SYLLABLES IN SPEECH PRODUCTION:
EFFECTS OF SYLLABLE PREPARATION
AND SYLLABLE FREQUENCY

ISBN: 90-76203-15-6

Cover design: Linda van den Akker & Inge Doehring
Cover illustration: Petra & Ottoline van Alphen, Joana Cholin & Kerstin Mauth
Printed and bound by Ponsen & Looijen bv, Wageningen

SYLLABLES IN SPEECH PRODUCTION: EFFECTS OF SYLLABLE PREPARATION AND SYLLABLE FREQUENCY

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
op gezag van de Rector Magnificus Prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op woensdag 16 juni 2004
des namiddags om 1.30 uur precies

door
Joana Cholin
geboren op 4 juli 1972
te Cagliari, Italïe

Promotor:	Prof. dr. Willem J. M. Levelt
Co-promotor:	Dr. Niels O. Schiller

Manuscriptcommissie:	Prof. dr. Herbert Schriefers Prof. dr. Antje S. Meyer (University of Birmingham, UK) Prof. dr. Rob Schreuder

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

für A. und J.

ACKNOWLEDGMENTS

During my PhD-time at the MPI, I remember that I thought several times "That person definitely deserves a big THANK YOU in the acknowledgments of my dissertation". Now it's time to write the acknowledgments and I hope I remember everybody who helped and supported me. First of all, I would like to thank Pim Levelt who was a reliable and untiring supervisor in the search for the "Mental Syllabary". Pim once said to me that writing a dissertation is not only mentally exhausting but also emotionally. I now can confirm that this is certainly the case and I would like to thank him for his support in both respects. Furthermore, I would like to thank Niels Schiller for his contributions to this thesis and for the good discussions we had.

I truly thank Antje Meyer and James McQueen for being there at the right moments. A warm thank you goes to Alissa Melinger for always being supportive and patient. Alissa was never too tired to read another version of various chapters and articles and she always gave useful comments without taking things over. I would like to thank the members and guests of the speech production group for listening to the many presentations of my frequency experiments and for their helpful comments. Ardi Roelofs was always happy to discuss WEAVER++'s predictions regarding the mental syllabary and his suggestions were always very useful. A big thank you goes to Evelyn Giering for making a lot of things easier. Simone Sprenger helped me especially to get started with all experimental matters and more. She was a great office mate and became a really good friend. Fermín Moscoso del Prado Martín, Harald Baayen, Frauke Hellwig, and Dirk Janssen helped me a lot with CELEX and AWK related things and always had fruitful ideas. Petra van Alphen translated the Dutch summary at the end of this thesis and spoke splendidly all of the syllable material (van 'bin' tot 'bing'...); and I also would like to thank all the participants in my experiments who managed to reproduce these 'bins' and 'bings'!

The support of the technical group was invaluable. John Nagengast's help with the 'little-loudspeaker'-method and his energy and patience with all the complicated EAT-

files was fantastic. Johan Weustink, Ad Verbunt, Rick van Viersen, and Reiner Dirksmeyer were always there whenever I needed technical material or help. Inge Doehring and Frauke helped me to prepare very professional posters. I would like to thank everybody from the administration for their unbureaucratic way of dealing with things. Thanks to Karin Kastens and the library team for getting all the articles I needed. Moreover, I thank Agnes de Boer and José Wannet, Pim Folten, Jan Achterberg, and Hans Gerrits for always being friendly and cheerful.

The students and post-docs from the MPI were great colleagues and also good friends. At first, I would like to mention Rasha Abdel Rahman, Rebecca Gross, Heidrun Bien and the other production girls with whom I had a lot of chats, coffees (& Wiener Mélange!), evenings, and parties. You all made sure that I did not forget that life is more than syllables, thank you. Mandana Seyfeddinipur participated happily in my piecemeal acquisition of phonological and phonetic encoding processes. As house mates and friends we shared a lot of the daily good and bad times and we had many lively discussions about syllables, gestures, and life in general. I would like to thank Keren Shatzman and Petra for their interest and help in working on syllables and for their friendship. With Fermín, Nivja de Jong, and Rachèl Kemps I spent the first LOT winter school in Amsterdam and I also really enjoyed the company of Rachèl and Mirjam Ernestus during the LabPhon8. With Pamela Perniss, Frank Seifart, and all the others I had a lot of nice Ph-beers, lunches, and ‘feestjes’. I would especially like to thank Britta Stolterfoth, Monika Lück, Angelika Wolf, and Barbara Wurm for the good time in Leipzig. I also owe a big thank you to my “old” friends and also to my former colleagues of the B7- and later C7-project in Düsseldorf. Kerstin Mauth had already been a good friend of mine before I started working at the MPI and she became an even better one during my time there. I thank Kerstin and Petra for being my ‘paranimfen’!

I would like to thank my parents, Ada and Jochen, and my brother Jochen for their support and trust over all these years.
Special thanks and more to Ansgar!

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION 1

A MODEL OF WORD FORM ENCODING 2

OUTLINE OF THE THESIS 5

CHAPTER 2

THE PREPARATION OF SYLLABLES IN SPEECH PRODUCTION 7

ABSTRACT 7

INTRODUCTION 8

Phonological and phonetic encoding in WEAVER++ 9

Priming studies of syllable access 12

Syllable frequency studies 13

The WEAVER++ predictions in more detail 13

The implicit priming paradigm 16

EXPERIMENT 2.1: PRODUCTION OF CVV TARGETS 18

Method 18

Results 21

Discussion 23

EXPERIMENT 2.2: PRODUCTION OF CCVV TARGETS 23

Method 24

Results 24

Discussion 25

GENERAL DISCUSSION 29

CONCLUSIONS 33

CHAPTER 3

EFFECTS OF SYLLABLE FREQUENCY IN SPEECH PRODUCTION **35**

ABSTRACT 35

INTRODUCTION 36

PHONOLOGICAL AND PHONETIC ENCODING 37

Predictions by WEAVER++ 40

Empirical evidence for syllabic units in speech production 41

Evidence for the mental syllabary and syllable frequency studies 43

Symbol - position association learning task 46

PRE-TEST INVESTIGATING THE PRODUCTION OF HIGH- AND LOW-FREQUENCY WORDS IN THE SYMBOL-POSITION ASSOCIATION LEARNING TASK 47

Method 47

Results 49

Discussion 50

EXPERIMENT 3.1: MONOSYLLABIC PSEUDO-WORDS 50

Method 51

Results 55

Discussion 56

EXPERIMENT 3.2: MONOSYLLABIC PSEUDO-WORDS 57

Method 57

Results 60

Discussion 61

EXPERIMENT 3.3: DISYLLABIC PSEUDO-WORDS FREQUENCY-MANIPULATION ON THE SECOND SYLLABLE 62

Method 63

Results 64

Discussion 64

EXPERIMENT 3.4: DISYLLABIC PSEUDO-WORDS FREQUENCY-MANIPULATION ON THE FIRST SYLLABLE 66

Method 66

Results 67

Discussion 68

GENERAL DISCUSSION 68

CONCLUSIONS 73

CHAPTER 4

EFFECTS OF SYLLABLE PREPARATION AND SYLLABLE FREQUENCY: A COMBINED STUDY **75**

ABSTRACT 75

INTRODUCTION 76

- The interface of phonological and phonetic encoding 77
- Phonological encoding 77
- Access to the mental syllabary and phonetic encoding 79
- WEAVER++ 80
- Syllable priming studies 81
- Effects of syllable frequency 84

EXPERIMENT 4.1: THE IMPLICIT PRIMING TASK AS A READING STUDY WITH FREQUENCY-MANIPULATED DISYLLABIC PSEUDO-WORDS 86

- Method 87
- Results and discussion 91

SUMMARY AND CONCLUSIONS 93

CHAPTER 5

GENERAL DISCUSSION **97**

SUMMARY OF RESULTS 98

IMPLICATIONS FOR A MODEL OF SPEECH PRODUCTION 101

FURTHER RESEARCH 103

REFERENCES **107**

APPENDICES **119**

SUMMARY **131**

SAMENVATTING **137**

ZUSAMMENFASSUNG **143**

CURRICULUM VITAE **148**

INTRODUCTION

CHAPTER 1

The output of spoken language is a continuous acoustic string, which unfolds incrementally over time without clear pauses, boundaries or delimitation. The acoustic output is not composed of discrete units that can be easily identified from spectrographic investigation. Even whole words do rarely have clear separations between them in the acoustic string. However, this acoustic output is the end result of a series of processes that do operate on discrete units of representation, such as words, morphemes, phonemes, and perhaps *syllables*.

What are syllables?

The syllable has proven difficult to define and in turn it has also been difficult to find strong evidence for its relevance to psycholinguistic processes of speaking and listening. Minimally, the syllable consists of a short string of segments with a single peak of sonority, usually a vowel, optionally flanked by one or more consonants. Phonologically, the syllable is organized into onset, nucleus, and coda, and these syllabic positions then play important roles in identifying domains for various phonological processes such as allophonic variation, coarticulation, assimilation and stress or accent assignment (see Blevins, 1995; Fujimura & Lovins, 1978; Hooper, 1972; Kenstowicz, 1994; Selkirk, 1982). According to Booij (1995) the syllable (a) is the most important domain of phonotactic restrictions, that is, a crucial domain for constraints on the co-occurrence of segments, (b) functions as the domain of certain phonological rules, and (c) functions as the bearer of stress properties. Phonetically, the syllable has been viewed as the minimal planning unit for articulation, possibly because it typically consists of a single opening and closing of the vocal tract (McNeilage, 1998) and/ or a single peak of prominence resulting from a combination of stress, pitch, length, and intrinsic sonority (Ladefoged, 1982).

In addition to the linguistic internal arguments for the syllable, there is also some off-line external evidence to support its central role in linguistic processes. First, while naïve speakers generally have shaky intuitions about linguistic structures, they tend to have good and accurate intuitions about the number of syllables in a given word (Liberman, Shankweiler, Fischer, & Carter, 1974). This suggests that syllables have more than just an abstract theoretical status relevant to phonological processes. Furthermore, syllables participate widely in speech errors. Syllabic positions, i.e., onset, nucleus, coda, are typically respected when segments in an utterance are exchanged (Fromkin, 1971; Garrett, 1975, 1982). Although syllables are often confounded with other units such as morphemes (e.g., "shoul*wards* for*ders*" instead of "shoulders forward"; Garrett, 1975) or combinations of phonemes in speech errors (e.g., "mon*kle*'s un*key*" instead of "monkey's uncle"; Garrett, 1975), there are a couple of speech errors that suggest the involvement of syllables as independent units (e.g., "tremenly" instead of "tremen*dous*ly" and "specifity" instead of "specifi*ci*ty"; Fromkin, 1971). Fromkin (1971, p. 48) assumed that "syllables constitute units in the production of a speech utterance". The usually tens of thousands of words in any given language are composed of a relatively small inventory of syllables. In fact, 500 syllables from English, from Dutch, and from German suffice to produce approximately 80% of all speech in those languages (Schiller, Meyer, Baayen, & Levelt, 1996). Thus, utilizing the syllable as a production unit has the potential to aid in the speedy production of the majority of words in a language.

However, despite the apparent relevance of syllables to a variety of linguistic phenomena, their involvement in speech errors (Berg, 1988; Fromkin, 1971; MacKay, 1970; Nooteboom, 1969; Shattuck-Hufnagel, 1979; Stemberger, 1982) and meta-linguistic games (Schiller, Meyer, & Levelt, 1997; Treiman, 1983; Treiman & Danis, 1988; Bagemihl, 1995) there is to date not a wealth of experimental evidence to support their relevance and involvement in the processes of speech production.

A MODEL OF WORD FORM ENCODING

There are several conceivable ways in which syllables could play a role in the processes of speech production. One possibility is that the phonological forms of words are pre-syllabified in their stored representations (Dell, 1986, 1988). Another possibility is that rough motor execution programs which correspond to segmentally specified syllables are stored, separate from the store of lexical word forms, and are retrieved to facilitate articulation.

The existence of such a store, called *mental syllabary*, is an inherent part of Levelt, Roelofs and Meyer's (1999) theory of spoken word production and its computer simulation WEAVER++ (Roelofs, 1997a, 1997b, 1998). This model outlines how words are generated from a first stage of conceptual preparation to the last stage, articulation. Word generation proceeds through different processing steps: syntactic, morphological, phonological and phonetic encoding, ending with articulation itself. The mental syllabary is assumed to function as a mediator at the interface between phonological and phonetic encoding.

During *phonological encoding*, an abstract phonological word form is retrieved from the mental lexicon. Two kinds of phonological information can be distinguished, segmental and metrical information. Segmental information specifies the phonemic structure of the word form, whereas metrical information specifies the number of syllables the word form consists of and possibly the word's stress pattern. Segmental and metrical information are retrieved separately (see Roelofs & Meyer, 1998) and merged in a process called prosodification, creating abstract phonological syllables. For common syllables, the output of phonological encoding serves as input to the mental syllabary; the abstract phonological syllables activate stored syllable motor programs. The output of the mental syllabary in turn serves as input to *phonetic encoding*, at which time contextually-driven phonetic fine-tuning of retrieved motor programs occurs. For new or very low-frequency syllables, articulatory plans are assembled using the segmental and metrical information specified in the phonological syllables. The laryngeal and supralaryngeal apparatus and its neural control transforms these articulatory plans into overt speech.

This model of word form encoding makes some crucial claims about how syllables are formed and their role in speech production:

(i) The word form which is retrieved from the mental lexicon during phonological encoding is not syllabified, i.e., not specified for syllable-internal positions such as onset or coda. The motivation for the assumption that the word form is not pre-syllabified arises from the phenomenon of *resyllabification*: Resyllabification is the process by which segments are reassigned to new syllable positions given the context in which they appear.

For example, the segment-to-syllable position assignments for a pre-syllabified word-form representation, such as *de.mand,* would need to be reassigned to new syllable positions in a different context, such as past tense (*de-man-ded*) or cliticization (*demand it – de-man-dit*). Given the frequency with which segments would need to be resyllabified in fluent speech, economy of storage and processing dictates that word forms should not be pre-syllabified.

(ii) Syllabification is an (late) on-line process, which occurs when metrical and segmental information are combined during prosodification. The domain for syllabification is the phonological word, which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticization (Booij, 1995). The incremental composition of segments to abstract phonological syllables follows, on the one hand, universal syllabification constraints (such as maximization of onsets and sonority gradations) and, on the other hand, language-specific rules, e.g., phonotactics. Together, these rules create well-pronounceable syllables. The output of phonological encoding is specified for its metrical, syllabic, and segmental properties.

