for the clitic sentences were 18 ms longer than those for the nonclitic sentences. Newman–Keuls pairwise comparisons showed that this difference was significant ($p < .05$ by subjects, $p < .01$ by items). Latencies for the clitic sentences were 24 ms longer than the control sentences ($p < .01$ by subjects and by items). The nonclitic and control conditions differed by a nonsignificant 6 ms. An analysis including the variable experiment half yielded a main effect of this variable, $F_1(1,17) = 33.0$, $p < .001$, $F_2(1,35) = 500.2$, $p < .001$, due to an 81-ms decrease in naming latencies in the second half of the experiment. However, experiment half did not interact with sentence type, F_1 & $F_2 < 1$. Percentage error rates also showed a main effect of experiment half $F_1(1,17) = 4.9$, $p < .05$, $F_2(1,35) = 6.5$ $p < .05$, due to a 1.3% decrease in the second half of the experiment. All other effects on percentage error rates were small and nonsignificant.

In contrast to the production latencies, utterance durations show an almost identical pattern of results to utterance durations in Experiment 1. As in Experiment 1, the nonclitic condition has the longest duration and the clitic sentences were 92 ms longer than the control sentences. The main effect of sentence type was again significant, $F_1(2,30) = 325.8$, $p < .001$, $F_2(2,22) = 148.6$, $p < .001$. All differences were highly significant ($p < .001$ in both subject and item analyses). Utterance durations decreased by 30 ms in the second half of the experiment. The main effect of experiment half was significant, $F_1(1,17) = 8.2$, $p < .01$, $F_2(1,35) = 173.7$, $p < .001$. There was also a significant interaction of experiment half and sentence type, $F_1(2,34) = 4.7$, $p < .05$, $F_2(2,70) = 3.3$, $p < .05$, but this was due to small differences in the size of the effects rather than in the pattern of results in each half of the experiment.

Discussion

As predicted, the on-line production task yielded a quite different pattern of results than the prepared speech production task. Production latencies were no longer determined by the total number of phonological words in the sentence but by the complexity of the first phonological word. Clitic sentences now took significantly longer to produce than both the nonclitic and the control sentences which did not significantly differ. Had the article not cliticized with the preceding verb to form a single phonological word, the first unit in all sentence conditions would have been [Ik zoek]$_\omega$ and latencies across conditions should not have differed. In contrast to the latency results, the utterance durations in Experiment 4 were very similar to those observed in Experiment 1, underlining the independence of sentence production latency from sentence duration. Similar to all prepared speech experiments, the results were robust with respect to repetition and cannot, therefore, be attributed to strategies developed by the subjects during the course of the experiment.

The experiment provides strong support for the proposal that the phonological word is the preferred unit of output during speech production (Levelt, 1989, 1992), as subjects clearly prefer to construct such a unit even at the cost of initiation speed.

The experiment also provides support for incremental models of speech production as the length of the entire sentence to be produced did not affect production latency. However, the results of Experiment 4 do not rule out the possibility of nonincremental generation of prosodic structure. The sentences to be produced were short and the boundaries of all prosodic units larger than the phonological word fell at utterance boundary. Moreover, all sentences were produced with the same declarative intonation pattern. Thus, prosodic encoding of these sentences required minimal processing with regards to larger prosodic structures and intonation. It is therefore possible that with longer and more complex sentences effects of whole sentence complexity may be observed in on-line sentence production latency. Nevertheless Experiment 4 demonstrates, that when it is possible to do so, speakers preferentially initiate articulation following the phonological encoding of the initial phonological word of an utterance.

GENERAL DISCUSSION

The experiments we have reported combine to provide good initial evidence that the phonological word forms a unit of processing in the later stages of speech production. Prepared sentence production latencies were a function of the number of phonological words that a sentence contained (Experiment 1) rather than the number of content words it contained (Experiment 2). Moreover, this effect was robust to changes in absolute stress levels, primary stress placement, intonation pattern, and phonological phrase structure (Experiment 3). These findings are consistent with the claim articulation is preceded by the generation of an abstract prosodic representation of an utterance.

