

Do speakers have access to a mental syllabary?

Willem J.M. Levelt*, Linda Wheeldon

Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525 XD, Netherlands

Abstract

The first, theoretical part of this paper sketches a framework for phonological encoding in which the speaker successively generates phonological syllables in connected speech. The final stage of this process, phonetic encoding, consists of accessing articulatory gestural scores for each of these syllables in a “mental syllabary”. The second, experimental part studies various predictions derived from this theory. The main finding is a syllable frequency effect: words ending in a high-frequent syllable are named faster than words ending in a low-frequent syllable. As predicted, this syllable frequency effect is independent of and additive to the effect of word frequency on naming latency. The effect, moreover, is not due to the complexity of the word-final syllable. In the General Discussion, the syllabary model is further elaborated with respect to phonological underspecification and activation spreading. Alternative accounts of the empirical findings in terms of core syllables and demisyllables are considered.

Introduction

The purpose of the present paper is to provide evidence for the notion that speakers have access to a *mental syllabary*, a repository of articulatory-phonetic syllable programs. The notion of stored syllable programs originates with Crompton (1982) and was further elaborated in Levelt (1989, 1992, 1993). The latter two papers introduced the terms “phonetic” and “mental syllabary” for this hypothetical mental store. Most current theories of speech production model the pre-articulatory form representation at a phonological level as consisting of

*Corresponding author. E-mail PIM@MPI.NL

The authors wish to thank Ger Desserjer for his assistance with the running and analysis of all of the experiments reported.

discrete segments or features (Dell, 1988; Shattuck-Hufnagel, 1979) and some models assume explicitly that this level of representation directly activates articulatory routines (Mackay, 1982). However, the actual phonetic realization of a phonological feature is determined by the context in which it is spoken. The fact that phonetic context effects can differ across languages means that they cannot all be due to the implementation of universal phonetic rules but most form part of a language-dependent phonetic representation (Keating, 1988). The mental syllabary was postulated as a mechanism for translating an abstract phonological representation of an utterance into a context-dependent phonetic representation which is detailed enough to guide articulation.

The present paper will provide experimental evidence that is consistent with the existence of a mental syllabary and provides a challenge to theories that assume (tacitly or otherwise) that the phonetic forms of all syllables are generated anew each time they are produced. We present here a more detailed model of syllable retrieval processes than has previously been attempted, and while we readily admit that much further evidence is required in order to substantiate it, we propose it as a productive framework for the generation of empirical research questions and as a clear target for further empirical investigation.

In the following we will first discuss some of the theoretical reasons for assuming the existence of a syllabary in the speaker's mind. We will then sketch a provisional framework for the speaker's encoding of phonological words – a framework that incorporates access to a phonetic syllabary. This theoretical section will be followed by an empirical one in which we present the results of four experiments that address some of the temporal consequences of a speaker's retrieving stored syllable programs during the ultimate phase of phonological encoding. In the final discussion section, we will return to a range of further theoretical issues that are worth considering, given the notion of a mental phonetic syllabary.

The syllabary in a theory of phonological encoding

Crompton's suggestion

As with so many notions in theories of speech production, the idea that a speaker retrieves whole phonetic syllable programs was originally proposed to account for the occurrence of particular speech errors. Crompton (1982) suggested the existence of a library of syllable-size articulatory routines to account for speech errors involving phonemes and syllable constituents. For example, an error like *guinea hig pair* (for *guinea pig hair*) arises when the mechanism of addressing syllable routines goes awry. The articulatory syllables [pIg] and [hɛəɾ] in the library are addressed via sets of phonemic search instruction such as:

onset = p	onset = h
nucleus = I	nucleus = εə
coda = g	coda = r

If these search instructions get mixed up, leading to the exchange of the onset conditions, then instructions arise for the retrieval of two quite different articulatory syllables, namely [hIɡ] and [pɛər]. This provides an elegant account for the phonetic “accommodation” that takes place: [hIɡ] is pronounced with the correct allophone [ɪ], not with [ɛ] that would have been the realization of [h] in *hair*. This addressing mechanism, Crompton argues, is fully compatible with Shattuck-Hufnagel’s (1979) scan copier mechanism of phonological encoding. According to that model, a word’s phonological segments are spelled out from the word’s lexical representation in memory, and inserted one-by-one into the slots of a syllabic frame for the word that is independently retrieved from the word’s lexemic representation. This copier mechanism in fact specifies the search instructions for each of a word’s successive syllables (i.e., onset of syllable 1, nucleus of syllable 1, coda of syllable 1, onset of syllable 2, etc.).

A functional paradox

However, in the same paper Crompton reminds us of a paradox, earlier formulated by Shattuck-Hufnagel (1979, p. 338), but not solved by either of them: “perhaps its [the scan copier’s] most puzzling aspect is the question of why a mechanism is proposed for the one-at-a-time serial ordering of phonemes when their order is already specified in the lexicon”. Levelt (1992) formulated this functional paradox as follows:

Why would a speaker go through the trouble of first generating an empty skeleton for the word, and then filling it with segments? In some way or another both must proceed from a stored phonological representation, the word’s phonological code in the lexicon. Isn’t it wasteful of processing resources to pull these apart first, and then to combine them again (at the risk of creating a slip)?

And (following Levelt, 1989) he argued that the solution of the paradox should be sought in the generation of connected speech. In connected speech it is the exception rather than the rule that a word’s canonical syllable skeleton is identical to the frame that will be filled. Instead, new frames are composed, not for lexical words (i.e., for words in their citation form), but for phonological words, which often involve more than a single lexical word. It is only at this level that syllabification takes place, not at any earlier “citation form” level. Let us now outline this framework in more detail (see Fig. 1).

Word from retrieval

A first step in phonological encoding is the activation of a selected word’s

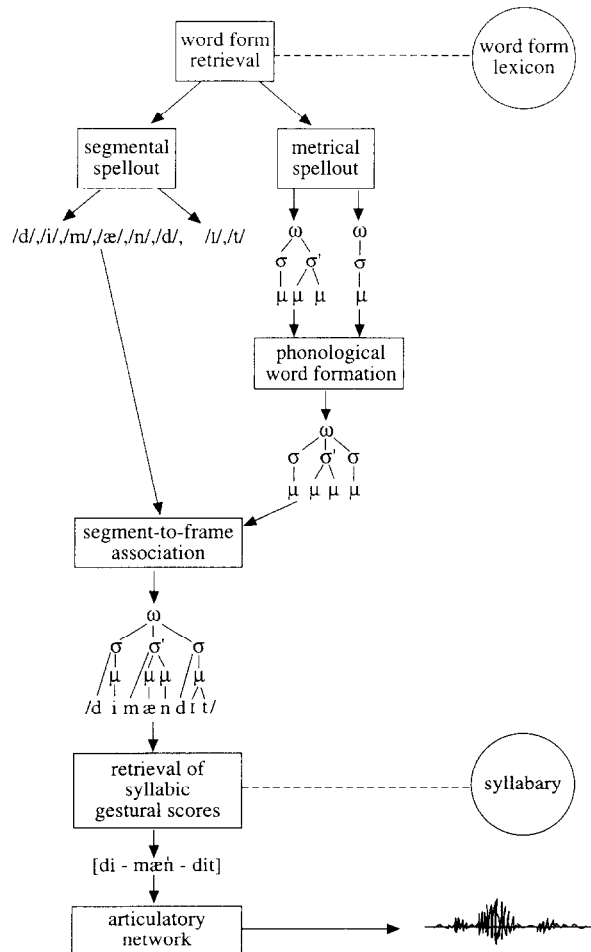


Figure 1. A framework for phonological encoding.

“lexeme” – the word’s form information in the mental lexicon. In Fig. 1 this is exemplified for two words, *demand* and *it*, as they could appear in an utterance such as *police demand it*. Although terminologies differ, all theories of phonological encoding, among them Meringer and Mayer (1895), Shattuck-Hufnagel (1979), Dell (1988) and Levelt (1989), distinguish between two kinds of form information: a word’s segmental and its metrical form.