(iii) Crucially, it is assumed that these phonological syllables have phonetic counterparts that are stored and accessed in the mental syllabary (Crompton, 1981; Levelt & Wheeldon, 1994). These speech syllables are the basic units for articulation (Goldstein & Fowler, 2003). The obvious advantage of a mental syllabary is that it greatly reduces the programming load relative to an on-line segment-by-segment assembly of phonetic forms.

These three claims serve as the theoretical background for the present dissertation.

Evidence for the syllable

Prior on-line investigations on the role of the syllable in speech production used a classic priming method (for Dutch: Baumann, 1995; Schiller, 1997, 1998; for French: Brand, Rey, & Peereman, 2003; Evinck, 1997; Schiller, Costa, & Colomé, 2002; for English: Schiller, 1999, 2000; for Spanish and an overview see Schiller et al., 2002). In these studies, a syllabic prime was given that was either congruent or incongruent with the syllabic structure of the target. The standard finding from these studies was that it was the number of overlapping segments between prime and target that drove the facilitative effect on the target production. This *segmental overlap effect* was also found to be independent of matching or mismatching syllable structure. No specific syllable priming effects were obtained with priming methods. This was interpreted as to be in agreement with the first crucial claim from the Levelt model, namely that word form representations are not pre-syllabified. On the basis of these converging results it was concluded that the syllable priming method taps into an early stage of the speech production process, where only *segmental* but no *syllabic* structure is available, namely the *segmental retrieval procedures*. It is assumed that the syllabic primes speed up the *segmental* retrieval from the mental lexicon. Given what these results indicate about the nature of the priming methodology, a new method must be developed that taps into later phonological and phonetic encoding processes if syllable effects are to be revealed.

OUTLINE OF THE THESIS

The first goal of the dissertation is to find evidence for the syllable playing a functionally important role in speech production. To this end, in Chapter 2 a new variant of the implicit priming paradigm was used for investigating effects of *syllable preparation*. Preparation effects have been reported in an experimental situation where speakers know in advance about the initial part of the to-be-produced utterance (e.g., Meyer, 1990, 1991; Janssen, Roelofs, & Levelt, 2002). Depending on the structure of the previously given information, speakers can prepare some initial part of the target utterance. I tested whether or not speakers can benefit from information about the syllabic structure of the target word in order to prepare their utterance, this in addition to their (already proven) use of segmental advanced knowledge. In two experiments investigating two different syllable structures, it could be shown that implicit knowledge of both the segmental and the syllabic structure of the target words produces faster responses compared to when only information about the segments was available. These experiments yielded evidence for the functional role of syllables at the interface of phonological and phonetic encoding in speech production. Given the nature of this task, the observed results are consistent with the first claim, i.e., that retrieved word-form representations are not pre-syllabified. Furthermore, the assumption that syllabification is a late on-line process (the second claim) is confirmed by these data. The results from Chapter 2 support the notion that the syllable is a relevant unit at the interface of phonological and phonetic encoding, but it does not directly support the notion of a mental syllabary.

In Chapter 3, effects of syllable frequency were investigated. Syllable frequency effects provide evidence for the existence of the mental syllabary because only stored units are likely to exhibit frequency effects (see Oldfield and Wingfield, 1965; Jescheniak and Levelt, 1994). In a series of experiments, a newly developed *Symbol-Position Association Learning Paradigm* was used to investigate the production of high and low-frequency syllables in mono- and disyllabic Dutch pseudo-words. Significant syllable frequency effects provide strong evidence for the notion of a mental syllabary, supporting the third crucial claim from the model.

In Chapter 4, effects of syllable preparation and syllable frequency were investigated in a combined study to disentangle the two effects. If the preparation effects includes access to the mental syllabary in sets where the syllable structure is shared, then we should obtain an interaction between the two factors *preparation* and *frequency*: Low-frequency syllables should yield a larger preparation benefit than high-frequency syllables because low-frequency phonetic syllables take longer to be retrieved from the syllabary, therefore yielding a larger gain when prepared.

In a reading variant of the classic implicit priming paradigm, the production of frequency-manipulated Dutch pseudo-words was investigated. It was shown that advanced preparation of syllables includes access to the mental syllabary. The results reported in this chapter converge with those reported in Chapters 2 and 3 and add further support to the second and third critical claims from the model.

Chapter 5 summarizes the empirical findings that lend support to the second and third crucial claims regarding the syllable, namely that phonological syllables are produced on-line and abstract phonetic syllables are retrieved from the mental syllabary. These results are also linked to their implications for theories of speech production and potential further research is discussed.

THE PREPARATION OF SYLLABLES IN SPEECH PRODUCTION

CHAPTER 2

(A slightly adapted version of this chapter is published as Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*, 47-61.)

ABSTRACT

Models of speech production assume that syllables play a functional role in the process of word-form encoding in speech production. In this study, we investigate this claim and specifically provide evidence about the level at which syllables come into play. We report two studies using an *odd-man-out* variant of the *implicit priming paradigm* to examine the role of the syllable during the process of word formation. Our results show that this modified version of the implicit priming paradigm can trace the emergence of syllabic structure during spoken word generation. Comparing these results to prior syllable priming studies, we conclude that syllables emerge at the interface between phonological and phonetic encoding. The results are discussed in terms of the WEAVER++ model of lexical access.

INTRODUCTION

The role of the syllable as a functional unit in the speech production process has been investigated in several psycholinguistic studies (Baumann, 1995; Chen, Chen, & Dell, 2002; Chen, Lin, & Ferrand, 2003; Ferrand, Segui, & Grainger, 1996; Ferrand, Segui, & Humphreys, 1997; Levelt, 1992, Levelt & Wheeldon, 1994; Meijer, 1996; Schiller, 1997, 1998, 1999, 2000; Schiller, Costa, & Colomé, 2002). Many of the off-line studies suggest the existence of the syllable as a production unit (Fromkin, 1971; Schiller, Meyer, & Levelt, 1997; Shattuck-Hufnagel, 1987, 1992; Treiman, 1983; Treiman & Danis, 1988). For example, speech error data suggest that segmental errors such as exchanges of segments only take place for identical syllable internal positions, i.e., onsets exchange with onsets, nuclei exchange with nuclei, etc. (Berg, 1988; MacKay, 1970; Nooteboom, 1969; Shattuck-Hufnagel, 1979; Stemberger, 1982). This is referred to as the *syllable position constraint*. However, a quantitative analysis showed that the majority of such errors occurs in the onset position. Thus, the syllable onset constraint may be a word-onset constraint (Shattuck-Hufnagel, 1987, 1992; see Meyer, 1992, for a critical review). Evidence from meta-linguistic tasks suggests that syllables play a role at some level of processing in speech production (Schiller et al., 1997; Treiman, 1983; Treiman & Danis, 1988; see Bagemihl, 1995 for a review) but makes no strong claim about where. Despite the (limited) off-line support for the syllable and the relevance of syllables to linguistic phenomena, on-line experiments do not provide evidence that the syllable is a production unit (Brand, Rey, & Peereman, 2003; Evinck, 1997; Schiller et al., 2002).

The majority of prior on-line studies used some form of priming as their experimental method. The experiments reported here use a different paradigm to investigate the syllable as a processing unit, i.e., the *implicit priming paradigm* (Meyer, 1990, 1991). The existence of syllabic units is assumed by two influential models of speech production, i.e., the Levelt, Roelofs, and Meyer model (1999) on the one hand and the model proposed by Dell (1986, 1988) on the other hand. Despite this general agreement, these models differ in the status of syllabic units in phonological encoding. Dell's (1986, 1988) model includes word forms that are already syllabified when retrieved from the mental lexicon, i.e., an abstract phonological representation which is specified not only for its segmental composition but also for its internal syllabic structure. In contrast, the model of spoken word production proposed by Levelt et al. (1999) assumes syllables play a crucial role at the interface of phonological and phonetic encoding. At this interface, abstract phonological syllables are generated which are subsequently mapped onto phonetic syllables. We will return to the issue of when syllables are predicted to play a role in speech production periodically throughout this chapter. Here, we introduce a version of the implicit priming paradigm that specifically taps into the preparation of syllable structure. The Levelt et al. (1999) model of lexical access, and its computer simulation WEAVER++, will be taken as the theoretical framework for the interpretation of our findings. This requires a short introduction to the model's phonological and phonetic encoding parts.

PHONOLOGICAL AND PHONETIC ENCODING IN WEAVER++

According to the WEAVER++ model (Roelofs, 1997b; Levelt et al., 1999), the preparation of a spoken word proceeds through a number of stages. After conceptually driven selection of the appropriate lemma from the mental lexicon, the target word is first phonologically encoded, which largely consists of computing its syllabification and prosody. This is incrementally followed by phonetic encoding, which includes the computation of the articulatory gestures for the target word's syllables in their phonetic context. Finally, the execution of these gestural scores by the laryngeal and supralaryngeal muscle systems produces the acoustic realization of the spoken word. The present chapter exclusively concerns the stages of phonological and phonetic encoding.

Phonological encoding. The first operation in phonological encoding is the retrieval of the target word's *phonological code* from the mental lexicon. The code consists of an ordered set of phonemic segments. For stress-timed languages such as English and Dutch the model also assumes the existence of sparse metrical markers in phonological codes. More specifically, the stress position is marked for those words whose stress does not appear in default position (but see Schiller, Fikkert, & Levelt, in press for a different position). For English, the default position is defined as the first full-vowel syllable of the word. Different from other models of spoken word production (in particular Dell, 1986, 1988), Levelt et al.'s retrieved phonological codes are not syllabified. The main argument for this assumption derives from the phenomenon of *resyllabification*. In connected speech, syllable boundaries often differ from a word's canonical syllabification. The domain of syllabification is the phonological word which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticization (Booij, 1995). If, for instance, the stored phonological code for the word *predict* would be syllabified (i.e., as *pre-dict*), then the speaker must 'resyllabify' the word when used in a different context, such as past tense (*pre-dic-ted*) or cliticization (*predict it* - *pre-dic-tit*). The ubiquity of such 'resyllabifications' in the normal use of English (or Dutch for that matter), would make this a highly inefficient procedure.[1] For a language like Mandarin Chinese, which has a small set of syllables and limited resyllabification processes, the story might be different. The issue of cross-linguistic differences will be revisited later in the chapter.

The alternative assumption is, therefore, that a word's syllabification is not retrieved, but generated 'on the fly', dependent on the context in which the word appears. During this process, called 'prosodification', spelled-out segments are incrementally combined to form successive syllables. Also, these successive syllables are incrementally assigned the appropriate metrical properties, either following default stress, or otherwise the retrieved non-default stress marking feature. The incremental composition of syllables follows, on the one hand, universal syllabification constraints (such as maximization of onsets and sonority gradations) and, on the other hand, language-specific rules, e.g., phonotactics. Together, these rules create maximally pronounceable syllables. The output of phonological encoding is a phonological word, specified for its metrical, syllabic, and segmental properties.

Before turning to the next processing step we will briefly describe the assumptions Dell's (1986, 1988) model makes with respect to the phonological encoding process. As already mentioned above, this model includes abstract phonological representations that are specified for internal syllabic positions, i.e., the word form retrieved from the mental lexicon activates not only segmental information but also syllabic frames. These syllabic

---

[1] Schiller, Meyer, Baayen, and Levelt (1996) estimated the occurrence of resyllabifications in a running text for Dutch. Approximately once every six words speakers of Dutch would have to resyllabify their lexical forms.

frames serve as placeholders into which the retrieved segments are inserted during the process of segment-to-frame-association.

Phonetic encoding. These fairly abstract, syllabified phonological words are incrementally translated into articulatory-motor programs. These programs consist in large part of specifications for subsequent syllabic gestures. One assumption of the theory is that speakers have access to a repository of syllabic gestures. This repository, coined the ‘mental syllabary’ (Levelt, 1992; Levelt & Wheeldon, 1994), contains the articulatory scores for at least the high-frequency syllables of the language. Schiller has computed that English speakers do some 85% of their talking with no more than 500 different syllables (out of some 12,000, see Schiller et al., 1996; Schiller, 1997). Hence, for normal speakers, the corresponding articulatory gestures may have become highly over-learned motor actions. The model assumes that as soon as a syllable emerges during incremental syllabification, the corresponding syllabic gesture will be selected from the repository in Broca's area or a pre-motor area (Dronkers, 1996; Indefrey & Levelt, 2000; Kerzel & Bekkering, 2000). Proposing the notion of a mental syllabary, however, was not intended to deny the existence of a mechanism for the generation of low-frequency or entirely new syllabic gestures. That mechanism is still to be modeled in detail within the framework of WEAVER++, but irrelevant for the present discussion.

In summary, the theory proposed by Levelt et al. (1999) takes syllables, rather than segments, to be basic programming units of speech articulation. This is entirely in line with traditional notions of speech generation. According to Fujimura and Lovins (1978) as well as Lindblom (1983) only the syllable can form the appropriate source for late phonological processes, such as allophonic variation, coarticulation and (as a result of this) assimilation. Phenomena such as word-initial aspiration of plosives in English or word-final devoicing in Dutch or German can be described conveniently with reference to the syllable as a unit (see Kenstowicz, 1994).

The WEAVER++ model specifies where syllabic patterns emerge in speech generation. First, syllables are not stored in the mental lexicon; they are not specified in the phonological codes speakers retrieve from their form lexicon. This predicts the absence of syllable-specific effects in priming paradigms because syllables are not represented as units in long-term memory. Below, we will discuss in more computational detail the basis for this prediction as well as the relevant evidence (and counter evidence). Second, phonological syllables first arise during incremental phonological encoding, i.e., during context-sensitive ‘syllabification’. Third, as phonological syllables arise, they trigger the retrieval of syllabic articulatory gestures (‘phonetic syllables’) from a repository of articulatory motor actions, the ‘mental syllabary’. The purpose of the present chapter is to trace the emergence of syllables in word generation by means of a paradigm which manipulates the speaker’s ability to do advance preparation of a syllable. It can provide the speaker with a head-start in syllabification and in retrieving a word’s first syllabic gesture.