In contrast, on-line speech production latencies were a function of the complexity of the initial phonological word to be produced (Experiment 4). This finding provides strong evidence that the phonological word is the preferred unit of output in fluent speech production. We will now attempt to relate these findings in a more detailed way to speech production processes.

As discussed in the introduction, Sternberg et al. (1978, 1980) provide a detailed model of processing in the prepared speech production paradigm. They model the retrieval process as a serial search through an unordered nondecreasing buffer (i.e., one in which elements remain even after they are unpacked and articulated). The Sternberg et al. model provides a good account of their data. Importantly, both the latency data and the duration data are accounted for by the same mechanism. However, two aspects of this model are unsatisfactory when we consider normal speech production. First, prosodic representations have intrinsic order. Many aspects of phonological and phonetic structure are determined by the position of an element in the prosodic representation and by the nature of the units flanking it. It therefore makes little sense to lose this ordering when these elements are buffered. Second, the idea of a nondecreasing buffer is somewhat difficult to reconcile with the fluency of speech production. Even when some advanced buffering occurs it seems inefficient to suggest that all buffered elements are articulated before the buffer clears for the next chunk of output. Nevertheless, it is possible that this buffering behaviour is specific to the prepared speech production situation. Further work is, of course, required to see if sentence materials show a similar duration effect and to provide independent evidence of a nondecreasing buffer.

An alternative account for the latency effect (a version of which was considered by Sternberg et al., 1978) avoids the first problem mentioned above. According to this account, subjects prepare an *ordered* phonological representation of the sentence during the preparation period. The elements of this representation are phonological words. All elements of the representation must be held active until they can be produced, but in order for an element to be selected for phonological encoding it must be activated beyond the level of the other elements. If all activated elements compete to be produced then this competition could lead to an increase in naming latency that is a function of the number of activated elements. If we are also willing to assume a nondecreasing buffer then naming duration could also be accounted for as the sum of retrieval latencies for each element in the representation.

Our account also differs from that of Sternberg et al. (1978, 1980) in the nature of representation that we claim is constructed during the preparation period. They propose a motor program—a phonetic–articulatory representation—whereas we suggest an abstract phonological representation. Monsell (1986) argues for a motor representation because of the limited impact of an additional short-term memory task on list preparation (subjects were given two lists to prepare and had to recall the second list without time pressure following the rapid production of the first list). It has been claimed that verbal short-term memory tasks involve retention of the sequence in a *phonological* buffer (Baddeley, Thomson, & Buchanan, 1975; Vallar & Baddeley, 1984). If the list prepared for rapid production is also phonologically encoded one would predict disrupted performance due to the extra demands on shared resources.

However, 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 et al., 1975; Cowan, Day, Saults, Keller, Johnson, & Flores, 1992 Exp 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). One possible interpretation of the finding that the *spoken* duration of the items affects performance is that the to-be-remembered-items are encoded at a quantitative phonetic level of representation. This interpretation is also consistent with research that demonstrates that phonological processes can survive articulatory suppression (Wheeldon & Levelt, 1995) although the duration effect disappears (Longoni, Richardson, & Aiello, 1993). Similar to Monsell (1986), we propose that the independence of prepared speech production from additional short-term memory requirements can be explained by proposing that these tasks rely on different levels of representation. However, our claim is that the prepared speech representation is abstract and phonological in nature, whereas the short-term memory representation is phonetic.

Returning to on-line speech production processes, we have argued that the results of Experiment 4 are consistent with an incremental model of sentence production. In this experiment sentences beginning with the phonological word *ik zoek het* (I seek the) yielded significantly longer production latencies than sentences beginning with the phonological word *ik zoek* (I seek) regardless of the number of phonological words required to complete the utterance. Nevertheless, we still do not know which aspect of the generation of the clitic phonological words caused the increase in naming latency. The generation of the phonological word, *ik zoek het,* is more complex than the generation of the phonological word, *ik zoek*, in a number of ways. First, one extra lexical item (*het*) must be retrieved and associated to the surface syntactic representation. Moreover, in the sentences used, the form of the determiner, *het,* is dependent on the gender of the noun, *water*. Therefore, before we can articulate the phonological word *Ik zoek het,* we need also to have constructed the noun phrase, *het water*. Finally, we must also have generated the phonological form for *het.* We need to do none of these things in order to produce the phonological word, *Ik zoek.* Further experimentation is required before we can determine the relative contribution of these factors to the latency effect. Finally, in our experiments we examined the production of short sentences with syntactic complexity held constant. Future work will involve the production of longer sentences to investigate how varying syntactic complexity interacts with the incremental phonological processing we have demonstrated.