The segmental information relates to the word’s phonemic structure: its composition of consonants, consonant clusters, vowels, diphthongs, glides, etc. Theories differ with respect to the degree of specification, ranging from minimal or underspecification (Stemberger, 1983) to full phonemic specification (Crompton, 1982), and with respect to the degree of linear ordering of segmental

information. Without prejudging these issues (but see Discussion below), we have represented segments in Fig. 1 by their IPA labels and as consonantal or vocalic (C or V).

The metrical information is what Shattuck-Hufnagel (1979) called the word's "frame". It specifies at least the word's number of syllables (its "syllabicity") and its accent structure, that is, the lexical stress levels of successive syllables. Other metrical aspects represented in various theories are: onset versus rest of word (Shattuck-Hufnagel 1992), the precise CV structure of syllables (Dell, 1988), the degree of reduction of syllables (Crompton, 1982) and (closely related) whether syllables are strong or weak (Levelt, 1993). Our representation in Fig. 1 follows Hays (1989) as far as a syllable's weight is represented in a moraic notation (one mora for a light syllable, two morae for a heavy one). This is not critical, though; weight could also be represented by branching (vs. not branching) the nucleus. But the mora representation simplifies the formulation of the association rules (see below). The representation of accent structure in Fig. 1 is no more than a primitive "stressed" (with ') versus unstressed (without ').

There is also general agreement that metrical information is, to some extent, independently retrieved. This is sometimes phenomenologically apparent when we are in a "tip-of-the-tongue" state, where we fail to retrieve an intended word, but feel pretty sure about its syllabicity and accent structure. This relative independence of segmental and metrical retrieval is depicted in Fig. 1 as two mechanisms: "segmental spellout" and "metrical spellout" (see Levelt, 1989, 1993, for more details).

An important aspect of form retrieval, which will play an essential role in the experimental part of this paper, is that it is frequency sensitive. Jescheniak and Levelt (in press) have shown that the word frequency effect in picture naming (naming latency is longer for pictures with a low-frequency name than for pictures with a high-frequency name) is entirely due to accessing the lexeme, that is, the word's form information.

Phonological word formation

A central issue for all theories of phonological encoding is how segments become associated to metrical frames. All classical theories, however, have restricted this issue to the phonological encoding of single words. However, when generating connected speech, speakers do not concatenate citation forms of words, but create rhythmic, pronounceable metrical structures that largely ignore lexical word boundaries. Phonologists call this the "prosodic hierarchy" (see, for instance, Nespor & Vogel, 1986). Relevant here is the level of phonological words (or clitic groups). In the utterance *police demand it*, the unstressed function word *it* cliticizes to the head word *demand*, resulting in the phonological word *demandit*. Of crucial importance here is that phonological words, not lexical

words, are the domain of syllabification. The phonological word *demandit* is syllabified as *de-man-dit*, where the last syllable straddles a lexical boundary. Linguists call this “resyllabification”, but in a processing model this term is misleading. It presupposes that there was lexical syllabification to start with (i.e., *de-mand + it*). There is, in fact, good reason to assume that a word’s syllables are not fully specified in the word form lexicon. If they were, they would regularly be broken up in connected speech. That is not only wasteful, but it also predicts the occurrence of syllabification speech errors such as *de-mand-it*. Such errors have never been reported to occur in fluent connected speech.

In short, there must be a mechanism in phonological encoding that creates metrical frames for phonological words. This is depicted in Fig. 1 as “phonological word formation”. Notice that this is an entirely metrical process. There are no known *segmental* conditions on the formation of phonological words (such as “a word beginning with segment *y* cannot cliticize to a word ending on segment *x*”). The conditions are syntactic and metrical. Essentially (and leaving details aside), a phonological word frame is created by blending the frames of its constituent words, as depicted in Fig. 1.

Segment-to-frame association

The next step in phonological encoding, then, is the association of spelled-out segments to the metrical frame of the corresponding phonological word. There is good evidence that this process runs “from left to right” (Dell, 1988; Meyer, 1990, 1991; Meyer & Schriefers, 1991; Shattuck-Hufnagel, 1979). But the mechanisms proposed still vary substantially. However, whatever the mechanisms, they must adhere to a language’s rules of syllabification.

Levelt (1992) presented the following set of association rules for English, without any claim to completeness:

- (1) A vowel only associates to μ , a diphthong to $\mu\mu$.
- (2) The default association of a consonant is to σ . A consonant associates to μ if and only if any of the following conditions hold:
 - (a) the next element is lower in sonority;
 - (b) there is no σ to associate to;
 - (c) associating to σ would leave a μ without associated element.

In addition, there is a general convention that association to σ , the syllable node, can only occur on the left-hand side of the syllable, that is, to the left of any unfilled morae of that syllable. See Levelt (1992) for a motivation of these rules.

On the assumption that spelled-out segments are ordered, and that association proceeds “from left to right”, a phonological word’s syllabification is created “on the fly” when these rules are followed. The reader can easily verify that for

demandit the syllabification becomes *de-man-dit*, where the last syllable straddles the lexical boundary.

It should be noticed that this is not an account of the *mechanism* of segment-to-frame association. It is doubtless possible to adapt Shattuck-Hufnagel's (1979) scan-copier mechanism or Dell's (1988) network model to produce the left-to-right association proposed here. The adaptations will mainly concern (i) the generation of phonological, not lexical word frames, and (ii) the use of more global syllable frames, that is, frames only specified for weight, not for individual segmental slots.

Accessing the syllabary

The final step of phonological encoding (which is sometimes called phonetic encoding) is to compute or access the articulatory gestures that will realize a phonological word's syllables. It is at this point that the notion of a mental syllabary enters the picture. But before turning to that, we should first say a few words about what it is that has to be accessed or computed.

We suggest that it is what Browman and Goldstein (1991) have called *gestural scores*. Gestural scores are, like choreographic or musical scores, specifications of tasks to be performed. Since there are five subsystems in articulation that can be independently controlled, a gestural score involves five "tiers". They are the glottal and the velar system, plus three tiers in the oral system: tongue body, tongue tip and lips. Example of a gestural task is to close the lips, as in the articulation of *apple*. The gestural score only specifies *that* the lips should be closed, but not *how* it should be done. The speaker can move the jaw, the lower lip, both lips, or all of these articulators to different degrees. But not every solution is equally good. As Saltzman and Kelso (1987) have shown, there are least-effort solutions that take into account which other tasks are to be performed, what the prevailing physical conditions of the articulatory system are (does the speaker have a pipe in his mouth that wipes out jaw movement?), etc. These computations are done by what they called an "articulatory network" – a coordinative motor system that involves feedback from the articulators. Relevant here is that gestural scores are *abstract*. They specify the tasks to be performed, not the motor patterns to be executed.

The gestural score for a phonological word involves scores for each of its syllables. The issue here is: how does a speaker generate these scores? There may well be a direct route here, as Browman and Goldstein have convincingly argued. A syllable's phonological specifications are, to some extent, already specifications of the gestural tasks that should be carried out in order to realize the syllable. One can present a reader with a phonotactically legal non-word that consists of non-existing syllables (such as *flitirp*), and the reader will pronounce it all right.

Still, there may be another route as well. After all, most syllables that a speaker uses are highly overlearned articulatory gestures. It has been argued time

and again that most (though not all) phenomena of allophonic variation, of coarticulation and of assimilation have the syllable as their domain (see, for instance, Fujimura & Lovins, 1978; Lindblom, 1983). In other words, if you know the syllable and its stress level, you know how to pronounce its segments. Or rather: phonetic segments have no independent existence; they are mere properties of a syllabic gesture, its onset, nucleus and offset. If these syllabic scores are overlearned, it is only natural to suppose that they are accessible as such, that is, that we have a store of syllabic gestures for syllables that are regularly used in speech.

This is depicted in Fig. 1 as the syllabary. According to this theory, the syllabary is a finite set of pairs consisting of, on the one hand, a phonological syllable specification and, on the other hand, a syllabic gestural score. The phonological specification is the input address; the gestural score is the output. As phonological syllables are, one by one, created during the association process, each will activate its gestural score in the syllabary. That score will be the input to the “articulatory network” (see above), which controls motor execution of the gesture. Crompton (1982) made the suggestion that articulatory routines for stressed and unstressed syllables are independently represented in the repository, and this was adopted in Levelt (1989). It should be noticed that the size of the syllabary will be rather drastically different between languages, ranging from a few hundred in Chinese or Japanese to several thousands in English or Dutch.