PRIMING STUDIES OF SYLLABLE ACCESS

Several cross-linguistic studies were conducted to investigate whether syllables could be primed and thereby identified as an independent unit in the process of speech production (for Dutch: Baumann, 1995; Schiller, 1997, 1998, 1999; for Mandarin Chinese: Chen et al., 2003; for French: Brand et al., 2003; Evinck, 1997; Ferrand et al., 1996; Schiller et al., 2002; for English: Ferrand et al., 1997; Schiller, 2000; Schiller & Costa, submitted; for Spanish and an overview see Schiller et al., 2002). One of the first studies that was conducted in order to test whether syllables can be primed in speech production was Baumann (1995). She investigated the time course of syllabification during phonological encoding in Dutch. In a series of priming experiments using a semantic-associate learning task, she tested whether a syllable priming effect could be obtained. A first finding of her experiments was that phonologically related primes, whatever their syllabic relation to the target word, facilitated the response relative to unrelated control primes. A second result was that, in all related conditions, CVC-primes were more effective than CV-primes. But, thirdly, no specific syllable priming effects were obtained.

Several subsequent studies have failed to find a syllable priming effect but rather confirmed the finding of a segmental overlap effect. Much discussion has been given to the results of the apparent syllable priming effect in French (Ferrand et al., 1997, Experiment 5). However, Brand et al. (2003)’s failure to replicate the Ferrand effects suggests that this should not be taken as strong evidence for the syllable effect (see also Evinck, 1997; and for a review Schiller et al., 2002). In sum, the evidence from syllable priming tasks may indicate this method is not tapping into the appropriate level of processing to reveal potential syllable effects.

SYLLABLE FREQUENCY STUDIES

Syllabic effects are, however, predicted for access to the hypothesized mental syllabary. Gestures for high-frequency syllables should be more accessible than gestures for low-frequency syllables. (The argument is further spelled out in the next section). In order to find empirical evidence for this, Levelt and Wheeldon (1994) investigated naming latencies for words consisting of high- versus low-frequency syllables. The prediction in three naming tasks was that, under the assumption of the existence of a mental syllabary, onset latencies for words that consist of high-frequency syllables should be shorter than those for words consisting of low-frequency. This expectation was inspired by the finding that word form access is sensitive to word frequency (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965), although that involves a different level of processing. Levelt and Wheeldon's core finding was that, when word frequency was controlled for, words with high-frequency syllables were named faster than words with low-frequency syllables. If syllables are computed on-line rather than retrieved from a repository, their frequency of use should be irrelevant. The obtained syllable frequency effects therefore seemed to support the notion of the mental syllabary, where syllables are stored separately from words.

One potential problem with this conclusion is that syllable frequency was correlated with segment frequency in some of Levelt and Wheeldon's experiments. It is hardly possible in Dutch to disentangle these effects. A replication of this syllable frequency effect with carefully controlled experimental material would be desirable to allow for any strong claims.

THE WEAVER++ PREDICTIONS IN MORE DETAIL

The WEAVER++ model provides an account for the absence of a syllable priming effect and for the presence of a syllable frequency effect. So far, the model has been more successful in the former case than in the latter. Here the computational rationale is discussed in some more detail, because it provides at the same time the motivation for the present experiments. *WEAVER* (*W*ord-form *E*ncoding by *A*ctivation and *VER*ification) is the spreading activation based computer network model developed by Roelofs (1992, 1996, 1997a, 1997b, 1998, 1999), which is based on Levelt's (1989, 1992) theory of speech production. WEAVER++ adopts Dell's (1986) assumption of word form retrieval by the spread of activation and Levelt's (1992) on-line syllabification and access to a syllabary (Levelt & Wheeldon, 1994).

In accordance with Levelt and Wheeldon (1994), WEAVER++ (Roelofs, 1997a, 1997b) assumes that the syllabification of a word is computed on-line during the speech production process. In the WEAVER++ model, segments in the retrieved phonological code are not specified for their syllable position, but only for their serial order within a word. The actual syllabic position of a segment is determined by the syllabification process. Each retrieved segment in the phonological code spreads activation to all syllabic gestures in which it partakes. Hence, upon retrieval of a phonological code, there are always multiple phonetic syllable programs in a state of activation. How is the appropriate syllable program selected? There are, first, selection conditions. The crucial one is that the syllable matches the phonological syllable that is incrementally composed; this involves a procedure of verification. Second, each syllable in the syllabary has a frequency dependent selection threshold. This causes the predicted syllable frequency effect on naming latencies. Notice, however, that the threshold assumption is a ‘modular’ one. Removing it does not affect the architecture of the system. Third, selection is subject to Luce's (1959) choice rule. During any smallest interval, the probability of selecting the (verified) target syllable equals the ratio of its activation to the summed activation of all syllable nodes. Given the choice ratio, the expected selection latency can be computed.

WEAVER++ does not predict any syllable priming effects, at least not for Dutch and English for the following reasons: First, there is no syllable structure in the phonological code. The code's retrieval cannot be specifically primed by a string of segments that matches the word's canonical syllable structure. Phonologically related primes in the masked priming paradigm (pre-)activate the phonological segments retrieved from the mental lexicon during segmental spell-out. As a consequence, the longer the prime the more segments get (pre-) activated during segmental spell-out and the shorter the phonological code's selection latency. Notice that primes never get articulated in priming tasks. Hence, there is no need to transform incoming segments into syllables, i.e., no syllable structure is imposed on the input (it is free to map onto all compatible syllables).

In addition, incremental syllabification is not specifically facilitated by syllable matching primes. Take a CV.CVC target word such as *lotus*. A masked visual CV-prime (LO) will activate the first two segments and all syllable programs in which they partake, including the syllable program [lo] but also the syllable program [lot]. A CVC-prime (LOT) will activate the first three segments of the phonological code, the syllable program [lo] to the same amount that the CV prime (LO) did, and in addition the syllable program [tus] to some extent. Hence, it primes the relevant syllable programs despite the fact that it does not correspond to the first syllable of the target word. Therefore, there will only be a number-of-segments effect. The same holds when the target word has a CVC.CVC structure, such as *cactus*. Here the prime (CA) will be less effective than the prime (CAC), for similar reasons. Thus, WEAVER++ predicts an effect of prime length or the so-called segmental overlap effect (Schiller, 1998, 1999, 2000) but no interaction of prime and target syllabic structure.

Following the arguments of the model, in order to identify the syllable as a processing unit, we have to investigate the late syllabification process with a method that involves the advance construction of a phonological word's first syllable and the corresponding advance syllabary access. The implicit priming paradigm (Meyer, 1990, 1991; Roelofs, 1996, 1998; Roelofs & Meyer, 1998) provides access to exactly these late steps in spoken word encoding. Whereas (explicit) priming is sensitive only to early stages of phonological encoding, the implicit priming paradigm exhibits effects that emerge at these early stages but also comprise later stages at the interface of phonological and phonetic encoding, i.e., on-line syllabification, possibly including syllabary access.

In the implicit priming paradigm, participants repeatedly produce a syllable shared by several response words (as in *lotus*, *local*, *loner*). As speakers know of the shared properties between response words they can prepare the first phonological syllable and the corresponding syllable program of the target word. The empirical issue is then whether a prepared syllable of the target word is a more effective preparation than a non-syllabic string of prepared segments, other factors being controlled. To answer this question, we designed a special version of the implicit priming paradigm.

THE IMPLICIT PRIMING PARADIGM

The basic paradigm. The implicit priming paradigm involves the production of words that are part of a list of previously learned paired associates. Participants learn a small set of prompt-response pairs. The response words are either phonologically related or not. Each experiment using the implicit priming paradigm consists of two types of sets, called the homogeneous and heterogeneous sets. In the homogeneous set, the response words share part of their form, e.g., the first syllable in *loner*, *local*, *lotus*, or the first syllable in *beacon*, *beadle*, *beaker*, or the first syllable in *major*, *maker*, *maple*. The heterogeneous sets are created by regrouping the pairs from the homogeneous sets, e.g., *loner*, *beacon*, *major* (etc.). Each word is thus tested under both the homogeneous and the heterogeneous conditions, hence each word is its own control in the experiment. Production latency (the time between onset of prompt and speech onset, measured by voice key) is the dependent variable. A preparation effect is said to have occurred if production latencies in the homogeneous condition are shorter than in the heterogeneous condition. Meyer (1990, 1991) reported such a preparation effect only when the response words in the homogeneous sets shared one or more word-initial segments. No effect was found for shared word-final segments demonstrating the incrementality of the process of syllabification. The preparation effect was found to increase with the length of the shared initial stretch.

Using this paradigm, Meyer (1991) reports that sets with open initial syllables (CV) that share only those two initial segments produced preparation effects that were equivalent to effects produced for sets with closed syllables (CVC) that shared three initial segments. This result was surprising because a pure segmental overlap effect would predict larger preparation effects in the CVC sets since they comprise one more shared segment. This finding supports the possibility of syllabic effects that are independent of segmental overlap[2].

The paradigm with an odd-man-out. To investigate a specific syllable preparation effect, we opted for a slightly different variant of the original implicit priming paradigm,

---

[2] However, contrary to Meyer's (1991) results, Roelofs (1996, Experiment 6) showed that the size of the preparation effect depends on the length of the shared syllable in terms of number of segments.

i.e., the implicit priming paradigm with an *odd-man-out* (Janssen, Roelofs, & Levelt, 2002). The term odd-man-out labels an item in the response list of a homogeneous set that has (compared to the other words in that list) a different feature, such as another syllabic structure. The homogeneous set containing an odd-man-out is the so-called *variable* set; the homogeneous set without an odd-man-out is called the *constant* set.

In a constant set, the response-words consist of items which share *two* phonological properties. One is always the shared word-initial segments. The other can, for instance, be the word's syllable structure. The constant set can be *beacon*, *beadle*, *beaker*, sharing both the initial CV and the initial syllable. In comparison with the constant set, the variable sets contain only one of these two phonological properties, namely the shared word onset. A related variable set could be *beacon*, *beatnik*, *beaker*, where all items share the initial CV, but where they do not share the initial syllable (*bea* vs. *beat*).

We chose the odd-man-out variant of the implicit priming paradigm rather than the classic version for its ability to keep one phonological property constant while systematically manipulating the other. In the present studies, the number of shared initial segments is held constant but the underlying syllabic structure differs. Thus, we can investigate what knowledge about the to-be-produced word speakers need in order to (successfully) prepare for it. If segmental information about syllabic structure is sufficient to prepare for the target word, then it should make no difference whether or not the response words in a set additionally share the syllabic structure. For this scenario, response times should be equally fast independent of syllabic structure; thus no effect of the manipulation should be observed. If speakers need the information of the current syllabic structure in addition to the information of the shared initial segments, then they should be faster in preparing for constant items than for variable items. In the experiments reported below, we argue that participants are only able to successfully prepare for the target word in those cases in which they have both types of information from the constant sets. In variable sets, in which the number of shared segments is invariable but the syllabic structure is different for one item in the response set, we argue that participants are not able to prepare for *any* members of that response set as they cannot predict the next upcoming syllable.

To ensure that it is in fact the case that not only the odd-man-out might be excluded from a preparation mechanism, the odd-man-out itself is excluded from the analysis. As already mentioned, we predict that the odd-man-out hinders participants from fully preparing for any of the first syllables within the response set; thus we should be able to find the effect even after exclusion of the odd-man-out.

EXPERIMENT 2.1: PRODUCTION OF CVV TARGETS

In order to test if the emergence of the syllable is traceable in the preparation of spoken Dutch words, two experiments were carried out using the odd-man-out variant of the implicit priming paradigm. Under the assumption that the syllable indeed represents a processing unit at the phonology/ phonetics interface, the odd-man-out should reduce the preparation effect in the variable set as compared to the constant set. The odd-man-out, with a syllable structure that differs from the other members in the set, is expected to spoil the preparation effect, not only for the odd-man-out, but also for the other set members. However, if speakers just need the segmental information for preparation, the reaction times for constant and variable sets should show no difference.

METHOD

Participants. Twenty-four native speakers of Dutch participated in the first experiment. They were randomly taken from the pool of participants of the Max Planck Institute in Nijmegen, The Netherlands and were paid for their participation.

Materials. Eight different Dutch verb stems served as base for constructing eight different experimental blocks (four verb stems for the constant and four verb stems for the variable sets). The verb stems within each set were presented in three different inflectional forms plus the corresponding noun, e.g., the infinitive form of the verb stem ‘leid-‘ *lei.den* (CVV; [to] lead), the corresponding noun *lei.der* (CVV; leader), the gerund *lei.dend* (CVV; leading), and the past tense form *lei.dde* (CVV; led). As all response words are derived from the same verb stem, segmental overlap within each set was assured. In constant sets, the first syllable of all four words in the set had the same syllable structure, as in the ‘*lei*.den’-set quoted above. The variable sets were constructed in the same way, but the past tense form of the verb stems selected for these sets had a different syllable structure compared to the other members in their sets, e.g., *ro.ken* (CVV; [to] smoke), *ro.ker* (CVV; smoker), *ro.kend* (CVV; smoking), but *ROOK.te* (CVVC; smoked)[3]. Thus, all items in the constant and the variable sets have the same syllable structure in the first syllable with the exception of the past tense form in variable sets; this item with the deviating syllable structure served as odd-man-out. A full list of items is given in the Appendix 2-A. Prompt words for each target word in each condition were derived from the strong verb *staan* ([to] stand), *staander* (stand), *staande* (standing), *stond* (stood). The use of this irregular verb with its four (irregular) inflectional forms guaranteed that there was no form overlap between prompt and target (as would have been the case by using a regular verb, e.g., wer.*ken* – ro.*ken*). The fact that the same prompt was used for all target words allowed for maximum comparability within and between sets and also simplified the participant's learning task.

Design. Constant and variable sets were presented in a four and in a three-item condition. The four-item sets contained all of the items mentioned above, whereas in the three-item sets the past tense form, i.e., the one that caused the odd-man-out in the variable set but not in the constant set, was excluded (see Table 2-1). Although the syllable structure is therefore constant in the three-item sets, we will denote the pair of a constant four-item set and its three-item derivative set the 'constant condition', and the pair of a variable four-item set and its three-item derivative the 'variable condition'. Hence, we crossed two factors: a factor "Word Type" (opposing the constant and variable conditions) and a factor "Set Size" (opposing the four- and three-item sets). In the data analysis, the past tense form in the four-item sets was also excluded. The voice onset latencies of the remaining three forms were compared to those of their corresponding three-item sets. As the odd man out is expected to spoil the preparation effect for the whole set, the original four-item set in the variable conditions should show larger latencies compared to their three-item sets as well as to the other constant sets. The resulting advantage of presenting the response words in different set sizes, i.e., three- vs. four-item sets rather than in heterogeneous and homogeneous sets, as in the original version of the implicit priming paradigm, is that the two types of sets are more comparable to each other

---

[3] Vowels in open syllables, as in *ro*.ken and vowels marked twice in orthography, as in *rook.te*, both have a long pronunciation in Dutch.

because we did not present different lemmas in the heterogeneous sets (as would happen by regrouping items from the homogeneous sets) but only items which are derived from the same stem. The two-by-two design was within-subjects. Each participant was presented four out of eight experimental sets, two sets being variable and two being constant.