In conclusion, the experiments we have reported provide evidence that articulation is preceded by the generation of prosodic structure and demonstrate that sentence production latencies can be used to gain insight into the processes by which we generate rhythmic connected speech.

APPENDIX 1: EXPERIMENTAL MATERIALS FOR EXPERIMENT 1 AND 4 WITH THEIR ENGLISH TRANSLATIONS

Clitic (two phonological words)	Nonclitic (three phonological words)	Control (two phonological words)	Question
(1) Ik drink de wijn (I drink the wine)	Ik drink Jans wijn (John's)	Ik drink wijn	WAT DRINK JE? *What drink you?* (What do you drink?)
(2) Ik lees de kranten (I read the newspapers)	Ik lees Jans kranten	Ik lees kranten	WAT LEES JE? (What do you read?)
(3) Ik mors de suiker (I spill the sugar)	Ik mors Jans suiker	Ik mors suiker	WAT MORS JE? (What do you spill?)
(4) Ik was de kleding (I wash the clothes)	Ik was Jans kleding	Ik was kleding	WAT WAS JE? (What do you wash?)
(5) Ik weeg het fruit (I weight the fruit)	Ik weeg vers fruit (fresh)	Ik weeg fruit	WAT WEEG JE? (What do you weigh?)
(6) Ik zoek het water (I seek the water)	Ik zoek vers water	Ik zoek water	WAT ZOEK JE? (What do you seek)
(7) Ik proef het ijs (I taste the ice cream)	Ik proef vers ijs	Ik proef ijs	WAT PROEF JE? (What do you taste?)
(8) Ik koop het eten (I buy the food)	Ik koop vers eten	Ik koop eten	WAT KOOP JE? (What do you buy?)
(9) Ik eet te snel (I eat too fast)	Ik eet heel snel (very)	Ik eet snel	HOE EET JE? (How do you eat?)
(10) Ik trap te hard (I kick too hard)	Ik trap heel hard	Ik trap hard	HOE TRAP JE? (How do you kick?)
(11) Ik krab te zacht (I scratch too softly)	Ik krab heel zacht	Ik krab zacht	HOE KRAB JE? (How do you scratch?)
(12) Ik verf te slecht (I paint too badly)	Ik verf heel slecht	Ik verf slecht	HOE VERF JE? (How do you paint?)

APPENDIX 2: EXPERIMENTAL MATERIALS FOR EXPERIMENT 2 WITH THEIR ENGLISH TRANSLATIONS

Clitic (two phonological words)	Nonclitic (three phonological words)	Pronoun (two phonological words)	Control (one phonological word)	Questions
(1) Ik drink de wijn (I drink the wine)	Ik drink veel wijn (much)	Ik drink het (it)	Ik drink	WAT DRINK JE? (what do you drink?) WAT DOE JE? (What do you do?)
(2) Ik mors de suiker (I spill the sugar)	Ik mors veel suiker	Ik mors het	Ik mors	WAT MORS JE? (what do you spill?) WAT DOE JE?
(3) Ik lees de krant (I read the newspaper)	Ik lees Jans krant (John's)	Ik lees het	Ik lees	WAT LEES JE? (what do you read?) WAT DOE JE?
(4) Ik proef het ijs (I taste the ice cream)	Ik proef Jans ijs	Ik proef het	Ik proef	WAT PROEF JE? (what do you taste?) WAT DOE JE?
(5) Ik weeg het fruit (I weight the fruit)	Ik weeg vers fruit (fresh)	Ik weeg het	Ik weeg	WAT WEEG JE? (what do you weigh?) WAT DOE JE?
(6) Ik zoek het water (I seek the water)	Ik zoek vers water	Ik zoek het	Ik zoek	WAT ZOEK JE? (what do you seek?) WAT DOE JE?
(7) Ik eet te snel (I eat too fast)	Ik eet heet snel (very)	Ik eet het	Ik eet	HOE EET JE? (How do you eat?) WAT DOE JE?
(8) Ik trap te hard (I kick too hard)	Ik trap heel hard	Ik trap het	Ik trap	HOE TRAP JE? (How do you kick?) WAT DOE JE?
(9) Ik krab te zacht (I scratch too softly)	Ik krab heel zacht	Ik krab het	Ik krab	HOE KRAB JE? (How do you scratch?) WAT DOE JE?