So far for the theoretical framework. It is obvious that many theoretical issues have not (yet) been raised. It is, in particular, not the intention of the present paper to go into much more detail about the initial processes of phonological encoding, segmental and metrical spellout, phonological word formation and segment-to-frame association. We will, rather, focus on the final step in the theory: accessing the syllabary. It is important to notice this step has a certain theoretical independence. Most theories of phonological encoding are not specific about phonetic encoding, and many of them would be compatible with the notion of a syllabary. Still, as will be taken up in the General Discussion, the syllabary theory may have interesting consequences for an underspecification approach to phonological encoding. It may provide an independent means of determining what segmental features should minimally be specified in the form lexicon.

The following four experiments were inspired by the notion of a syllabary. Their results are compatible with that notion, but alternative explanations are by no means excluded. Still, they provide new evidence about the time course of phonetic encoding that has not been predicted by other theories.

EXPERIMENT 1: WORD AND SYLLABLE FREQUENCY

According to the theory outlined above, there are two steps in phonological encoding where the speaker accesses stored information. The first one is in

retrieving word form information, that is, the lexeme. The second one is in retrieving the syllabic gestural score. The former involves the form part of the mental lexicon, the latter the syllabary. We have modelled these two steps as successive and independent.

It has long been known that word form access is sensitive to word frequency. Oldfield and Wingfield (1965) and Wingfield (1968) first showed that naming latencies for pictures with low-frequency names are substantially longer than latencies for pictures with high-frequency names. The effect is, moreover, not due to the process of recognizing the picture; it is a genuinely lexical one. Jescheniak and Levelt (in press) have further localized the effect in word form access. Accessing a low-frequent homophone (such as *wee*) turned out to be as fast as accessing non-homophone controls that are matched for frequency to the corresponding high-frequent homophone (in case, *we*). Since homophones, by definition, share their word form information, but not their semantic/syntactic properties, the frequency effect must have a form-level locus: the low-frequent homophone inherits the form-accessing advantage of its high-frequent twin.

It is, however, enough for the rationale of the experiment to know that there is a genuinely lexical frequency effect in word retrieval, and to assume that accessing the syllabary is a later and independent step in phonological encoding. Similar to word retrieval, accessing the store of syllables might also involve a frequency effect: accessing a syllable that is frequently used in the language may well be faster than accessing a syllable that is less frequently used.

The experiment was designed to look for an effect on word production latency of the frequency of occurrence of a word's constituent syllables. High- and low-frequency bisyllabic words were tested which comprised either two high-frequency syllables or two low-frequency syllables. Whole-word frequency of occurrence was therefore crossed with syllable frequency, allowing us to test for any interaction. The syllabary theory predicts that the effects should be additive and independent.

Method

In the following experiments the linguistic restrictions on the selection of experimental materials were severe. It is, in particular, impossible to obtain the relevant naming latencies by means of a picture-naming experiment; there are simply not enough depictable target words in the language. We therefore designed another kind of naming task, that would put minimal restrictions on the words we could test. In the preparation phase of the experiment, subjects learned to associate each of a small number of target words to an arbitrary symbol. During the experiments, these symbols were presented on the screen and the subjects produced the corresponding target words; their naming latencies were measured.

Notice that we decided against a word-reading task, which would always involve linguistic processing of the input word.

Frequency counts

All frequency counts were obtained from the computer database CELEX¹, which has a Dutch lexicon based on 42 million word tokens. The word frequency counts we used are two occurrences per million counts from this database: word form frequency, which includes every occurrence of that particular form, and lemma frequency, which includes the frequencies of all word forms with the same stem. Syllable frequencies were counted for phonetic syllables in Dutch. The phonetic script differentiates the reduced vowel schwa from full vowel forms, giving approximately 12,000 individual syllable forms. Syllable frequencies were calculated for the database from the word form occurrences per million count. Two syllable frequency counts were calculated: overall frequency of occurrence and the frequency of occurrence of the syllable in a particular word position (i.e., first or second syllable position). The syllable frequencies range from 0 to approximately 90,000 per million words, with a mean frequency of 121. In all of the experiments reported the same criteria were used in assigning words to frequency conditions. All low-frequency words had a count of less than 10 for both word form and lemma counts. All high-frequency words had both counts over 10. Low-frequency syllables had counts of less than 300 in both overall and position-dependent counts; high-frequency syllables had both counts over 300. Most low-frequency syllables, therefore, had above-average frequency of occurrence in the language. This is important as our model claims that very low-frequency syllables will be constructed on-line rather than retrieved from store. We are aware of the fact that we have been counting citation form syllables, not syllables as they occur in connected speech. But if the latter frequency distribution deviates from the one we used, this will most likely work against our hypothesis; our distinct HF and LF syllable classes will tend to be blurred in the “real” distribution.

Vocabulary

The experimental vocabulary comprised four groups of 16 bisyllabic Dutch words. These groups differed in the combination of word frequency and syllable frequency of their constituent words. Average frequencies for each word group are given in Table 1. Each group contained 13 nouns and three adjectives. Groups

¹The Centre for Lexical Information (CELEX), Max Planck Institute, The Netherlands.

Table 1. *Log syllable and word frequencies and number of phonemes of words in each of the Word × Syllable frequency groups of Experiment 1*

log syllable frequency:	High	Low	High	Low
Log word frequency:	High	High	Low	Low
Word form	3.3	3.2	0.4	0.3
Lemma	3.6	3.6	0.6	0.4
1st syllable position dependent	7.4	4.7	7.5	4.2
1st syllable total	7.9	5.0	8.1	4.6
2nd syllable position dependent	7.3	4.1	7.3	3.3
2nd syllable total	8.2	4.1	8.0	3.6
Number of phonemes	5	6	5	6

were also matched for word onset phonemes and mean number of phonemes. Each group was divided into four matched subgroups which were recombined into four experimental vocabularies of 16 words (four from each condition; see Appendix 1). Within each vocabulary four groups of four words (one from each condition) were selected to be elicited in the same block. These groups contained words which were phonologically and semantically unrelated and each group contained at least one word with second syllable stress.

Symbols

Four groups of four symbol strings were constructed. Each symbol consisted of a string of six non-alphabetic characters. The four groups of symbols were roughly matched for gross characteristics as follows:

Set 1	Set 2	Set 3	Set 4
))))))	\\ \\ \\ \\ \\ \\	} } } } } }	[[[[[[[
% % % % % %	& & & & & &	# # # # # #	@ @ @ @ @ @
> > > > >	: : : : : :	= = = = = =	; ; ; ; ; ;
" " " " " "	~ ~ ~ ~ ~ ~	' ' ' ' ' '	^ ^ ^ ^ ^ ^

Design

Subjects were assigned to one of the vocabularies. Their task was to learn to produce words in response to symbols. Subjects learned one block of four words at a time. The experiment consisted of 12 blocks of 24 naming trials – three blocks for each four-word set. Within a block subjects produced each word six times. The first production of each word in a block was a practice trial. Order of presentation

was random, with the condition that no symbol occurred twice in a row. This condition was included in order to eliminate the potentially large facilitation effect due to immediate repetition and to encourage subjects to clear their minds at the end of each trial. Within a vocabulary the order of presentation of block groups was rotated across subjects. Within a vocabulary each block group was assigned a symbol set. The assignment of symbols to words within sets was also rotated across subjects.

Procedure

Subjects were tested individually. They were given a card on which four words with associated symbols were printed. They were asked to practise the relationship between the symbols and the words until they thought they could accurately produce the words in response to the symbols. When each subject was confident that they had learned the associations they were shown each symbol once on the computer screen and asked to say the associated word. If they could do this correctly they then received three blocks of 24 trials. The events on each trial were as follows. A fixation cross appeared on the screen for 300 ms. The screen then went blank for 500 ms, after which a symbol appeared on the screen and remained there for a further 500 ms. Subject then had 2 s in which to respond, followed by a 3 s interval before the onset of the next trial. This procedure was repeated for all four groups of words. The printed order of the words from each frequency group was rotated across block groups. Both naming latencies and durations were recorded for each trial.