Table 2-1. *Response set and prompts within a constant and a variable set in a three- and a four-item set.*

	Word Type			
	Constant sets		*Variable sets*	
Prompts	Four-item set	Three-item set	Four-item set	Three-item set
staan ([to] stand)	lei.den ([to] lead)	lei.den	ro.ken ([to] smoke)	ro.ken
stond (stood)	lei.dde (led)		rook.te (smoked)	
staander (stand)	lei.der (leader)	lei.der	ro.ker (smoker)	ro.ker
staande (standing)	lei.dend (leading)	lei.dend	ro.kend (smoking)	ro.kend

Procedure and Apparatus. The participants were tested individually in a quiet room. They were given detailed written instruction specifying that they had to respond as accurately and as quickly as possible. The experiment consisted of alternating learning and test phases. In the learning phase, participants were shown the four (or three) pairs of prompt-response words of a set on the computer screen (NEC Multisync3FG). When they indicated that they had studied the pairs sufficiently, the experimenter started the practice phase in which participants saw all four prompts together on the computer screen and they had to produce the corresponding responses in a row. When they failed, the learning phase was started again and the session was rehearsed to ensure that they learned the sets accurately. After they successfully completed the practice phase, the experimenter started the test phase. Each trial started with an attention sign (asterisk) marking the position of the prompt. The asterisk was displayed for 500 ms and after a pause the prompt was presented. Four different presentation times (after a 350/ 600/ 850/ 1300 ms pause) equally distributed across conditions were chosen to prevent speakers from producing the prepared target onsets before the prompt was actually displayed. Simultaneously with prompt presentation the voice key was activated for 1500 ms. The prompt disappeared after the response with a delay of 500 ms. The asterisk of the next trial appeared after 100 ms. Prompts within each set were repeated five times in a random order, resulting in response sets of 15 items for the three-item sets, and 20 items for the four-item sets.

The presentation of the stimuli and the measuring of the reaction times were controlled by the NESU software package. The spoken reactions were registered by a Sennheiser MD211N microphone, which fed into a NESU-box voice key device and a DAT recorder (Sony DTC-55ES). The experimenter sat in the same room and took note of hesitations, voice key errors, wrong naming responses, and time outs. After the completion of each item set, the number of successful trials and the corresponding mean reaction time were displayed on the participant's screen. The total duration of the experiment varied as a function of participant's learning time. On average, an experimental session lasted 30 minutes.

RESULTS

Only those test items for which a correct response was obtained were included in the reaction time analysis. Test items leading to wrong or invalid responses were not included (all wrong naming responses, voice key errors, and hesitations). Time outs (RTs > 1500 ms) and extreme outliers (i.e., naming latencies shorter than 300 ms) were also removed. Two participants were excluded from the analysis because of high error rates (more than 20% errors). The mean voice onset latencies, standard deviations, error rates, and preparation effects are summarized in Table 2-2.

Table 2-2. *Mean voice onset latencies (in ms), standard deviations, percentage errors (in parentheses), and preparation effects (in ms) in Experiment 2.1.*

	Set Size						
Word Type	Three-item sets			Four-item sets			Preparation Effects
	M	SD	% Err	M	SD	% Err	
Constant	606	152	(5.0)	631	163	(2.4)	25
Variable	595	170	(0.3)	686	182	(4.7)	91

Analyses of variance were run with Set Size (three-item sets versus four-item sets) and Word Type (constant versus variable) as independent variables. As mentioned before, we excluded the past tense form in the four-item sets from the analyses and compared only the remaining three forms (infinitive, noun, and gerund) to their corresponding three-item sets. The term ‘four-item sets’ always refers to the original four-item sets with the excluded past tense form to distinguish them from the corresponding three-item sets.

Error rates. In this experiment, there were 3.1% trials excluded altogether. As none of the main effects or interactions were significant, the error analysis is not reported.

Reaction times. As a first result, we expected an effect of Set Size, i.e., the three-item sets were expected to be produced faster compared to their four-item sets. The reduced set-size was expected to make it easier for the participants to recall the items, thus resulting in shorter voice onset latencies. This is confirmed by the data. Participants responded on average 58 ms faster in the three-item sets compared to their four-item sets. The main effect of Set Size was significant ($F_1(1,21) = 53.12$, $MS_e = 1463.17$, $p < .01$; $F_2(1,22) = 147.89$, $MS_e = 281.82$, $p < .01$). The main effect of Word Type (variable vs. constant sets) was significant by participants but not by items ($F_1(1,21) = 8.87$, $MS_e = 1873.97$, $p < .01$; $F_2(1,22) < 1$).

The crucial prediction tested in this experiment was that of a significant interaction between the factors Word Type and Set Size. More precisely, beside the fact that the voice onset latencies in the four-item sets were larger because of the additional item, the variable four-item sets were predicted to be slower in comparison to the constant four-item sets due to the odd-man-out. We predicted the difference between the three-item sets and the four-item sets to be larger in the variable condition than in the constant condition. This prediction was confirmed by the data. The preparation effect (the difference between the production latencies of the three and the four-item sets within one condition) was much larger in the variable condition (686 ms – 595 ms = 91 ms) than in the constant condition (631 ms – 606 ms = 25 ms). The interaction of Word Type and Set Size was marginally significant for participants ($F_1(1,21) = 4.14$, $MS_e = 4925.09$, $p = .055$) and highly significant for items ($F_2(1,22) = 45.48$, $MS_e = 281.89$, $p < .01$).

DISCUSSION

The results of Experiment 2.1 show that there is in fact a preparation effect for the syllable. The effect of 66 ms can be attributed to a syllable structure effect since the overlap of initial segments is the same in constant and in variable sets. Thus, the larger onset latencies in the variable four-item sets can only be explained by the change of the syllabic structure in those sets. As we excluded the past tense forms from the four-item sets and analyzed only the remaining three forms, the effect cannot be due to the odd-man-out itself, the item with a deviating syllable structure in variable sets. As we expected, the odd-man-out spoils the preparation effect for the whole (variable) set.

The syllable structure, CV(V), that we investigated in this experiment is the most basic and simplest syllable structure universally. According to phonological theory, syllables universally ‘prefer’ to have simple onsets and no coda (Hooper, 1972; Selkirk, 1982; Vennemann, 1988). Therefore it is possible that the preparation effect we found is specific to CV-syllables and will not generalize to typologically more complex syllables. That is why we decided to try and replicate the obtained preparation effect for syllables with different properties, long vowels and complex onset clusters. Specifically, we constructed sets of target items beginning with CCVV-syllables.

EXPERIMENT 2.2: PRODUCTION OF CCVV TARGETS

The second experiment tested the same predictions as the previous experiment with different participants and materials. Materials in this experiment also consisted of disyllabic Dutch words but contained word stems which had a different syllable structure. In this experiment, all words except the odd-man-out shared a CCVV-structure as first syllable as in *pra.ten* (CCVV; [to] speak) or *sle.pen* (CCVV; [to] drag), the syllable structure of the first syllable for the odd-man-out was CCVVC as in *sleep.te* (dragged)[4]. Design, procedure and apparatus of Experiment 2.2 were the same as in Experiment 2.1.

METHOD

Participants. Twenty-four undergraduate students from the same population described in the previous experiment participated in Experiment 2.2.

RESULTS

Only those test items for which a correct response was obtained were included in the reaction time analysis. Test items leading to wrong or invalid responses were not included (all wrong naming responses, voice key errors, and hesitations). Time outs (RTs > 1500 ms) and extreme outliers (i.e., naming latencies shorter than 300 ms) were also removed. The mean voice onset latencies, standard deviations, error percentages, and preparation effects are summarized in Table 2-3.

Table 2-3. *Mean voice onset latencies (in ms), standard deviations, percentage errors (in parentheses), and preparation effects (in ms) in Experiment 2.2.*

	Set Size						
Word Type	Three-item sets			Four-item sets			Preparation Effects
	M	SD	% Err	M	SD	% Err	
Constant	601	160	(2.9)	633	174	(3.2)	32
Variable	582	164	(3.6)	675	197	(4.7)	93

---

[4] Please note, that as in the roken/ rookte example, vowels in open syllables, e.g., *sle.pen* and vowels marked twice in orthography, e.g., *slee.pte*, both have the same long pronunciation in Dutch.

Analyses of variance were run again with Set Size (three-item sets versus four-item sets) and Word Type (constant versus variable) as independent variables.

Error rates. In this experiment, 3.6% of the trials were errors. None of the main effects or interactions were significant.

Reaction times. There was a difference of 62 ms between the three-item sets and the four-item sets. This difference was significant for the factor Set Size by participants and by items ($F_1(1,23) = 48.29$, $MS_e = 1966.95$, $p < .01$; $F_2(1,22) = 128.3$, $MS_e = 369.51$, $p < .01$). The main effect of Word Type was not significant ($F_1(1,23) = 3.4$, $MS_e = 1139.52$, n.s.; $F_2(1,22) < 1$).

The preparation effect (the difference between the production latencies of the three and the four-item sets within one condition) was much larger in the variable condition (674 ms – 582 ms = 93 ms) than in the constant condition (633 ms – 601 ms = 32 ms). The interaction of Word Type and Set Size was just significant for participants ($F_1(1,23) = 4.27$, $MS_e = 5806.93$, $p = .05$) and highly significant for items ($F_2(1,22) = 30.7$, $MS_e = 369.51$, $p < .01$).

DISCUSSION

Experiment 2.2 again confirms the prediction of a preparation effect and replicates the effect of Experiment 2.1. Participants produced longer voice onset latencies in the variable four-item sets than in the three-item sets. As in the first experiment, this result can be interpreted as a syllable structure effect, because the overlap of the beginning segments was the same for variable and constant sets. The difference between the constant and variable three-item sets in this experiment showed the same pattern as in the first experiment namely that the variable three-item sets yield faster reaction times than the constant three-item sets. In Experiment 2.1, the reaction times in the variable three-item sets were on average 11 milliseconds faster in comparison to the constant three-item sets (606 ms – 595 ms); in Experiment 2.2 this difference was even stronger. Here, participants produced the variable three-item sets on average 19 ms faster than the constant three-item sets. Thus, the longer response latencies in the variable four-item sets cannot be attributed to the fact that the variable item sets were more difficult to learn or to produce. Consequently, the larger reaction times in the variable four-item sets must be explained by the syllabic change in those sets. In other words, the odd-man-out spoiled the preparation effect for the whole set.

As the crucial effect was only marginally significant for participants in the first experiment, we decided to enhance test power by combining the data of both experiments in one ANOVA, with Experiment as a between-subjects factor. This collapsed analysis shows that the effect is reliable, also for subjects. The main effect of Set Size was significant ($F_1(1,45) = 102.04$, $MS_e = 1691.21$, $p < .01$; $F_2(1,46) = 283.42$, $MS_e = 313.99$, $p < .01$). The main effect for Word Type is significant by participants but not by items ($F_1(1,45) = 11.89$, $MS_e = 1512.63$, $p < .01$; $F_2(1,46) = 1.65$, $MS_e = 4408.40$, n.s.). The interaction of Word Type and Set Size was significant for both participants ($F_1(1,45) = 8.57$, $MS_e = 5267.09$, $p < .01$) and items ($F_2(1,46) = 76.88$, $MS_e = 313.99$, $p < .01$). Neither the main effect nor any of the interactions with the factor Experiment were significant.

Experiments 2.1 and 2.2 demonstrate that response times in the implicit priming paradigm are faster when speakers have advanced knowledge about both segmental and syllabic content of an upcoming word. Prior studies suggest that increased segmental overlap leads to larger preparation effects (Meyer, 1991) and thus we believe that the variable sets in the present experiments were being prepared by participants but not to the same extent as the constant sets. However, this needs to be empirically demonstrated.

To test this claim, we carried out two control experiments in which we contrasted the constant and variable (four-item) sets both comprising initial segmental overlap to sets where there was no overlap of initial segments across items. This experimental design corresponds to the classic version of the implicit priming paradigm (Meyer, 1991) where homogeneous sets are compared to heterogeneous sets. The same materials as in Experiment 2.1 and 2.2 were used to conduct Control-Experiment 2.1 and 2.2. The former constant and variable four-item sets served as (constant and variable) homogeneous sets. The heterogeneous sets were created by regrouping the items from the homogeneous sets such that no two words in a set shared word onsets. See Table 2-4 for an example of a homogeneous and a heterogeneous set.

Table 2-4. *Example for a homogeneous and a heterogeneous set within a constant and a variable set.*

Word Type			
Constant sets		*Variable sets*	
Homogeneous set	Heterogeneous set	Homogeneous set	Heterogeneous set
lei.den ([to] lead)	lei.den ([to] lead)	ro.ken ([to] smoke)	ro.ken ([to] smoke)
lei.dde (led)	haa.tte (hated)	rook.te (smoked)	huil.de (cried)
lei.der (leader)	po.ter (person who plants)	ro.ker (smoker)	boe.ner (person who polishes)
lei.dend (leading)	wa.dend (wading)	ro.kend (smoking)	ha.kend (croching)

Note. Heterogeneous sets are created by regrouping the response words from the four different homogeneous sets (for a full list of homogeneous item sets please see item lists of Experiment 2.1 for Control Exp. 2.1 and item list of Experiment 2.2 for Control Exp. 2.2 respectively in the Appendix 2-A/ B).

A preparation effect for the homogeneous sets compared to their heterogeneous sets (independent of whether they are constant or variable) is expected as all items share phonological properties, i.e., the first segments, and allow for advanced construction of the phonological word. No preparation is possible in heterogeneous sets as they consist of all different items.

To summarize the results of these control experiments, we found an overall preparation effect for homogeneous versus heterogeneous sets in both experiments, confirming that segmental overlap, independent of syllable structure, leads to a preparation effect. The two homogeneous sets yielded in both control-experiments faster reaction times than their corresponding heterogeneous sets (for Control-Exp.1: constant and variable *homogeneous* sets: 414 ms and 456 ms, constant and variable *heterogeneous* sets: 846 ms and 818 ms; for Control-Exp. 2.2: constant and variable *homogeneous* sets: 424 ms and 445 ms, constant and variable *heterogeneous* sets: 853 ms and 829 ms).