REFERENCES

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.

Berendsen, E. (1986). *The phonology of cliticization.* Dordrecht: Foris Publications.

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* Academic Press, San Diego.

Booij, G., & Lieber, R. (1993). On the simultaneity of morphological and prosodic structure. In S. Hargus and E. M. Kaisse (Eds.), *Studies in lexical phonology: Phonetics and phonology* (Vol. IV). Academic Press, San Diego.

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.

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.

Cooper, W. E., & Paccia-Cooper, J. M. (1980). *Syntax and speech.* Cambridge, MA: Harvard Univ. Press.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production.* New York: Springer.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production.* Berlin Mouton.

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.

De Pijper, J. R. (1983). *Modelling British English intona-*

tion: An analysis by resynthesis of British intonation. Dordrecht: Foris.

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

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language,* **30,** 210–233.

Ferreira, F. (1994). The creation of prosody during sentence production. *Psychological Review,* **100,** 233–253.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.) *Language production* (Vol 1., pp. 177–220). London: Academic Press.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: Academic Press.

Garrett, M. F. (1990, March). *Processing vocabularies in language production.* Paper presented at the CUNY Conference on Human Sentence Processing, New York.

Gussenhoven, C. (1989). *Cliticization in Dutch as phonological word formation.* Unpublished manuscript. Univ. of Nijmegen.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & Youmans (Eds.), *Phonology and meter: Phonetics and phonology* (Vol 1, pp. 201–260). San Diego: Academic Press.

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,* **41**A, 321–327.

Inkelas, S., & Zec, D. (1990). *The phonology–syntax connection.* Chicago: Univ. of Chicago Press.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formation. *Cognitive Science,* **11,** 201–258.

Klapp, S. T. (1974). Syllable-dependent pronunciation latencies in number naming, a replication. *Journal of Experimental Psychology,* **102,** 1138–1140.

Klapp, S. T., Anderson, W. G., & Berrian, R. W. (1973). Implicit speech in reading reconsidered. *Journal of Experimental Psychology,* **100,** 368–374.

Lahiri, A., Jongman, A., & Sereno, J. A. (1990). The pronominal clitic [der] in Dutch: Theoretical and experimental approach. *Yearbook of Morphology,* **3,** 115–127.

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

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry,* **8.2,** 249–336.

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

Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In C. Fromm and H. Heuer (Eds.) *Generation and modulation of action patterns.* Berlin: Springer.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

Odden, D. (1980). *The phrasal phonology of Kimatuumbi.* Unpublished manuscript, Yale Univ., New Haven, CT.

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

Saffran, E. M., Schwartz, F., & Martin, O. S. M. (1980). Evidence form aphasia: Isolating the components of a production model. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 221–241). London: Academic Press.

Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Arnoff and M. L. Kean (Eds.), *Juncture* (Studia linguistica et philologica) (pp. 107–29). Saratoga, CA: Anma Libri.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook,* 371–405.

Selkirk, E. (1995). The prosodic structure of function words. In J. Beckman, L., Walsh Dickey, and S. Urbanczyk (Eds.) *Papers in Optimality Theory, UMASS Occasional Papers in Phonology.* Amherst: GLSA.

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

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning,* (pp. 117–152). New York: Academic Press.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *The perception and production of fluent speech* (pp. 507–534). Englewood Cliffs, NJ: Erlbaum.

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

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

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

(Received June 5, 1995)
(Revision received January 28, 1997)