Subjects

Thirty-two subjects were tested, 24 women and 8 men. All were native speakers of Dutch. They were voluntary members of the Max-Planck subjects pool, between the ages of 18 and 34. They were paid for their participation.

Results

Exclusion of data

Data from two subjects were replaced due to high error rates. The first production of a word in each block was counted a practice trial and excluded from the analysis. Correct naming latencies following error trials were also excluded from the latency analysis as errors can often perturb subject's responses on the

following trial. 3.3% of the data points were lost due to these criteria. Data points greater than two standard deviations from the mean were counted as outliers and were also excluded. This resulted in the loss of only 1.6% of the data points. Missing values in all experiments reported were substituted by a weighted mean based on subject and item statistics calculated following Winer (1971, p. 488).

Naming latency

Collapsed across syllable frequency, high and low word frequency latencies were 592.0 ms and 607.2 ms respectively. The main effect of word frequency (15.2 ms) was significant, $F_1(1, 28) = 14.9$, $p < .001$, $F_2(1, 48) = 4.2$, $p < .05$. Collapsed across word frequency, high and low syllable frequency latencies were 592.3 ms and 606.8 ms respectively. The main effect of syllable frequency (14.5 ms) was also significant, $F_1(1, 28) = 17.7$, $p < .001$, $F_2(1, 48) = 3.8$, $p = .052$. Mean naming latencies for words in each of the frequency groups are shown in Fig. 2. The size of the syllable frequency effect is similar in both word frequency groups and vice versa: the interaction of word and syllable frequency was insignificant, F_1 and $F_2 < 1$.

There was a significant effect of vocabulary in the materials analysis, $F_1(3, 28) = 1.2$, $F_2(3, 48) = 7.8$, $p < .001$, but no interactions of this variable with either syllable or word frequency.

Effects of practice were evident in the significant decrease in naming latencies across the three blocks of a word group, $F_1(2, 56) = 203.1$, $p < .001$, $F_2(2, 96) =$

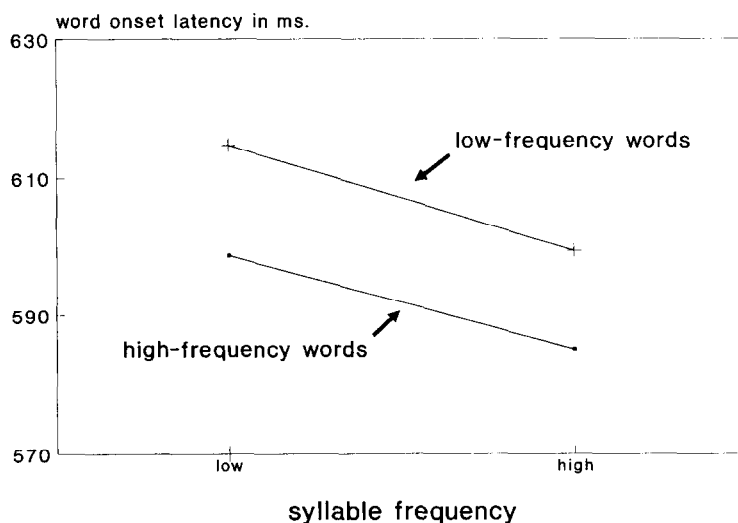


Figure 2. Naming latencies in Experiment 1. Syllable versus word frequency.

318.8, $p < .001$, and across the five repetitions of a word within a block, $F_1(4, 112) = 25.7$, $p < .001$, $F_2(4, 192) = 23.8$, $p < .001$. The effect of block did not interact with either word or syllable frequency effects (all $F_s < 1$). The effect of trial, however, showed an interaction with syllable frequency that approached significance by subjects, $F_1(4, 112) = 2.3$, $p < .06$, $F_2(4, 192) = 1.5$. However, this interaction was due to variation in the size of the priming effect over trials but not in the direction of the effect and does not qualify the main result.²

Percentage error rate

High and low word frequency error rates were 2.6% and 3.0% respectively. High and low syllable frequency error rates were 2.7% and 2.9% respectively. A similar analysis carried out on percentage error rate (arc sine transformed) yielded no significant effects.

Naming duration

A similar analysis was carried out on naming durations. High and low word frequency durations were 351.4 ms and 344.7 ms respectively. The 6.7 ms difference was significant over subjects, $F_1(1, 28) = 8.8$, $p < .01$, $F_2 < 1$. High and low syllable frequency durations were 326.8 ms and 369.3 ms respectively. The 42.5 ms difference was significant, $F_1(1, 28) = 253.7$, $p < .001$, $F_2(1, 48) = 15.6$, $p < .001$. Word and syllable frequency did not interact, F_1 and $F_2 < 1$.

Regression analyses

Regression analyses were carried out on the means data of the experimental words. In all regressions mean naming latency is the dependent variable. Simple regressions with both log word form frequency and log lemma frequency failed to reach significance ($R = 0.2$, $p > .05$). Of the syllable frequency counts only second syllable frequency counts yielded significant correlations: total log frequency ($R = 0.3$, $p < .01$) and position-dependent log frequency ($R = 0.4$, $p < .001$). Similarly number of phonemes in the second syllables and log second syllable CV structure frequency showed significant correlations with naming latency (both $R = 0.3$, $p < .05$). A multiple regression of naming latency with these three

²Main effects of block and trial were observed in the analyses of all the dependent variables reported. These practice effects were always due to a decrease in naming latencies, durations and error rates as the experiment progressed. In no other analysis did they significantly interact with frequency effects and they will not be reported.

second syllable variables showed only a significant unique effect of log syllable frequency ($p < .05$). This pattern of results remained when only words with initial syllable stress were included in the regressions ($n = 32$).

Discussion

Apart from the expected word frequency effect, the experiment showed that there is a syllable frequency effect as well, amounting to about 15 ms. Bisyllabic words consisting of low-frequency syllables were consistently slower in naming than those consisting of high-frequency syllables. Moreover, this syllable frequency effect was independent of the word frequency effect, as predicted by the syllabary theory.

The post hoc regression analyses suggest that second syllable frequency is a better predictor of naming latency than the frequency of first syllable. Experiments 2 and 3 will explore this possibility in more detail. Not surprisingly, syllable complexity affected word durations, but there was also some evidence that complexity of the second syllable has an effect on naming latency. This issue will be taken up in Experiment 4.

EXPERIMENT 2: FIRST AND SECOND SYLLABLE FREQUENCY

There are theoretical reasons to expect that in bisyllabic word naming the frequency of the second syllable will affect naming latency more than the frequency of the first syllable. It is known that in picture naming bisyllabic target words are produced with longer naming latencies than monosyllabic target words. In a study by Klapp, Anderson, and Berrian (1973) the difference amounted to 14 ms. The effect cannot be due to response initiation, as the difference disappears in a delayed production task where subjects can prepare their response in advance of the “Go” signal to produce it. It must therefore have its origin in phonological encoding. Levelt (1989, p. 417) suggests that if in phonetic encoding syllable programs are addressed one by one, the encoding duration of a phonological word will be a function of its syllabicity. But the crucial point here is that, apparently, the speaker cannot or will not begin to articulate the word before its phonetic encoding is complete. If articulation was initiated following the phonetic encoding of the word’s first syllable, no number-of-syllables effect should be found. Wheeldon and Lahiri (in preparation) provide further evidence that during the production of whole sentences articulation begins only when the first phonological word has been encoded.