By subtracting the mean response time for the homogeneous *variable* set from the heterogeneous *variable* set (for Control-Exp. 2.1: 456 ms – 818 ms; for Control-Exp. 2.2: 445 ms – 829 ms) we can see the magnitude of the segmental preparation effect in the absence of syllabic overlap (362 ms for Control-Exp. 2.1; 384 ms for Control-Exp. 2.2). The same calculation for the constant sets (homogeneous sets minus their corresponding heterogeneous sets: for Control-Exp. 2.1: 414 ms – 846 ms; for Control-Exp. 2.2: 424 ms – 853 ms) shows the preparation effects in case of segmental *and* syllabic overlap (for Control-Exp. 2.1: 432 ms; for Control-Exp. 2.2: 429 ms). Thus, if we further subtract the preparation effect for the constant sets from the preparation effect for the variable sets we can determine the additional preparation benefit provided by the constant syllabic structure (for Control-Exp. 2.1: 432 ms – 362 ms = 70 ms; for Control-Exp. 2.2: 429 ms – 384 ms = 45 ms). These syllabic effects replicate the syllable preparation effects of Experiments 2.1 and 2.2.

The effect for the segmental preparation is much larger than the effect for syllabic preparation (approximately 400 ms versus approximately 58 ms). However, it is unlikely that these large differences are due to segmental overlap alone. Please note that there are four different lemmas in heterogeneous sets which have to be learned and recalled whereas in homogeneous sets there is only one single verb stem in four different inflectional forms. Thus, the learning and processing load in heterogeneous sets was four times as great. Furthermore, responses in heterogeneous sets were additionally hampered by the rotation of the different inflectional forms of a single verb stem across sets (e.g., in constant set 1, participants learn *leiden* as the infinitive form but in the subsequent set, they learn that the same verb must now be produced in the past tense, *leidde*).

Finally, we can additionally extract from the data of the control-experiments the following: Under the assumption that a shared phonological property within one response set should lead to faster reaction times, one could predict that the heterogeneous constant sets should be faster compared to their variable counterparts because they share (as in their homogeneous sets) an abstract syllabic structure (not the segments) and in variable sets they do not. But the opposite is the case: Heterogeneous *variable* sets (818 ms for Control-Exp. 2.1; 829 ms for Control-Exp. 2.2) are in both control-experiments faster than the *constant* sets (847 ms for Control-Exp. 2.1; 852 ms for Control-Exp. 2.2). This shows once again (see also Roelofs & Meyer, 1998) that a shared abstract syllabic structure (CV-structure) without segmental overlap does not give rise to benefit in preparing for a target word.

GENERAL DISCUSSION

Two experiments were reported that investigated the role of the syllable in the process of spoken word production. The implicit priming paradigm seems to be an appropriate method to tap into the late processes of syllabification and syllabary access. In the experiments, participants produced previously learned target words repeatedly within an experimental block. As we chose item sets in which the response words were all derived from the same word stem, segmental overlap within all sets was provided. The homogeneity of segmental overlap in constant and in variable sets was a crucial requirement to test whether, in addition to segmental information, speakers use information about the syllabic structure in order to prepare the response. If speakers only prepare for the segmental structure of the target word, there should have been no difference between the reaction times in constant and variable sets. However, this is not what we found. In Experiment 2.1, we investigated items with a constant syllable structure of the form CVV in the first syllable, while the odd-man-out in variable sets had a deviating syllable structure in the first syllable, namely a CVVC-structure. The 'variable' three-item sets – which contained the items of the variable four-item sets, minus their odd-man-out – yielded reaction times that were on average shorter than the 'constant' three-item sets. Thus, the larger reaction times in the variable four-item sets (in comparison to constant four-item sets) cannot be explained by the nature of the items in variable sets themselves, i.e., those items being more complex or more difficult to learn or produce than those in constant sets. The same holds for the findings in Experiment 2.2, where we investigated items with a constant CCVV-structure in the first syllable and a variable syllabic structure, which consisted of a CCVVC-structure. Here, variable three-item sets were on average even shorter compared to their counterparts in constant sets.

Another difference between the items in the constant and the variable sets lies not in the syllable structure of the past tense forms of the verbs but in the complexity of the segmental transition between the two syllables. Past tense forms in variable four-item sets consist of a consonant-consonant between-syllable sequence, e.g., *klaa**g.de**,* that may be inherently more difficult to articulate than the vowel-consonant sequence found in the constant four-item sets, e.g. *knee.**dde***. This difference is an unavoidable characteristic of the nature of the present stimulus set. But while it may be that the words containing the more complex consonant clusters are harder to articulate than the words with the simpler transition and that this difference might lead to a response time difference between the two types of past tense forms, this possible difference cannot account for the longer reaction times of the *whole* (four-item) variable set. Remember that all the past tense forms are excluded from the analysis and therefore response time differences to the two types of past tense forms did not contribute to the response time difference between the variable and constant sets.

Another possibility is that the reported effects could be attributed to general memory retrieval processes rather than processes specific to speech production. Specifically, since items in the constant sets have overlap at multiple tiers (i.e., the segmental and the syllabic levels), they may be easier to retrieve from memory than items in the variable sets. However, in contrast to findings using the implicit priming paradigm, findings from immediate serial recall tasks report *slower* response times when items share phonological or phonetic features (see Baddeley, 1997, for a review). Thus, it is unlikely that the present results are due to memory effects.

Rather, the most plausible explanation for the present results is that the odd-man-out, the one variable item with the deviating syllable structure, spoiled the preparation effect for its whole (variable four) item set. Speakers could no longer prepare for the target word's first syllable. Still, we have to consider what explicitly happens in the implicit priming paradigm and which mechanisms of phonological and phonetic encoding (possibly involving syllabary access) contribute to the observed effect. In both constant and variable (four-item) sets, initial segments, which are constant across all items within a set, can be spelled-out from memory even before the prompt is displayed. However, only in the constant sets, in which the syllabic structure is also constant across items in a set, can the syllabification process begin to incrementally put the segments together and create the first syllable. This abstract phonological syllable can then be fed into the mental syllabary and activate its motor program. Thus, in constant sets, the first syllable is fully prepared for articulation. For variable sets, which lack a consistent syllable structure across items, the preparation cannot go beyond the retrieval of the initial segments. Only with the appearance of the prompt word does the further needed information about the appropriate syllable structure become available and only then can the segments be assembled for the first syllable.

Accordingly, we can conclude that, for constant sets, the first syllable can be fully prepared for articulation, including the retrieval of the corresponding gestural score from the mental syllabary. Thus, in constant sets, preparation can even go beyond on-line syllabification. The variable (four-item) sets can only be partially prepared for articulation (relative to the heterogeneous baselines), as can be concluded from the results of the two control-experiments. While advanced phonological encoding of the segmental information can occur, the absence of predictive syllabic structure prevents completion of phonetic encoding, namely, the retrieval of the corresponding gestural score from the mental syllabary.

An additional component of the observed effects may reside in the residual activation of the syllables in the mental syllabary. The retrieval time for the one gestural score required for the constant set may be faster due to the repeated selection of the same representation for all items within a set. But in variable sets, residual activation can only speed retrieval of the syllable half of the time (both the odd-man-out and the item immediately following the odd-man-out cannot benefit from the residual activation). While this may account for some of the observed differences, post-hoc analyses demonstrated that this is not the whole story. If only items which share the initial syllable with the immediately preceding trial are considered (thus allowing for a benefit of residual activation), the difference between the constant and variable sets remains. The remaining items in the constant four-item sets still show faster reaction times (633 ms, Experiment 2.1; 626 ms, Experiment 2.2) than the remaining items in the variable four-item sets (676 ms, Experiment 2.1; 670 ms, Experiment 2.2). Thus, even when the differences in residual activation of syllable representations in the mental syllabary are maximally matched, a syllable preparation effect is still observed. This post-hoc analysis demonstrates that all items in the constant sets benefit from the syllable overlap while none of the items in the variable do.

Using the implicit priming paradigm we successfully identified syllables as functional units in speech production. However, why was this task successful in finding a syllable effect when a decade of syllable priming studies failed? We would like to argue that the crucial difference lies in the explicit articulation of each word. Note that in all prior syllable priming studies, the prime was never overtly produced. If indeed syllables emerge late in speech production, primes that are not articulated may not reach the relevant stage in production where syllables are encoded, resulting in no priming effect.

Thus, our results confirmed the predictions made by WEAVER++. This model predicted that the two different tasks (priming versus preparation) affect different aspects of the process of phonological encoding. In syllable priming studies, the primes are assumed to speed up the segmental spell out, i.e., the moment when segments become available as part of the stored word form, which is retrieved from the mental lexicon. At the stage of segmental spell-out there is only segmental, but no syllabic information available according to the theory of lexical access proposed by Levelt et al. (1999; but see Dell, 1986, 1988). There can be no primed syllable retrieval. The finding that the magnitude of the priming effect increases with an increase of the number of shared segments, independent of a syllable match or mismatch with the target's first syllable, confirms the assumption that only shared segments can be primed. Actual syllable information such as syllable-internal positions for the current segments, i.e., which segment is assigned to the onset, the nucleus, or the coda position, only becomes available when phonological syllables induce the retrieval of the corresponding phonetic syllables from the mental syllabary.

The implicit priming paradigm allows for maximum preparation given the item set. In constant sets, where initial segments and also the syllabic structure of the first syllable are shared between items in one response set, the first syllable is fully prepared for articulation. Thus, all stages prior to articulation, including segmental spell-out, on-line syllabification and possibly access to the mental syllabary, can contribute to the preparation effect. In variable sets, responses can be prepared up to on-line syllabification or prosodification. Thus, all stages preceding on-line syllabification contribute to the preparation effect. In the current task, syllabic information is relevant in the sense that any deviating syllabic structure, i.e., an item with a different syllable structure, reduces the preparation effect. This account is confirmed – at least for a language like Dutch – by the results of the present study. While Dell's model can account for the results presented here, it can not account for the absence of syllable priming effects reported repeatedly in the literature (Schiller, 1997, 1998, 1999, 2000; Schiller et al., 2002).

However, as argued above, the assumption that syllabification is a late process may not hold cross-linguistically. In a recent study, Chen et al. (2002) investigated the process of word-form encoding in Mandarin Chinese by means of the (standard) implicit priming paradigm. In their experiments, they tested whether they could find preparation effects when the first syllable was shared, with or without shared tone. Indeed, they obtained preparation effects in both conditions, though the preparation effect was substantially stronger in case the tone was shared. The latter interaction is in line with findings reported by Roelofs and Meyer (1998). They tested for Dutch whether the standard first syllable priming effect requires that the target words share their stress pattern. Using iambic words (the non-default stress pattern in Dutch), they showed that an odd-man-out item (with trochaic stress pattern) entirely annihilated the preparation effect. This is in line with the Chinese data, except that if tone is not shared there is still a small preparation effect remaining. Chen et al. (2002) rightly raise the question what is exactly prepared when syllables are primed that do not share their tones. It cannot be the syllable's full articulatory gesture, because that must involve the tone. This can be constructed as another argument for whole syllable retrieval in Chinese. The remaining preparation effect would then be due to facilitation of retrieving a syllabified phonological code. Notice that this presupposes that in Mandarin Chinese syllables stored with the lexical item are not specified for tone; an item's tone pattern is, in some way, independently specified in lexical form memory. The retrieved tone pattern gets assigned to the appropriate phonological syllables during incremental syllabification. This is all highly speculative, but it shows that it is worth exploring the mechanism of syllabification in much more depth cross-linguistically.

To summarize, the different results found in Mandarin Chinese and Dutch (as well as in other Indo-European languages) could be due to the contrasting properties of the respective languages. In Mandarin Chinese, with a syllabary inventory of a much smaller size and no need for resyllabification, the syllabification process may be different in the sense that syllabic units are represented earlier in the process of speech production.

CONCLUSIONS

The results of the odd-man-out variant of the implicit priming task reported in this study show specific preparation of syllabic articulatory units is possible. Taking the results of prior syllable priming studies into account, we conclude that the used paradigm taps into the right level of processing where syllables are in fact encoded, i.e., the interface of phonological and phonetic encoding. This is in agreement with predictions from the spoken word production model by Levelt et al. (1999) and its computer simulation, WEAVER++ (Roelofs, 1996, 1997a, 1997b, 1998, 1999). The results from the Mandarin Chinese study (Chen et al., 2002) do not contradict the proposed syllabification process in Dutch, but rather suggest that a language user's word form encoding architecture will, at least in part, be tuned to specific requirements of target language phonology.

EFFECTS OF SYLLABLE FREQUENCY IN SPEECH PRODUCTION

CHAPTER 3

(An adapted version of this chapter has been submitted for publication as Cholin, J., Levelt, W. J. M., & Schiller, N. O., 'Effects of syllable frequency in speech production')

ABSTRACT

In the speech production model proposed by Levelt et al. (1999), syllables play a crucial role at the interface of phonological and phonetic encoding. At this interface, abstract phonological syllables are translated into phonetic syllables. It is assumed that this translation process is mediated by a so-called *Mental Syllabary*. Rather than constructing the motor programs for each syllable on-line, the mental syllabary is hypothesized to provide precompiled gestural scores for the articulators. In order to find evidence for such a repository, we investigated syllable frequency effects: If the mental syllabary consists of retrievable representations corresponding to syllables, then the retrieval process should be sensitive to frequency differences. In a series of experiments using a symbol-position association learning task, we tested whether high-frequency syllables are retrieved and produced faster compared to low-frequency syllables. We found significant syllable frequency effects with monosyllabic pseudo-words and disyllabic pseudo-words in which the first syllable bore the frequency manipulation; no effect was found when the frequency manipulation was on the second syllable. The implications of these results for the theory of word form encoding at the interface of phonological and phonetic encoding, especially with respect to the access mechanisms to the mental syllabary in the speech production model by Levelt et al. (1999) are discussed.

INTRODUCTION

The aim of the present chapter is to provide evidence for the assumption that speakers access a separate mental store containing pre-compiled motor-programs of syllabic size. The notion of such a *Mental Syllabary* that is accessed during speech production is an inherent part of the theory of spoken word production by Levelt, Roelofs and Meyer (1999) and is also implemented in its computer simulation WEAVER++ (Roelofs, 1997a, 1997b, 1998, 1999).

In fluent speech, the individual sounds of a word are bundled together to form optimally pronounceable units, namely syllables, which serve as the basis for motor execution. The tens of thousands of words in any given language are composed of a relatively small inventory of syllables. In fact, 500 syllables from English, Dutch, and German suffice to produce approximately 80% of all speech in those languages (Schiller, Meyer, Baayen, & Levelt, 1996). To account for the easy and rapid production of these frequently occurring syllables, Levelt and Wheeldon (1994) proposed a repository of ready-made syllabic motor programs, called the *Mental Syllabary*.