Making the same assumption for the present case – that is, that initiation of

articulation will wait till both syllables have been accessed in the syllabary – it is natural to expect a relatively strong second syllable effect. The association process (see Fig. 1) creates phonological syllables successively. Each new syllable triggers access to the syllabary and retrieval of the corresponding phonetic syllable. Although retrieving the first syllable will be relatively slow for a low-frequency syllable, that will not become apparent in the naming latency; the response can only be initiated after the second syllable is retrieved. Retrieving the second syllable is independent of retrieving the first one. It is initiated as soon as the second syllable appears as a phonological code, whether or not the first syllable's gestural score has been retrieved. And articulation is initiated as soon as the second syllable's gestural code is available. First syllable frequency will only have an effect when retrieving that syllable gets completed only *after* retrieving the second syllable. This, however, is a most unlikely state of affairs. Syllables are spoken at a rate of about one every 200 ms. Wheeldon and Levelt (1994) have shown that phonological syllables are generated at about twice that rate, one every 100 ms. Our syllable frequency effect, however, is of the order of only 15 ms. Hence it is implausible that phonetic encoding of the second syllable can "overtake" encoding of the first one due to advantageous frequency conditions.

In this experiment we independently varied the frequency of the first and the second syllable in bisyllabic words. In one sub-experiment we did this for high-frequency words and in another one for low-frequency words.

Method

The vocabulary consisted of 96 bisyllabic Dutch nouns: 48 high word frequency, 48 low word frequency. Within each word frequency group there were four syllable frequency conditions (12 words each) constructed by crossing first syllable frequency with second syllable frequency (i.e., high–high, high–low, low–high and low–low). The criteria for assigning words to frequency groups were the same as in Experiment 1. Mean log frequencies and number of phonemes for the high- and low-frequency words in each syllable condition are given in Table 2. Two high word frequency vocabularies and the two low word frequency vocabularies were constructed, each with six words from each syllable frequency condition. Each vocabulary was then divided into six four-word groups with one word from each condition. As in Experiment 1, these groups contained words which were phonologically and semantically unrelated. Each group was assigned a symbol set with four rotations and each of 48 subjects were assigned to one vocabulary and one symbol set. Each subject received 18 blocks of 24 trials: three blocks for each word group.

In this experiment word frequency was a between-subjects variable. This was necessary because of the extra syllable frequency conditions and the limited

Table 2. *Log syllable and word frequencies and mean number of phonemes for high- and low-frequency words in each of the First × Second syllable frequency groups of Experiment 2*

Syllable freq.		No. phonemes		Syllable 1		Syllable 2		Word	
1st	2nd	Syl. 1	Syl. 2	POS	TOT	POS	TOT	WRD	LEM
<i>High word frequency</i>									
High × High		2.8	2.7	7.3	7.8	7.8	8.7	3.8	4.0
High × Low		2.8	3.2	7.6	7.9	5.1	5.3	3.6	4.0
Low × High		3.1	2.8	4.9	5.2	8.3	8.9	3.8	4.0
Low × Low		3.0	3.6	4.9	5.3	4.8	5.2	3.7	4.0
<i>Low word frequency</i>									
High × High		2.8	2.6	7.0	7.5	8.1	8.7	1.5	1.9
High × Low		2.7	3.2	7.5	8.1	4.0	4.5	1.2	1.5
Low × High		3.1	2.6	4.1	4.7	8.3	8.9	1.4	1.7
Low × Low		2.9	3.3	4.3	4.7	4.1	4.5	1.0	1.5

number of words a subject could accurately memorize and produce within an hour. Moreover, our major interest was in the pattern of results over the syllable frequency conditions for both high- and low-frequency words, rather than in the word frequency effect itself. In order to be able to compare baseline naming speed of subjects who received the high and low word frequency vocabularies, each subject received a calibration block of the same four words at the end of the experiment.

The rest of the procedure was exactly the same as in Experiment 1. Forty-eight subjects were run; 24 received a high word frequency vocabulary (20 women and 4 men) and 24 received a low word frequency vocabulary (18 women and 6 men).

Results

Exclusion of data

Data from four subjects were replaced due to high error rates. Data points were excluded and substituted according to the same principles as in Experiment 1. The first production of a word in each block was again counted a practice trial and excluded from the analysis. 2.8% of data points were correct naming latencies following error trials. 1.8% of the data points were greater than two standard deviations from the mean.

Naming latency

Mean naming latency for the high word frequency group was 641.6 ms – 5.7 ms

Table 3. Mean naming latency and percentage error (in parentheses) for words in the four syllable frequency conditions of Experiment 2. Means are shown for all words and for high- and low-frequency words separately. The effect of syllable frequency (low minus high) is also shown.

	Syllable frequency		Effect Low – high
	Low	High	
<i>All words</i>			
1st syllable	637.4 (1.9)	640.1 (2.1)	–2.7 (–0.2)
2nd syllable	644.5 (2.3)	633.0 (0.2)	11.5 (0.5)
<i>High-frequency words</i>			
1st syllable	641.8 (2.0)	641.1 (2.4)	0.7 (–0.4)
2nd syllable	646.5 (2.5)	636.5 (2.0)	10.0 (0.5)
<i>Low-frequency words</i>			
1st syllable	632.8 (1.8)	638.9 (1.8)	–6.1 (0.0)
2nd syllable	642.3 (2.1)	629.3 (1.5)	13.0 (0.6)

slower than the low word frequency group, 635.9 ms (see Table 3). This reverse effect of word frequency was insignificant, F_1 and $F_2 < 1$, and can be attributed to the random assignment of slower subjects to the high-frequency vocabularies. Mean naming latencies for the calibration block were: high word frequency, 659.3 ms; low word frequency, 624.5 ms. Subjects who received the high word frequency vocabularies were, therefore, on average 34.8 ms slower than the subjects who received the low word frequency vocabularies. This difference was also significant by words, $F_1(1, 46) = 1.9$, $F_2(1, 3) = 52.1$, $p < .01$.

Mean naming latencies and error rates for the syllable frequency conditions are shown in Table 3; the latency data are summarized in Fig. 3. The –2.7 ms effect of first syllable frequency was, unsurprisingly, insignificant, $F_1(1, 44) = 1.1$, $F_2 < 1$. The 11.5 ms effect of second syllable frequency was significant by subjects, $F_1(1, 44) = 18.6$, $p < .001$, and again marginally significant by words, $F_2(1, 80) = 3.8$, $p = .053$. The interaction of first and second syllable frequency was not significant, F_1 and $F_2 < 1$. However, there was a significant three-way word frequency by first and second syllable frequency interaction, but only in the subject analysis, $F_1(1, 44) = 6.1$, $p < .05$, $F_2(1, 80) = 1.3$. This was due to a by-subjects only interaction of first and second syllable frequency in the low-frequency word set, $F_1(1, 22) = 5.6$, $p < .05$, $F_2(1, 40) = 1.2$; words with high-frequency first syllables showed a smaller effect of second syllable frequency than words with low-frequency first syllables (5 ms and 21 ms respectively). Words with high-frequency second syllables showed a reverse effect of first syllable frequency (–14 ms) compared to a 2 ms effect for words with low-frequency second syllables.

There was no main effect of vocabulary, F_1 and $F_2 < 1$. However, there was a

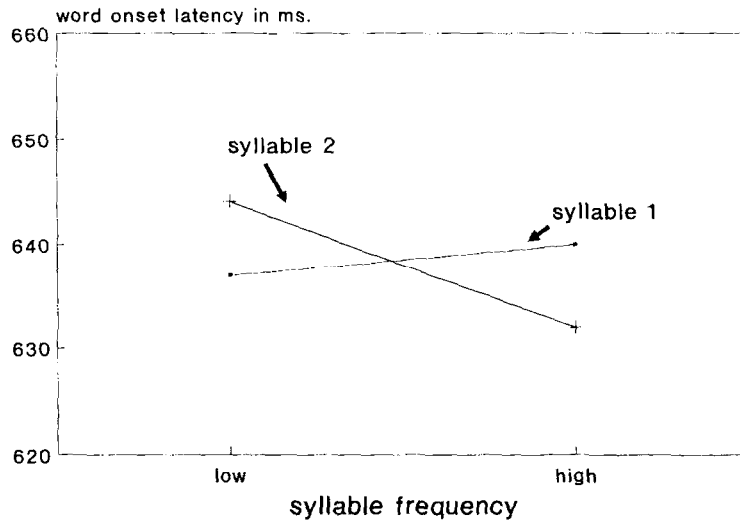


Figure 3. Naming latencies in Experiment 2. Syllable position (word-initial, word-final) versus syllable frequency.

significant interaction of second syllable frequency with vocabulary in the by-subject analysis, $F_1(1, 44) = 6.8$, $p < .05$, $F_2(1, 80) = 1.4$, due to differences in the size of the effect in the two vocabularies in both the high- and low-frequency word sets.

Naming duration

Naming durations for high- and low-frequency words were 346.8 ms and 316.6 ms respectively. The 50.2 ms effect was significant by words, $F_1(1, 44) = 3.5$, $p > .05$, $F_1(1, 80) = 20.1$, $p < .001$. There were also significant effects of first syllable frequency (high 329.1 ms, low 334.3 ms, $F_2(1, 44) = 12.7$, $p > .01$, $F_2 < 1$) and second syllable frequency (high 321.1 ms, low 342.3 ms, $F_1(1, 44) = 167.0$, $p > .001$, $F_2(1, 80) = 9.8$, $p < .01$). The interaction of first and second syllable frequency was only significant by subjects, $F_1(1, 44) = 12.0$, $p > .001$, $F_2 < 1$; the effect of frequency on second syllable durations was restricted to words with high first syllable frequencies.

Percentage error rate

Error rates are also shown in Table 3. They yielded only a significant effect of second syllable frequency over subjects, $F_1(1, 44) = 6.0$, $p < .05$, $F_2(1, 80) = 2.4$.

Discussion

Although not all vocabularies in this experiment yielded significant syllable frequency effects, the main findings were consistent with our expectations. Whatever there is in terms of syllable frequency effects was due to the second syllable only. The frequency of the first syllable had no effect on naming latencies. Although the average size of the frequency effect (12 ms) was of the order of magnitude obtained in Experiment 1 (15 ms), the complexity of the experiment apparently attenuated its statistical saliency.

An interaction of first and second syllable frequency effects is not predicted by our model of syllable retrieval. This experiment did yield some indication of such an interaction. However, it was observed in one vocabulary only and never approached significance over items. While further investigation is necessary to rule out such an effect, we do not feel it necessary to amend our model on the basis of this result.

The next experiment was designed to isolate the effect of second syllable frequency.

EXPERIMENT 3: SECOND SYLLABLE FREQUENCY

Method

Vocabulary

The experimental vocabulary consisted of 24 pairs of bisyllabic Dutch words. Members of a pair had identical first syllables but differed in their second syllable: one word has a high-frequency second syllable and one word had a low-frequency second syllable (e.g., *ha-mer/ha-vik*). High and low second syllable frequency words were matched for word frequency. No attempt was made to match second syllables for number of phonemes (see Table 4). Two matched vocabularies of 12 word pairs were constructed.

Design

Twelve pairs of abstract symbols of the form used in Experiment 1 were constructed. Each pair consisted of one simple symbol (e.g., -----) and one more complex symbol (e.g., } } } } }). The symbol pairs were assigned to one word pair in each vocabulary. Two sets for each vocabulary were constructed such that each word in a word pair was assigned to each symbol in its associated pair once.

Table 4. *Log syllable and word frequencies for high- and low-frequency second syllable words in Experiment 4*

	2nd syllable frequency	
	High	Low
<i>Log frequency</i>		
Word form	1.9	2.0
Lemma	2.1	2.2
1st syllable position dependent	6.8	6.8
1st syllable total	7.2	7.2
2nd syllable position dependent	7.8	4.0
2nd syllable total	8.7	4.7
Number of phonemes	2.8	3.3

Within a vocabulary, words were grouped into six blocks of four words. Only one member of a word pair occurred within a block. None of the words within a block had the same initial phoneme and they were semantically unrelated. The associated symbol groups in each set were the same in each vocabulary. Each subject was assigned randomly to a vocabulary and a word set. Each subject received 24 blocks of 24 trials: three blocks for each word group. Presentation of the blocks within a set was rotated.

Procedure and subjects

Each subject was assigned randomly to a vocabulary and a word set. Presentation of the blocks within a set were rotated. The procedure was the same as in Experiments 1 and 2. Twenty-four subjects were tested: 18 women and 6 men.

Results

Exclusion of data

2.2% of the data were trials following an error and 1.8% of the data were greater than 2 standard deviations from the mean. These data were again excluded from the analyses.

Naming latencies

Mean naming latency for words with high-frequency second syllables was

622.7 ms, and for low-frequency second syllable 634.5 ms. The 11.8 ms effect of syllable frequency was significant, $F_1(1, 22) = 12.6$, $p < .01$, $F_2(1, 44) = 4.7$, $p < .05$.

There was a main effect of vocabulary by words, $F_1 < 1$, $F_2(1, 44) = 18.0$, $p < .001$, due to slower reaction times to vocabulary A (640.1 ms) compared to vocabulary B (617.1 ms). There was also a significant interaction between syllable frequency and vocabulary by subjects only, $F_1(1, 22) = 4.5$, $p < .05$, $F_2(1, 44) = 1.7$, due to a larger frequency effect in vocabulary A (high 630.7 ms, low 649.5 ms) than in vocabulary B (high 614.8 ms, low 619.5 ms).

Naming durations

Mean naming duration for words with high-frequency second syllables was 351.5 ms, and for low-frequency second syllable 370.0 ms. The 18.5 ms difference was significant, $F_1(1, 22) = 106.0$, $p < .001$, $F_2(1, 44) = 4.5$, $p < .05$.

The effect of vocabulary was significant by words, $F_1(1, 22) = 2.8$, $F_2(1, 44) = 26.0$, $p < .001$ (vocabulary A, 338.4 ms, vocabulary B 383.0 ms), but there was no interaction of vocabulary with syllable frequency, $F_1(1, 22) = 3.1$, $F_2 < 1$.

Percentage error rate

Mean percentage error rates were, for high-frequency second syllable 1.2%, and for low-frequency second syllable 1.6%. The only significant effect was of vocabulary (vocabulary A 1.8%, vocabulary B 1.0%), $F_1(1, 22) = 5.2$, $p < .05$, $F_2(1, 44) = 5.1$, $p < .05$.

Discussion

The present experiment reproduced the 12 ms second syllable effect obtained in Experiment 2, but now with satisfying statistical reliability. Together with the previous experiments, it supports the notion that the bulk, if not the whole of the syllable frequency effect, is due to the word-final syllable.

Let us now turn to the other issue raised in the discussion of Experiment 1. Could it be the case that what we are measuring is not so much an effect of syllable frequency, but rather one of syllable complexity? In all of the previous experiments the second syllable frequency effect on naming latencies is accompanied by a similar effect on naming durations; that is, words with low-frequency second syllables have significantly longer naming durations than words with high-frequency second syllables. Moreover, the regression analyses of Experiment

1 showed that a syllable's frequency of occurrence correlates with the number of phonemes it contains. It is possible, therefore, that syllable complexity (defined in terms of number of phonemes to be encoded or in terms of articulation time) underlies the effects we have observed.

EXPERIMENT 4: SYLLABLE COMPLEXITY

The complexity issue is a rather crucial one. In the theoretical section of this paper we compared a direct route in phonetic encoding and a route via stored syllable programs. If any of these, the former but not the latter would predict an effect of syllable complexity. The more complex a syllable's phonological structure, the more computation would be involved in generating its gestural score afresh from its phonological specifications. But no such thing is expected on the syllabary account. The syllabic gesture need not be composed; it is only retrieved. There is no reason to suppose that retrieving a more complex gestural score takes more time than retrieving a simpler one. There will, at most, be a mediated relation to complexity. There is a general tendency for more complex syllables to be less frequent in usage than simpler syllables. If indeed frequency is a determinant of accessing speed, then – even on the syllabary account – simple syllables will be faster than complex syllables.

The present experiment was designed to test second syllable complexity as a potential determinant of phonetic encoding latency, but we controlled for syllable frequency in order to avoid the aforementioned confounding. We also controlled for word frequency.

Method

Vocabulary

The vocabulary consisted of 20 pairs of bisyllabic nouns. Each pair of words had the same initial syllable but differed in the number of phonemes in their second syllable (e.g., *ge-mis* [CVC]; *ge-schreeuw* [CCCVVC]). Word pairs were also matched for word and syllable frequency (see Table 5). The 20 pairs were divided into two vocabularies of 10 pairs matched on all the above variables.