The idea of a phonetic store of syllables was originally proposed by Crompton (1981) and adopted in Levelt (1989). To account for certain types of speech errors (e.g., phoneme and syllable substitutions and blends), Crompton suggested that speakers retrieve syllable-sized programs from a library of articulatory routines. However, despite the general agreement that syllables constitute an important linguistic and phonological/ phonetic unit (see Blevins, 1995; Fujimura & Lovins, 1978; Hooper, 1972; Kenstowicz, 1994; Selkirk, 1982) and the off-line support from speech error data and meta-linguistic tasks (Fromkin, 1971; Shattuck-Hufnagel, 1992; Treiman & Danis, 1988), there is to date relatively little psycholinguistic on-line evidence that the syllable serves as a functional unit in speech production. This also holds for the data reported in Levelt and Wheeldon (1994) – see below. Furthermore, what evidence there is for the functional importance of the syllable is orthogonal to the issue of whether or not there is mental storage for precompiled syllabic programs.

First, we introduce the process of word-form encoding in the framework of the model proposed by Levelt et al. (1999) and its computer simulation WEAVER++ (Roelofs, 1997a, 1997b, 1998, 1999) with particular attention to the encoding steps at the interface of phonological and phonetic encoding and crucially how access to the mental syllabary is involved. Second, an overview of the experimental evidence for syllabic representation in speech production and stored syllabic units at different processing levels is presented.

PHONOLOGICAL AND PHONETIC ENCODING

As soon as a speaker has selected a target word from the mental lexicon, phonological encoding can be initiated. It largely consists of computing the word's syllabification and prosody in the larger context of the utterance. This, in turn, is incrementally followed by phonetic encoding, which includes the computation of the articulatory gestures for the target word's syllables in their phonetic context. Finally, the execution of these gestural scores by the laryngeal and supralaryngeal muscle systems produces the spoken word. The present chapter exclusively concerns the stages of phonological and phonetic encoding.

Phonological Encoding

The first operation in phonological encoding is the retrieval of the target word's *phonological code* from the mental lexicon. The code consists of an ordered set of phonemic segments. For stress-timed languages such as English and Dutch the model also assumes the existence of sparse metrical markers in phonological codes. In these languages, the default position for lexical stress is defined as the first full-vowel syllable of a word. Thus, for polysyllabic words that do not have the stress on the first stressable syllable, the metrical structure is stored as part of the phonological code; but, for monosyllabic words and for all other polysyllabic words, it is not stored but computed. Metrical structures describe abstract groupings of syllables (σ) into feet (Σ) and feet into phonological words (ω). A phonological word was defined as the minimal unit above the metrical foot, to which clitics, such as unstressed words, can attach (as in "she's" where the reduced form of 'is' attached to 'she', see Levelt, 1989; Wheeldon & Lahiri, 1997). Crucially, at the stage of phonological encoding, phonological segments are not yet assigned to syllabic positions, nor is the C(onsonsant)-V(owel) (hereafter CV-) structure for the word specified. This is in contrast to other models of spoken word production (in particular Dell, 1986, 1988), which assume that the retrieved phonological codes are pre-syllabified. Internal syllabic positions such as onset, rhyme and coda are pre-specified in Dell's model but not in the Levelt model. The main argument for not pre-determining syllable positions in the phonological codes stored in the lexicon results from the phenomenon of *resyllabification*. In connected speech, syllable boundaries often differ from a word's or morpheme's canonical syllabification. The domain of syllabification is the phonological word, which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticization (Booij, 1995). If, for instance, the stored phonological code for the word *defend* would be syllabified (i.e., as *de-fend*), then the speaker must 'resyllabify' the word when used in a different context, such as the past tense (*de-fen-ded*) or cliticization (*defend it* - *de-fen-dit*). The ubiquity of such 'resyllabifications' in the normal use of English (or Dutch for that matter) renders pre-specification of segments to syllable positions highly inefficient.[5] The alternative assumption, therefore, is that a word's syllabification is not retrieved but computed on-line depending on the context in which the word appears. During this process, called 'prosodification', retrieved segments are incrementally combined to form successive syllables. Also, these successive syllables are incrementally assigned the appropriate metrical properties, either following default stress, or otherwise the retrieved non-default stress marking feature. The incremental composition of syllables follows, on the one hand, universal syllabification constraints (such as maximization of onsets and sonority gradations) and, on the other hand, language-specific rules, e.g., phonotactics. Together, these rules create well-pronounceable syllables. The output of phonological encoding is a phonological word, specified for its metrical, syllabic, and segmental properties.

Dell's (1986, 1988) model makes different assumptions with respect to the phonological encoding process. As already mentioned above, his model includes abstract phonological representations that are specified for internal syllabic positions, i.e., the word form retrieved from the mental lexicon activates not only segmental information but also syllabic frames. These syllabic frames serve as placeholders into which the retrieved segments are inserted during the process of segment-to-frame-association. The discussion

---

[5] The claim that phonological codes are not pre-syllabified is, in part, a language-specific claim. For a language like Mandarin Chinese, which has a small set of syllables and limited resyllabification processes, the story might be different (see, Chen, Chen, & Dell, 2002).

of whether or not the syllabic structure is part of word form encoding will be revisited under the section ‘Empirical evidence’.

Phonetic Encoding

The output of the phonological encoding process consisting of fairly abstract, syllabified phonological words is incrementally translated into articulatory-motor programs. These programs consist in large part of specifications for subsequent syllabic gestures.

A crucial assumption of Levelt’s theory is that speakers have access to a repository of syllabic gestures. This repository, coined the ‘mental syllabary’ (Levelt, 1992; Levelt & Wheeldon, 1994), contains the articulatory scores for at least the high-frequency syllables of the language. The model assumes that as soon as a syllable emerges during incremental syllabification, the corresponding syllabic articulatory gesture will be selected from the repository in Broca's area or a pre-motor area (Dronkers, 1996; Indefrey & Levelt, 2000; Kerzel & Bekkering, 2000).

The output of the mental syllabary in turn serves as input to further phonetic encoding, at which time contextually-driven phonetic fine-tuning of retrieved motor programs occurs: The motor programs are still rather abstract representations of the articulatory gestures which have to be performed at different articulatory tiers, a glottal tier, a nasal tier and an oral tier. The gestural scores are abstract in the way that their performance is highly context-dependent (due to allophonic variation, coarticulation and - as a result of this - assimilation). The actual details of the movements in realizing the scores, such as lip protrusion and jaw lowering, is within the domain of the articulatory system (Goldstein & Fowler, 2003). According to Levelt (1989), the stored syllable can be pronounced with more or less force, with shorter or longer duration, and different kinds of pitch movements. These are free parameters, which have to be set from case to case. For new or very low-frequency syllables it is proposed that articulatory plans are assembled using the segmental and metrical information specified in the phonological syllables. This route[6] can also be used for the assembly of high-frequency syllables, for instance in cases when more conscious on-line control of speech production is required. For example, when

---

[6] The terminology of ‘indirect’ or ‘direct’ route to refer to the computational process that bundles phonetic segments to form syllabic scores was avoided as there is no consistent use of these terms in the literature.

speakers give a lecture or self-correct *slips-of-the-tongue* or speech errors in general. However, the retrieval of gestural scores from the syllabary is usually faster and less error-prone.

In the final step, the articulatory network, a coordinative motor system that includes feedback mechanisms (Saltzman, 1986; Saltzman & Kelso, 1981; Goldstein & Fowler, 2003) transforms these articulatory plans into overt speech.

PREDICTIONS BY WEAVER++

WEAVER (*W*ord-form *E*ncoding by *A*ctivation and *VER*ification) is the spreading activation based computer network model developed by Roelofs (1992, 1996, 1997a, 1997b, 1998, 1999, 2002), which is based on Levelt's (1989, 1992) theory of speech production. WEAVER++ adopts Dell's (1986) assumption of word form retrieval by the spread of activation and Levelt's (1992) on-line syllabification and access to a syllabary (Levelt & Wheeldon, 1994).

WEAVER++ (Roelofs, 1997a, 1997b, 2002) assumes that the syllabification of a word is computed on-line during the speech production process. In the WEAVER++ model, segments in the retrieved phonological code are not specified for their syllable position, but only for their serial order within a word. The actual syllabic position of a segment is determined by the syllabification process. Each retrieved segment in the phonological code spreads activation to all syllabic gestures in which it partakes. Hence, upon retrieval of a phonological code, there are always multiple phonetic syllable programs in a state of activation. How is the appropriate syllable program selected? There are, first, selection conditions. The crucial one is that the syllable matches the phonological syllable that is incrementally composed; this involves a procedure of verification. Second, each syllable in the syllabary has a frequency-dependent selection threshold. This causes the predicted syllable frequency effect on naming latencies. Notice, however, that the threshold assumption is a 'modular' one. Removing it does not affect the architecture of the system as a whole. Third, syllable selection is subject to Luce's (1959) choice rule. During any smallest interval, the probability of selecting the (verified) target syllable equals the ratio of its activation to the summed activation of all syllable nodes. Given the choice ratio, the expected selection latency can be computed.

EMPIRICAL EVIDENCE FOR SYLLABIC UNITS IN SPEECH PRODUCTION

The question of whether or not the syllable constitutes a relevant unit in speech production has been the subject of many psycholinguistic off- and on-line studies. Many of the off-line studies, such as speech error analyses, suggested that phonological speech errors could be explained to a large extent within a syllabic framework. It was often argued that segmental speech errors such as exchanges of segments only occur for identical syllable internal positions, i.e., onsets exchange with onsets, nuclei exchange with nuclei, etc. (Berg, 1988; MacKay, 1970; Nooteboom, 1969; Shattuck-Hufnagel, 1979; Stemberger, 1982); this is referred to as the *syllable position constraint*. A quantitative analysis, however, showed that the majority of such errors take place in the onset position. Thus, the syllable onset constraint may be a word-onset constraint (Shattuck-Hufnagel, 1987, 1992; see Meyer, 1992, for a critical review). Another source of off-line evidence are meta-linguistic tasks, which suggest that syllables play a role at some level of processing in speech production (Schiller, Meyer, & Levelt, 1997; Treiman, 1983; Treiman & Danis, 1988; see Bagemihl, 1995 for a review) but no strong claims are made about which level is involved. Despite the (limited) off-line support for the syllable and the relevance of syllables to linguistic phenomena, the support from the on-line studies over the past decade is rather scarce.

A series of cross-linguistic studies used a syllable priming task to identify the syllable as a relevant unit in speech production (for Dutch: Baumann, 1995; Schiller, 1997, 1998; for Mandarin Chinese: Chen, Lin, & Ferrand, 2003; for French: Brand, Rey, & Peereman, 2003; Evinck, 1997; Ferrand, Segui, & Grainger, 1996; Schiller, Costa, & Colomé 2002; for English: Ferrand, Segui, & Humphreys, 1997; Schiller, 1999, 2000; Schiller & Costa, submitted; for Spanish and an overview see Schiller et al., 2002).

In all these studies, a syllabic prime was given which was either congruent with the target's syllabic structure (e.g., *ba.* as prime for *ba.lade* or the prime *bal.* for *bal.con;* a dot indicates syllable boundaries) or was incongruent (e.g., *ba.* as a prime for *bal.con* or the prime *bal.* for *ba.lade*). However, rather than demonstrating a syllable priming effect, each of these studies found a segmental overlap effect, i.e., phonologically related primes, whatever their syllabic relation to the target word, facilitated the response relative to unrelated control primes, with longer CVC-primes being more effective than CV-primes.

Moreover, no specific syllable priming effects were obtained[7]. The original segmental overlap effect (Schiller, 1998) has recently been specified in more detail (Schiller, in press).

Several subsequent studies have failed to find a syllable priming effect but rather confirmed the finding of a segmental overlap effect. The converging outcome of all of the syllable priming studies points to the conclusion that, at the stage where the (explicit) priming taps into the speech production process, there is only segmental but no syllabic structure available. If one considers that the primes in the syllable-priming task never get overtly articulated, one can imagine that these syllabic primes may not reach the late stage where syllables are computed on-line. Instead, the primes are assumed to speed up the segmental retrieval from the mental lexicon. At that stage of segmental spell-out there is only segmental, but no syllabic information according to Levelt et al. (1999) (but see Dell, 1986, 1988 and also Costa & Sebastián-Gallés, 1998; Sevald, Dell, & Cole, 1995; Ashby & Rayner, in press). The finding that the magnitude of the priming effect increases with an increase in the number of shared segments, independent of a syllable match or mismatch with the target's first syllable, confirms the assumption that only shared segments can be primed.

The conclusion that the explicit priming paradigm is not appropriate to detect syllabic representations in speech production made it necessary to find one which taps into later stages of word form encoding. A paradigm which is known to be sensitive also to later stages of word form encoding is the "*implicit* priming paradigm": Whereas the explicit priming is sensitive only to early stages of phonological encoding, the implicit priming paradigm exhibits effects that emerge at these early stages but also comprise later stages at the interface of phonological and phonetic encoding, i.e., on-line syllabification, possibly including syllabary access (Meyer, 1990, 1991; Roelofs, 1996, 1998; Roelofs & Meyer, 1998).

In a recent study, Cholin, Schiller, and Levelt (2004) used an *odd man out* variant (Janssen, Roelofs, & Levelt, 2002) of the implicit priming technique to demonstrate that

---

[7] Much discussion has been given to the results of the apparent syllable priming effect in French (Ferrand et al., 1997, Experiment 5). However, Brand et al.'s (2003) failure to replicate the Ferrand effects suggests that this should not be taken as strong evidence for the syllable effect (see also Evinck, 1997; and for a review Schiller et al., 2002).

speakers can benefit from a shared syllable structure. In this study, subjects learned sets of prompt-target pairs. Two types of response sets were compared, namely constant and variable response sets: Constant sets had overlapping initial segments *and* a constant CV-structure (as in **spui**.en, *to drain*; **spui**.de*, drained;* **spui**.er*, person who drains*; **spui**.end*, draining).* Variable sets had an overlap of the first segments but did not have a constant syllable structure (e.g., **spoe**.len, *to rinse*; **spoel**.de, *rinsed*; **spoe**.ler, *person who rinses*; **spoe**.lend, *rinsing*). Note that the second item of this set shares the same initial segments but has a different initial syllable structure; it is the *odd-man out* for the set. The prediction was that – under the assumption that the syllable is a relevant processing unit in speech production and is stored and accessed independently during word form encoding – speakers need knowledge about the current syllabic structure in order to prepare for a target utterance. In two studies investigating two different CV-structures (CVV, Exp. 2.1; CCVV, Exp. 2.2), Cholin et al. found a significantly larger preparation effect for constant sets as compared to variable sets. The authors argue that this preparation effect, which cannot be attributed to a segmental overlap effect, offers strong evidence for the relevance of syllables in word form encoding.