Design

As in Experiment 3, pairs of abstract symbols were constructed and assigned to one word pair in each vocabulary. Two sets for each vocabulary were again

Table 5. *Log syllable and word frequencies and mean number of phonemes for short and long words in Experiment 4*

	2nd syllable	
	Short	Long
<i>Log frequency</i>		
Word form	1.9	2.0
Lemma	2.3	2.4
1st syllable position dependent	9.3	9.3
1st syllable total	9.4	9.4
2nd syllable position dependent	3.7	3.6
2nd syllable total	5.0	5.3
Number of phonemes	3	5

constructed such that each word in a word pair was assigned to each symbol in its associated pair once. Each vocabulary consisted of five blocks of four words. The rest of the design was the same as in Experiment 3, except that each subject received 15 blocks of 24 trials: three blocks for each word group.

Procedure and subjects

Each subject was again assigned randomly to a vocabulary and a word set. Presentation of the blocks within a set were rotated. The procedure was the same as in Experiments 1 and 2. Twenty subjects were tested: 13 women and 7 men.

Results

Exclusion of data

Two subjects were replaced due to high error rates. Exclusion of data resulted in the loss of 5.6% of the data: 4.1% were trials following an error and 1.5% were outliers.

Analyses

Naming latencies and percentage error rates were, for simple words, 681.3 ms (4.2%), and for complex words 678.7 ms (3.3%). The effect of complexity on naming latency was insignificant, F_1 and $F_2 < 1$, as was the effect on error rates,

$F_1 = 1.0$, $F_2 = 1.5$. Clearly, the complexity (number of phonemes) of a word's second syllable does not affect its naming latency.

Mean word duration for the simple words was 270.0 ms, compared to 313.0 for the complex words; this difference was significant, $F_1(1, 18) = 99.5$, $p < .0001$, $F_2(1, 36) = 15.5$, $p < .001$.

Discussion

When syllable frequency is controlled for, second syllable complexity does not affect naming latency. This shows that complexity cannot be an explanation for the syllable frequency effect obtained in the previous three experiments. In addition, the lack of a complexity effect shows that either the direct route in phonetic encoding (see above) is not a (co-)determinant of naming latencies in these experiments, or that the computational duration of gestural scores is, in some way, not complexity dependent.

GENERAL DISCUSSION

The main findings of the four experiments reported are these: (i) syllable frequency affects naming latency in bisyllabic words; (ii) the effect is independent of word frequency; (iii) the effect is due to the frequency of the word's ultimate syllable; (iv) second syllable complexity does not affect naming latency, and hence cannot be the cause of the frequency effect.

What are the theoretical consequences of these findings? We will first consider this issue with respect to the theoretical framework of phonological encoding sketched above. We will then turn to alternative accounts that may be worth exploring.

The syllabary theory reconsidered

It needs no further discussion that the experimental findings are in seamless agreement with the syllabary theory as developed above. In fact, no other theory of phonological encoding ever predicted the non-trivial finding that word and syllable frequency have additive effects on naming latency. The theory, moreover, provides natural accounts of the dominant rule of the word-final syllable and one of the absence of a syllable complexity effect. These explanations hinge on the theoretical assumption that syllabification is a late process in phonological encoding (in particular that there is no syllabification in the word form lexicon) and that gestural scores for syllables are retrieved as whole entities.

It is, however, not the case that the findings are also directly supportive for other aspects of the theory, such as the details of segmental and metrical spellout, the metrical character of phonological word formation and the particulars of segment-to-frame association (except for the assumption that this proceeds on a syllable-by-syllable basis). These aspects require their own independent justification (for some of which see Levelt, 1989, 1993). But there is one issue in phonological encoding that may appear in a new light, given this framework and the present results. It is the issue of underspecification.

As pointed out above, Stemberger (1983) was amongst the first to argue for underspecification in a theory of phonological encoding. It could provide a natural account for speech errors such as *in your really gruffy – scruffy clothes*. Here the voicelessness of /k/ in *scruffy* is redundant. The lexicon might specify no more than the “archiphoneme” /K/, which can have both [k] and [g] as phonetic realizations; that is, the segment is unspecified on the voicing dimension. In the context of /s-r/, however, the realization has to be voiceless. But when, in a slip, the /s/ gets chopped off, the context disappears, and /K/ may become realized as [g]. The notion of underspecification was independently developed in phonological theory. Archangeli (1988) in particular proposed a theory of “radical underspecification”, which claims that only unpredictable features are specified in the lexicon.

But a major problem for any underspecification theory is how a full specification gets computed from the underspecified base. The solutions need not be the same for a structural phonological theory and for a process theory of phonological encoding. Here we are only concerned with the latter, but the proposed solution may still be of some relevance to phonological theory.

The syllabary theory may handle the completion problem in the following way. There is no need to complete the specifications of successive segments in a word if one condition is met. It is that each phonological syllable arising in the process of segment-to-frame association (see Fig. 1) corresponds to one and only one gestural score in the syllabary. In other words, even if a syllable’s segments are underspecified, their combination can still be unique.

This condition puts empirical constraints on the degree and character of underspecification. Given a theory of underspecification, one can determine whether uniqueness is preserved, that is, whether each phonological syllable that can arise in phonological encoding corresponds to only one phonetic syllable in the syllabary. Or in other words, the domain of radical redundancy should be the syllable, not any other linguistic unit (such as the lexical word). Moreover, the domain should not be *potential* syllables, but syllables that occur with sufficient frequency in the speaker’s language use as to have become “overlearned”. Different cut-off frequency criteria should be considered here. Another variant would be to limit the domain to core syllables, ignoring syllable suffixes (see below).

The syllabary theory is, of course, not complete without a precise characterization of how the syllabary is accessed, given a phonological syllable. What we have said so far (following Crompton, 1982, and Levelt, 1989) is that a syllable gesture is selected and retrieved as soon as its phonological specification is complete. In a network model (such as in Roelofs, 1992, or Levelt, 1992, but also *mutatis mutandis* in Dell's, 1988, model), this would require the addition of a bottom layer of phonetic syllable nodes. A syllable node's frequency-dependent accessibility can then be modelled as its resting activation.

A strict regime has to be built in, in order to select phonetic syllables in their correct order, that is, strictly following a phonological word's segment-to-frame association. Although a word's second syllable node may become activated before the first syllable has been selected, selection of syllable one must precede selection of syllable two (and so on for subsequent syllables). Unlike phonological encoding, which involves the slightly error-prone process of assigning activated phonemes to particular positions in a phonological word frame, there are no frames to be filled in phonetic encoding. It merely involves the concatenation of successively retrieved syllabic gestures. This difference accounts for the fact that exchanges of whole syllables are almost never observed. Modelling work along these lines is in progress. The successive selection of articulatory gestures does not exclude a certain overlap in their motor execution. Whatever there is in between-syllable coarticulation may be due to such overlap. The articulatory network probably computes an articulatory gesture that is a weighted average of the two target gestures in the range of overlap.

Alternative accounts

Let us now turn to possible alternative accounts of our data. They can best be cast as ranging over a dimension of "mixed models", which includes our own. The one extreme here is that all phonological encoding involves access to a syllabary. The other extreme is that a phonological word's and its syllables' gestural scores are always fully computed. Our own syllabary theory, as proposed above, is a mixed model in that we assume the computability of all syllables – new, low or high frequency. But there is always a race between full computation and access to stored syllable scores, where the latter process will normally win the race except for very low-frequency or new syllables. Hence, our theory predicts that there *should* be a syllable complexity effect for words that end on new or very low-frequency syllables.

But the balance between computation and retrieval may be a different one. More computation will be involved when one assumes that only core syllables are stored, whereas syllable suffixes are always computed. What is a core syllable? One definition is that it is a syllable that obeys the *sonority sequencing principle*.

This states that syllable-initial segments should be monotonically increasing in sonority towards the syllable nucleus (usually the vowel), and that syllable-final segments should be monotonically decreasing from the nucleus (see Clements, 1990, for a historical and systematic review of “sonority sequencing”). Phonetically a segment’s sonority is its perceptibility, vowels being more sonorant than consonants, nasals being more sonorant than stops, etc. But sonority can also be defined in terms of phonological principles (Clements, 1990). On either of these sonority accounts the syllable /plant/ is a core syllable, whereas /lpatn/ is not; the latter violates the sequencing principle both in its onset and its offset. Though /lpatn/ is not a syllable of English, violations of sonority sequencing do occur in English syllables, as in *cats*, *task* or *apt*.

Fujimura and Lovins (1978) proposed to treat such and similar cases as combinations of a core syllable plus an “affix”, such as /c t + s/, etc. Here the core obeys sonority sequencing, and the affix is added to it. The authors also gave other, more phonological reasons for distinguishing between core and affixes, not involving sonority. They proposed that English syllables can have only one place-specifying consonant following the nucleus. So, in a word like *lens*, *s* is a suffix, although the sonority principle is not violated here. A similar notion of “syllable appendix” was proposed by Halle and Vergnaud (1980).

It is clear where such affixes can arise in the process of segment-to-frame association discussed earlier. This will most naturally occur in word-final position when there is a “left over” consonantal segment that cannot associate to a following syllable (Rule 2b). The present version of a mixed theory would then be that as soon as a phonological core syllable is created in left-to-right segment-to-frame association, its phonetic score is retrieved from the syllabary. Any affixes will be computationally added to that score.

An advantage of this theory is that the syllabary will drastically reduce in size. In the CELEX database for English (i.e., for citation forms of words) there are about 12,000 different syllables (counting both full and reduced syllables). But most of them have complex offset clusters. These will all be eliminated in a core syllabary.

But a disadvantage is that the theory predicts the complexity effect that we didn’t find in Experiment 4. There we varied syllables’ complexity precisely by varying the number of segments in their consonant clusters (onset or coda), and this should have computational consequences on the present theory. Still, the experiment was not explicitly designed to test the affix theory; it is therefore premature to reject it without further experimentation.

Where Fujimura and Lovins (1978) only proposed to distinguish between syllable core and affix(es), Fujimura (1979) went a step further, namely to split up the core as well. In order to account for the different types of vowel affinity of the initial and final parts of the syllable (already observed in the earlier paper) he introduced the notion of *demissyllable*. The syllable core consists of an initial

demisyllable consisting of initial consonant(s) plus vowel, and a final demisyllable consisting of vowel plus following consonants. Hence, these demisyllables hinge at the syllabic nucleus. In this model, demisyllables are the domains of allophonic variation, of sonority and other relations between consonants and the vowels they attach to. Or more precisely, as Fujimura (1990) puts it, demisyllables, not phonemes, are the “minimal integral units”. Consonantal features are, in actuality, features of demisyllables.

On this account “the complete inventory for segmental concatenation will contain at most 1000 entries and still reproduce natural allophonic variation” (Fujimura, 1976). We could call this inventory a *demisyllabary*, and we have another mixed model here. The speaker might access such a demisyllabary and retrieve syllable-initial and syllable-final gestures or gestural scores. Fujimura’s model requires that, in addition, further computation of syllable affixes should be necessary.

This latter part of the model will, of course, create the same complexity problem as discussed above. But as far as the demisyllable aspect is concerned, we can see no convincing arguments to reject such a model on the basis of our present results. It cannot be excluded *a priori* that our syllable frequency effect is, in actuality, a demisyllable frequency effect. In order to test this, new experiments will have to be designed, where demisyllable frequency is systematically varied.

In conclusion, although we have certainly not yet proven that speakers do have access to a syllabary, our theory has been productive in making non-trivial predictions that found support in a series of experiments. Any alternative theory should be able to account for the syllable frequency effect, its independence of word frequency, and the absence of syllable complexity effects.

References

- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.
- Browman, C.P., & Goldstein, L. (1991). Representation and reality: Physical systems and phonological structure. *Haskins Laboratory Status Report on Speech Research*, SR-105/106, 83–92.
- Clements, G.N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M.E. Beckman (Eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech* (pp. 58–71). Cambridge, UK: Cambridge University Press.
- Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109–162). Berlin: Mouton.
- 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.
- Fujimura, O. (1976). Syllables as concatenated demisyllables and affixes. *Journal of the Acoustical Society of America*, 59 (Suppl. 1), S55.
- Fujimura, O. (1979). An analysis of English syllables as cores and suffixes. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 32, 471–476.

- Fujimura, O. (1990). Demisyllables as sets of features: Comments on Clement's paper. In J. Kingston & M.E. Beckman (Eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech* (pp. 377–381). Cambridge, UK: Cambridge University Press.
- Fujimura, O., & Lovins, J.B. (1978). Syllables as concatenative phonetic units. In A. Bell & J.B. Hooper (Eds.), *Syllables and segments* (pp. 107–120). Amsterdam: North-Holland.
- Halle, M., & Vergnaud, J.-R. (1980). Three dimensional phonology. *Journal of Linguistic Research*, 1, 83–105.
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20, 253–306.
- Jescheniak, J.D., & Levelt, W.J.M. (in press). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: LMC*, 20.
- Klapp, S.T., Anderson, W.G., & Berrian, R.W. (1973). Implicit speech in reading, reconsidered. *Journal of Experimental Psychology*, 100, 368–374.
- Keating (1988). Underspecification in phonetics. *Phonology*, 5, 275–292.
- 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. (1993). Timing in speech production: With special reference to word form encoding. *Annals of the New York Academy of Sciences*, 682, 283–295.
- Lindblom, B. (1983). Economy of speech gestures. In P.F. MacNeilage (Ed.), *The production of speech* (pp. 217–245). New York: Springer.
- Mackay, D. (1982). The problems of flexibility, fluency and speed–accuracy tradeoff. *Psychological Review*, 89, 483–506.
- Meringer, R., & Mayer, K. (1985). *Versprechen und Verlesen*. Stuttgart: Goschensche Verlag. (Reissued, with introductory essay by A. Cutler and D.A. Fay, 1978. Amsterdam: John Benjamins.)
- Meyer, A.S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, 29, 524–545.
- Meyer, A.S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.
- Meyer, A.S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1146–1160.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273–281.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Saltzman, E., & Kelso, J.A.S. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, 94, 84–106.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42, 213–259.
- Stemberger, J.P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana Linguistics Club.
- Wheeldon, L.R., & Lahiri, A. (in preparation). Prosodic units in language production.
- Wheeldon, L., & Levelt, W.J.M. (1994). Monitoring the time-course of phonological encoding. Manuscript submitted for publication.
- Winer, B.J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, 81, 226–234.

Appendix 1. Vocabularies in Experiment 1

The four experimental vocabularies split into blockgroups containing one word from each of the four frequency groups. Within a blockgroup, words are phonologically and semantically unrelated.

	VOCAB A	VOCAB B	VOCAB C	VOCAB D
GROUP 1	constant (HH)	nadeel (HH)	geding (HH)	roman (HH)
	neutraal (HH)	gordijn (LH)	triomf (LH)	borrel (LH)
	cider (HL)	takel (HL)	kakel (HL)	hoeder (HL)
	tarbot (LL)	concaaf (LL)	neuraal (LL)	soldeer (LL)
GROUP 2	arme (HH)	koning (HH)	stilte (HH)	versie (HH)
	client (LH)	sleutel (LH)	rapport (LH)	praktijk (LH)
	nader (HL)	volte (HL)	bever (HL)	neder (HL)
	vijzel (LL)	absint (LL)	horzel (LL)	causaal (LL)
GROUP 3	boter (HH)	toren (HH)	natuur (HH)	teder (HH)
	heuvel (LH)	nerveus (LH)	gratis (LH)	advies (LH)
	kandeel (HL)	gemaal (HL)	proper (HL)	combo (HL)
	giraffe (LL)	berber (LL)	concours (LL)	geiser (LL)
GROUP 4	pater (HH)	heelal (HH)	kussen (HH)	gebaar (HH)
	techniek (LH)	crisis (LH)	vijand (LH)	kasteel (LH)
	gewei (HL)	reiger (HL)	adder (HL)	tegel (HL)
	rantsoen (LL)	pingel (LL)	trofee (LL)	narcis (LL)