The virtue of the implicit priming paradigm is that it taps into later stages of word-form encoding than explicit priming paradigms do, but its shortcoming is that it is sensitive to all processing stages prior to articulation, including segmental retrieval, on-line syllabification and possibly access to the mental syllabary. Thus, the above effect could arise at any one of these stages or, possibly, at a combination of these stages. These results do not necessarily imply access to the hypothesized mental syllabary. Effects that would certainly provide evidence for the existence of the mental syllabary are effects of syllable frequency because only the retrieval of stored units produces frequency effects.

It should, finally, be noticed that Dell's model can account for the results of the implicit priming studies. However, it cannot handle the above-mentioned absence of syllable priming effects (Schiller, 1997, 1998, 1999, 2000; Schiller et al., 2002).

EVIDENCE FOR THE MENTAL SYLLABARY AND SYLLABLE FREQUENCY STUDIES

As already stated, the mental syllabary is conceived as a store for abstract motor routines of syllabic size. The economic advantage of such a repository becomes apparent if one considers the computational load that would evolve when the single sub-syllabic or phonemic motor-patterns had to be computed anew each time in speech production. The above mentioned analyses by Schiller et al. (1996) demonstrated that speakers re-use a rather small set of syllables over and over again, even for languages such as English, German or Dutch which have more than 10.000 different syllables. It would be efficient to store these overlearned syllabic patterns and retrieve them as wholes, instead of computing them anew time and again as they appear during syllabification.

Levelt and Wheeldon (1994) studied this storage hypothesis by comparing access latencies of high- versus low-frequency syllables. They had subjects produce disyllabic words consisting of high- versus low-frequency syllables. Participants learned to associate symbols with written target words (e.g., ###### = koning *king*). During the test phase the learned symbols were repeatedly presented in random order and each time participants responded with the associated target word.

Levelt and Wheeldon's core finding was that, when word frequency was controlled for, words consisting of high-frequency syllables were named faster than words consisting of low-frequency syllables. If all syllables are computed on-line rather than retrieved from a repository, their frequency of use should be irrelevant. The obtained syllable frequency effects therefore seemed to support the notion of the mental syllabary. However, a potential problem with this conclusion is that syllable frequency was correlated with segment frequency in some of Levelt and Wheeldon's experiments (Hendricks & McQueen, 1996).

Evidence for the assumption that syllables are retrieved as whole entities come from recent studies in Spanish and French. Carreiras & Perea (in press) reported syllable frequency effects in Spanish speech production tasks (see also Perea & Carreiras, 1996, 1998). Brand, Rey, Peereman, and Spieler (2002) found also significant syllable frequency effects in a larger-scale study. Laganaro (2003) investigated effects of syllable frequency in French from a psycho- and neurolinguistic perspective. In her psycholinguistic studies, she found that syllable frequency affected naming latencies in a picture-naming experiment (see also Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, in press) and the reaction times in non-word reading. In her neurolinguistic study, she obtained a significant frequency effect on segmental substitution errors and a syllable frequency effect on non-word repetition performance in aphasics. Furthermore, Aichert and Ziegler (2004) found a syllable frequency effect in patients with apraxia of speech (but see Wilshire & Nespoulous, 2003 and Howard & Smith, 2002). These studies are supportive of the notion of a mental syllabary.

To summarize, a speech production model was outlined that incorporates syllables as functional units in word form encoding. We presented empirical evidence supporting this assumption and moreover, we could make a good case for the stage at which syllables come into play. It was argued that the absence of syllable priming effects on the one hand and the positive syllable preparation effects on the other hand suggest that (phonologically abstract) syllables are not stored units (as argued by Dell, 1986, 1988) but rather generated by an on-line syllabification process.[8] Evidence that these abstract phonological syllables trigger the selection of their phonetic representations from a mental syllabary is still rather scarce. As already mentioned, the results by Levelt and Wheeldon (1994) have to be handled with care as some of the syllable frequency effects were confounded with segment frequency. The syllable frequency effects in Spanish (Carreiras & Perea, in press; Perea & Carreiras, 1996, 1998) and also the positive results from French (Brand et al, 2002) and German (Aichert & Ziegler, 2004) are promising but have to be seen before the background that syllables in Romance languages might have a different processing status than syllables in Germanic languages such as Dutch, German, and English. Romance languages (or *syllable-timed* languages, Abercrombie, 1967; for psycholinguistic evidence see Cutler, 1997; Cutler, Mehler, Norris, & Segui, 1986) have syllables as their primary rhythmic units. Germanic languages (or *stress-timed* languages) have a stress-based rhythm and are less prone to use the syllable as primary segmentation unit in speech perception (however, see Zwitserlood, Schriefers, Lahiri, & van Donselaar, 1993, but also Vroomen & de Gelder, 1994; for a critical, cross-linguistic review see Cutler, McQueen, Norris, Somejuan, 2001).

The question of whether or not speakers can access pre-compiled syllabic units from a repository is still under discussion. The aim of Chapter 3 is to gather additional evidence for the existence of such a mental syllabary.

---

[8] Note that the idea of a mental syllabary is also compatible with Dell's syllabified phonological code, the existence of stored units does not deny a store for pre-compiled syllabic programs.

SYMBOL - POSITION ASSOCIATION LEARNING TASK

A Symbol – Position Association Learning Task was used to investigate effects of syllable frequency, a technique which was comparable to the one used in the Levelt and Wheeldon (1994) study. However, in the Levelt and Wheeldon (1994) study, participants had to associate sets of four strings consisting of six non-alphabetic characters (as "%%%%%%" or "&&&&&&") to four written disyllabic Dutch words (containing high- and/ or low-frequency syllables). In the present experiments, participants had to associate an auditorily presented word to one of four positions on the screen (two positions in the final experiments). In learning phases, participants were presented with a symbol, namely an icon of a little loudspeaker, in one of four potential positions and were simultaneously presented with the to-be-associated word via headphones.

In production phases, i.e., practice and test phases, the symbol was presented either on the left, the right, the top or the bottom of the screen to prompt the previously associated target. One of the main advantages of this technique compared to the one applied by Levelt and Wheeldon (1994) lies in the presentation of the spoken target item instead of a printed word during learning phases. The auditory presentation of the target ensures that potential confounds deriving from orthographic factors (e.g., grapheme frequency) can be excluded.

In order to test whether this method is at all sensitive to frequency effects, a pre-test was conducted that tested the production latencies of high-frequency words compared to low-frequency words. The fact that high-frequency words are produced significantly faster than low-frequency ones is a stable finding in the literature (e.g., Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965). If a positive frequency effect can be obtained for words, then this method should also be appropriate to detect *syllable* frequency effects.

PRE-TEST INVESTIGATING THE PRODUCTION OF HIGH- AND LOW-FREQUENCY WORDS IN THE SYMBOL-POSITION ASSOCIATION LEARNING TASK

A pre-test was run with high- and low-frequency disyllabic Dutch words.

METHOD

Participants. Twenty-four native speakers of Dutch participated in the Experiment. They were randomly taken from the pool of participants of the Max Planck Institute in Nijmegen, The Netherlands and were paid for their participation. They had no known hearing deficit, and they had normal or corrected-to-normal vision.

Materials. 32 disyllabic Dutch words, 16 high- and 16 low-frequent were used. High- and low-frequency counterparts were chosen that had matching number of segments, the same syllabic structure and optimally a maximum number of identical segments. High-frequency sets had a mean word-form frequency of 194.25 (SD = 4.35) and for low-frequency sets with a mean word-form frequency of 0.93 (SD = 0.15) per one million words. A full list of items is given in the Appendix 3-A.

Procedure and Apparatus. The experiment consisted of alternating learning, practice and test phases. The participants were tested individually in a quiet room. They were given a detailed written instruction specifying that they had to respond as accurately and as quickly as possible.

In the learning phases, icons of little white loudspeakers (4 by 4 cm) were presented on four positions on a black computer screen (iiyamaLM704UT). The icons consisted of interactive fields presented on a right, left, upper, and lower position. By clicking with the mouse (cordless Wheel Mouse, Logitech) on one of the interactive fields, the target-word corresponding to this position was presented auditorily via headphones (Sennheiser HD250). Participants were asked to explore and to memorize the association of target-words to their positions by clicking with the mouse on the different positions in a self-paced fashion. When they indicated that they had studied the associations sufficiently, the experimenter started the practice phase.

In practice phases, one out of the four positioned icons was presented on the computer screen and participants were required to produce the corresponding target-word as fast and as accurately as possible. Each position was presented three times. When participants responded erroneously or a time out occurred (RTs > 1500 ms), the experimenter started the learning phase again and the session was rehearsed to ensure that they learned the sets accurately. When participants successfully completed the following practice phase, the experimenter started the test phase.

The structure of the practice and the test phases was as follows: The symbol was presented on one of the four positions and disappeared after the response with a delay of 500 ms. Each of the four target items was presented eight times in pseudo randomized order, resulting in a response block of 32 items. The experiment started with a short practice session of one set, practice items mimicked the experimental materials, in which participants could get acquainted with different phases and the experimental setting.

The presentation of the stimuli and the measuring of the reaction times were controlled by the NESU software package. The spoken reactions were registered by a Sennheiser MD211N microphone, which fed into a NESU-box voice key device and a DAT recorder (Sony DTC-55ES). The experimenter sat in the same room as the participant and took note of hesitations, voice key errors, wrong naming responses, and time outs. On average, an experimental session lasted 30 minutes.

Design. The two-level variable *Frequency* (high-frequency versus low-frequency) was tested within participants and between items. Each participant produced each of the 32 items 8 times during one experimental block, resulting in a total of 256 experimental trials. Within each experimental block, sets of four frequency-homogeneous items, that is, four high-frequency or four low-frequency items were presented as one set. We presented high- and low-frequency items in separate rather than in mixed sets as effects are more likely to arise in homogeneous or pure sets than in heterogeneous sets (see Meyer, Roelofs & Levelt, 2003 and also Lupker, Brown & Colombo, 1997). In total, there were eight sets, four high-frequency sets and four low-frequency sets. It was taken care of that the four items within one set did not have segmental overlap, were phonologically as distinct as possible and did not have any semantic relation to each other. High- and low-frequency sets always consisted of high- and low-frequency item counterparts, see Table 3-1 for an example, to have maximum comparability between high- and low-frequency item sets.

Table 3-1. *Example of a corresponding high- and a low-frequency experimental set.*

Experimental Set	
high-frequency	low-frequency
wa.ter (water)	wa.sem (steam)
ra.pport (report)	ra.dijs (radish)
ta.fel (table)	tan.dem (tandem)
fo.to (photo)	fa.kir (fakir)

The order of the item sets were as follows: High- and low-frequency sets alternated. The succession of the high- and low-frequency sets was counterbalanced over 24 participants by applying a Latin-Square design. The order of the second four experimental sets was always a repetition of the first four sets by reversing the frequency-assignment to ensure a maximum distance between the two corresponding high- and low-frequency sets (e.g., A_{high}, B_{low}, C_{high}, D_{low}, & A_{low}, B_{high}, C_{low}, D_{high}). There was a restriction of the immediate repetition of the same item within sets. The assignment of a specific item to one of the four positions on the screen rotated between participants, thus, overall every item was presented on every position.

RESULTS

Observations leading to wrong or invalid responses (mispronunciations, voice key errors, and hesitations) were not included in the reaction time analysis. Time outs (RTs > 1500 ms) and reaction times shorter than 300 ms were also removed. Observations deviating from a participant's and an item's mean by more than two standard deviations were considered as outliers and also discarded from the reaction time analysis. 274 (4.5 %) trials were treated as errors and 129 (2.1%) as outliers.

The mean of the high-frequency items and the mean of the low-frequency items were submitted to *t*-tests. Two complementary analyses were computed, one treating participants (t_1) and one treating items (t_2) as random factor (Clark, 1973).

The mean voice onset latencies, standard deviations and error rates for this experiment are summarized in Table 3-2.

Table 3-2. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) for the pre-test.*

Frequency	M	(SD)	% Err	(SD)
High	694	(82)	4.5	(3.8)
Low	731	(100)	4.5	(3.5)
Difference scores	-37		0.0	

The analysis of reaction times showed that high-frequency monosyllabic pseudo-words were produced significantly faster than low-frequency monosyllabic pseudo-words ($t_1(23) = 4.821$, $p < .001$; $t_2(30) = 3.327$, $p < .01$). The analysis of errors showed no significant difference (both *t*s < 1).

DISCUSSION

The pre-test shows a clear word-frequency effect. This result can be taken as replication of the standard finding that high-frequency words are produced faster than low-frequency words, which reflects faster retrieval processes for high-frequency words from the mental lexicon. As this Symbol-Position Association Learning Paradigm seems to be sensitive to *word* frequency it seems very likely that it is also sensitive to *syllable* frequency, even though the two effects are assumed to arise at two different levels of processing. Thus, the next experiment will investigate effects of syllable frequency.

EXPERIMENT 3.1: MONOSYLLABIC PSEUDO-WORDS

In this experiment, the production of high-frequency versus low-frequency syllables is investigated. The high-frequency monosyllabic pseudo-words have all the same CV-structure, namely CVC. The outcome that high-frequency items are produced faster than low-frequency items would support the notion of a mental syllabary.

METHOD

Participants. Sixteen native speakers of Dutch participated in the Experiment. They were randomly taken from the pool of participants of the Max Planck Institute in Nijmegen, The Netherlands and were paid for their participation. They had no known hearing deficit, and they had normal or corrected-to-normal vision.

Materials. All frequencies counts were obtained from the computer database CELEX (CEntre for LEXical Information; Baayen, Piepenbrock, & Gulikers, 1995), which has a Dutch lexicon based on 42 million word tokens. Syllable frequency was counted for phonetic syllables in Dutch. The phonetic script differentiates the reduced vowel schwa from full vowel forms, giving approximately 12,000 individual syllable types. Syllable frequencies were calculated for the database from the word form occurrences per one million. Two syllable frequency counts were calculated: The number of occurrences of each syllable (independent of the frequency of occurrence of the syllable in a particular word position, i.e., first or second syllable position within a word) and the number of the summed frequency of occurrence of each syllable (within words). The syllable frequency ranges from 0 to approximately 90,000 per one million words, with a mean frequency of 121.[9]

For the following experiments, the experimental high-and low-frequency items should only be different in their syllable frequency. Therefore, it was crucial to construct an experimental item set that was controlled for length in phonemes, phoneme frequency, CV

---

[9] The CELEX lexical database includes a list of Dutch syllables and their frequencies, based on syllabification of isolated word forms. In connected speech, as discussed above however, context-dependent phonological rules can modify the syllables and accordingly their token frequency. Schiller et al. (1996) carried out an empirical investigation in order to estimate the changes syllables may undergo in connected speech. A large Dutch newspaper corpus (TROUW) was transcribed, word-level rules were applied and then syllabified. In a first step the resulting *lexeme* syllables were compared to the corresponding entries in the CELEX database. In a second step, additional phonological sentence-level rules were applied to the TROUW corpus and then the frequencies of the resulting *connected-speech* syllables were compared to the lexeme syllables again. The overall correlation between lexeme and connected speech syllables was very high.

structure and bigram frequency. We decided to administer only *CVC*-syllables to the list of experimental items for the following reasons: (i) the CV-structure *CVC* is one of the most frequent syllable structure in Dutch[10] and (ii) CVC syllables provided for the best match on various features between the high- and low-frequency sets. Furthermore, the CVC-syllables could fulfill all of the criteria listed below.

We applied the following search criteria to the list of syllables: For a given high-frequency syllable a low-frequency counter part should be found that was different from the given syllable only in one segment. For example, if the high-frequency syllable is *kem* [kɛm], the corresponding low-frequency syllable should be different by only the last segment, thus, a possible counterpart is *kes* [kɛs]. Thus, we have one pair of syllables which has the same onset and the same nucleus, namely the same short vowel. The only difference lies in the deviating coda, that is, the final consonant differentiates the two high- and low-frequency syllables. The next search device was to take those two syllables and to look again for counterparts but this time for each syllable in the opposite direction: For the high-frequency syllable *kem* [kɛm] this involved looking for a low-frequency syllable that had an identical offset but a different onset. The last two positions should be matching. The low-frequency syllable *wem* [ʋɛm] has an offset-overlap with the high-frequency syllable *kem* [kɛm] and was taken as third member in the syllabic quadruple. The final step was to look for a high-frequency syllable that had a) the same onset as the low-frequency syllable *wem* [ʋɛm] and b) the same offset as the other low-frequency syllable *kes* [kɛs]. A successful quadruple was constructed when a high-frequency syllable was found which fulfilled this criterion, which in this case was the high-frequency syllable *wes* [ʋɛs]. See Table 3-3 for a schematic depiction of the described quadruple.

---

[10] In the inventories (CELEX and TROUW) the three most frequently occurring syllable types are, in order of frequency, CVV, CVC and CVVC, together accounting for more than 70% of all syllables. For more details and exact values of the two inventories, please see Schiller et al. (1996).

Table 3-3. *Example of an experimental quadruple consisting of two high- and two low-frequency syllables, onsets and offsets are frequency controlled.*

Syllables within one quadruple	
high-frequency	*low-frequency*
kem [kɛm]	kes [kɛs]
wes [ʋɛs]	wem [ʋɛm]

One remaining criterion had be taken into account before such a quadruple was actually chosen to serve as an experimental quadruple. It was carefully controlled that none of the syllables within one quadruple were in fact existing words as it was crucial to investigate syllable frequency that was independent of word frequency. As already mentioned, two different syllable frequency counts were calculated, the number of occurrences and the summed frequency of occurrences; only instances that had in both scores comparable values were taken. Eventually, eight quadruples were taken as experimental item set. Thus, as every quadruple consisted of two high- and two low-frequency syllables each, the item set contained sixteen high-frequency and sixteen low-frequency items. The low-frequency items ranged in the count for the number of occurrence (per one million words) from a value of 0.02 to 1.19 with an average of 0.31 (SD = 0.36) and for the count of the summed frequency of occurrence (per one million words) from a value between zero and 3.67 with an average of 1.28 (SD = 1.28). For the high-frequency-items the values in both counts were as follows: For the count number of occurrence (per one million words), the high-frequency items ranged from 1.48 to 23.50 with an average of 5.32 (SD = 5.23). For the count of summed frequency (per one million words) of occurrence, high-frequency items ranged from 34.12 to 1,192.57 with an average of 242.32 (SD = 287.22). For an overview of all experimental quadruples and their values in both frequency counts see the Appendix 3-B. We decided not to present an experimental quadruple as one experimental item set. Rather, the all experimental sets were composed of syllables from different quadruples, this to avoid phonological overlap between the target items in one experimental set.

Procedure and Apparatus. Procedure and Apparatus were the same as in the Pre-Test.

Design. The two-level variable *Frequency* (high-frequency vs. low-frequency) was tested within participants and within quadruples[11]. Each participant produced each of the 32 items 8 times, resulting in a total of 256 experimental trials. The 32 syllables were regrouped to form in total 8 experimental sets of 4 syllables with the same frequency, resulting in four high- and four low-frequency sets. It was taken care of that none of these experimental sets contained more than one syllable stemming from the same quadruple (to avoid phonological overlap between syllables in one set). See the Appendix 3-C for a distribution of syllables to experimental sets. High- and low-frequency sets always consisted of high- and low-frequency item counterparts, see Table 3-4 for an example. This method ensured that the high- and low-frequency items were presented in a maximally comparable environment.

Table 3-4. *Example of a corresponding high- and a low-frequency experimental set.*

Experimental Set	
high-frequency	*low-frequency*
ket	keg
wes	wem
luk	lur
tur	tug

The succession of high- and low-frequency sets alternated. The direct succession of specific sets (sets 1 and 2 as well as sets 3 and 4, see the Appendix 3-C for a detailed overview) should be avoided as these sets contained members of the same quadruples that comprised segmental overlap. This direct repetition of these (overlapping) forms could have led to undesired facilitation effects. The restriction of direct succession of identical

---

[11] Since the selection of one item determined the selection of three other items resulting in one quadruple of four interrelated items, the quadruple and not the single item was treated as random factor in the analysis.

onsets taken into account, 16 different experimental versions were constructed. Every item sets occurred at every position across experimental versions. The assignment of the four different positions to items within sets was counterbalanced between participants. There was a restriction of the immediate repetition of the same item within sets. Every experiment was preceded by a practice trial that mimicked the experimental trials.

RESULTS

Observations leading to wrong or invalid responses (mispronunciations, voice key errors, and hesitations) were not included in the reaction time analysis. Time outs (RTs > 1500 ms) and. reaction times shorter than 300 ms were also removed. Observations deviating from a participant's and an item's mean by more than two standard deviations were considered as outliers and also discarded from the reaction time analysis. 218 (5.3 %) trials were treated as errors and 67 (1.6%) as outliers.

The mean of the high-frequency items and the mean of the low-frequency items were submitted to *t*-tests. Two complementary analyses were computed, one treating participants (t_1) and one treating items (t_2) as random factor (Clark, 1973).

The mean voice onset latencies, standard deviations and error rates for Experiment 3.1 are summarized in Table 3-5.

Table 3-5. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) in Experiment 3.1.*

Frequency	M	(SD)	% Err	(SD)
High	703	(65)	4.9	(3.3)
Low	704	(83)	5.8	(3.8)
Difference scores	-1		.-0.9	

The analysis of reaction times did not show any effect (both *t*s < 1). The analysis of errors showed no significant differences. ($t_1 < 1$; $t_2(7) = -1.186$, $p = .274$).

DISCUSSION

The outcome of Experiment 3.1 investigating whether high-frequency syllables are produced faster than low-frequency syllables shows no effect at all. This finding contradicts our prediction. Interestingly, however, a post-hoc test analyzing the practice trials separately, i.e., the very first production of the syllables, yielded a descriptive difference of 32 ms: Mean reaction times for high-frequency sets amounted to 720 ms whereas the average response time for low-frequency sets was 752 ms. Participants have an unequal number of learning phases and depending on that also an unequal number of practice trials as they had individual learning behaviors and performance in practice phases. Therefore, no further statistical analysis of the practice trials was done.

Note that participants learn the assignment of (loudspeaker) position and items within sets in a totally self-paced manner. There was no limitation on how often participants could click on specific positions, furthermore, participants often rehearsed items overtly while clicking on the corresponding position to memorize and exercise the assembly of position and item. If participants failed the following practice phase or felt insecure and wished to improve their performance by a rehearsal of the learning phase, the learning phase and practice phase were started again, and participants again articulated the target items before the actual test phase started. By the time the actual test phase started, both the high-frequency and the low-frequency syllables might have resulted in articulatory training effects wiping out any smaller frequency effect.

Moreover, the amount of repeated learning and practice trials and the rather long reaction times of 700 ms for the very short and highly exercised syllables might indicate that the presentation of four positions and accordingly four to-be-associated items overextended participants' capacity. However, the fact that there is an effect in the practice phases raises the possibility to detect syllable frequency effects with an improved and better controlled experimental technique[12]. The results of Experiment 3.1 should not be taken as evidence for or against the notion of a mental syllabary. This issue needs more appropriate testing.

---

[12] Notice that participants needed considerably less learning and practice phases and trials for the existing words used in the pre-test than in the current experiment.

EXPERIMENT 3.2: MONOSYLLABIC PSEUDO-WORDS

This experiment is a replication of the previous experiment with the exact same materials by using a modified experimental setup. The number of targets was reduced from four to two target items in response sets to minimize the learning load. A modification of the practice phase ensured that there is no overt articulation before targets have to be produced in the test phase, thereby guaranteeing that the first overt production will be measured as a valid response in the test phase. Since the immediate repetition of items could not be prevented in the present two-item design, we introduced fillers in the form of different numbers presented in the center of the computer screen that have to be named in between two trials. This number naming should distract participants' attentiveness to any expectation of trial succession and beyond that it should help to 'neutralize' the articulators as the immediate repetition of two identical items could have huge facilitation effects.

METHOD

Participants. Sixteen native speakers of Dutch participated in the experiment. They were randomly taken from the pool of participants of the Max Planck Institute in Nijmegen, The Netherlands and were paid for their participation. They had no known hearing deficit, and they had normal or corrected-to-normal vision.

Procedure and Apparatus. As in the previous version, the experiment consisted of alternating learning, practice and test phases. Participants were tested individually in a quiet room. They were given a detailed written instruction specifying that they had to respond as accurately and as quickly as possible.

In the learning phases, the participant's task was to associate an auditorily presented target-word to one of *two* positions on the computer screen. An icon of a little white loudspeaker (4 by 4 cm) was presented on one of two possible positions on a black computer screen (iiyamaLM704UT) while the to-be-learned target word was simultaneously presented auditorily via headphones (SennheiserHD250). The two positions of the loudspeaker icons were on the left or the right side on the screen. Participants were instructed to listen carefully to the spoken pseudo-words and to memorize the position on which the current two target words were presented. Each target word was presented two times on its specific position.

In practice phases, the loudspeaker symbols were simultaneously presented on both sides, namely the left and right position on the screen. Both icons were presented as interactive fields. While displaying the two loudspeakers, one of the target items was presented via headphones and participants were instructed to associate this target item to its position by clicking with the left mouse button (cordless Wheel Mouse, Logitech) on the correct loudspeaker on one of the two positions. The practice phase contained eight trials in which the target items had to be correctly associated to their corresponding position; each target was auditorily presented four times. During practice phases, erroneous trials were counted and were displayed by a little white number on a black computer screen succeeding the practice phase. Only if participants could pass the practice phase with zero errors, the experimenter started the test phase. When they failed, the learning phase was started again and the session was rehearsed to ensure that they learn the sets accurately. Participants were explicitly asked to refrain from rehearsing the target items by articulating them.

In test phases, one position (left or right) was presented on the screen and had to be named as accurately and fast as possible. At the beginning of each test phase and between two subsequent target trials, i.e., responding to the left or the right position on the screen, one out of four numbers (1, 3, 5 or 8) was presented in the middle of the computer screen, white on black, which had to be named. Simultaneously with the presentation of a loudspeaker symbol on the left or the right side of the screen the voice key was activated for 1500 ms. The symbol disappeared after the response with a delay of 500 ms. One of the numbers appeared after 100 ms. Loudspeaker symbols triggering target items within one set were repeated eight times in a random order, resulting in response blocks of 16 items plus the additional 4 number items, which were repeated 4 times each. Thus, one response block involved 32 items in total. The experiment started with a short practice session of two sets, practice items mimicked the experimental materials, in which participants could get acquainted with different phases and the experimental setting.

The presentation of the stimuli and the measuring of the reaction times were controlled by the NESU software package. The spoken reactions were registered by a Sennheiser MD211N microphone, which fed into a NESU-box voice key device and a

DAT recorder (Sony DTC-55ES). The experimenter sat in the same room and took note of hesitations, voice key errors, wrong naming responses, and time outs. On average, an experimental session lasted 30 minutes.

Materials. The same basic syllabic materials as in Experiment 3.1 were used. As already mentioned, 4 numbers were taken to serve as ‘fillers’: At the beginning of each set and between two subsequent targets a number which appeared in the middle of the screen had to be named. The following numbers were used: 1 (*een*), 3 (*drie*), 5 (*vijf*), and 8 (*acht*). These numbers were all monosyllabic and there was no phonological overlap between the numbers and the experimental test items.

Acoustic versions of the syllables were spoken by a female native speaker of Dutch. The spoken syllables were digitized at a sampling rate of 22 kHz, to be used during the learning phase of the experiment. They varied in duration from 382 ms to 586 ms with an average of 484.69 ms (SD = 48.23). There was no difference in duration between high-frequency and low-frequency syllables (both *t*s < 1).

Design. The two-level variable *Frequency* (high-frequency vs. low-frequency) was tested within participants and within quadruples. Each participant produced each of the 32 syllables 8 times, resulting in a total of 256 experimental trials. One experimental set consisted of two syllables that were presented as a pair. These pairs were constructed as follows: Only items of the same frequency, i.e., either high- or low-frequency items, were combined to build one item pair. These frequency-homogeneous pairs were constructed on the basis that they were as distinct as possible from each other, that is, it was taken care of that the two syllables within each set had no segmental overlap. That implied that no two items were taken from the same quadruple. The pairing of two syllables resulted in 16 sets, that is, 8 high- and 8 low-frequency sets. The pairing of two items into one frequency-homogeneous set was the same for high- and low-frequency pairs, that is, once a pair was built, take for example the high-frequency pair *lug* [lʏx]- *bin* [bɪn], this pairing determined also the pairing of their counterparts, which is in this case the low-frequency pair *lur* [lʏr] – *bing* [bɪŋ]. A full list of experimental item pairs is given in the Appendix 3-D.

For the order of the 16 experimental sets two constraints had to be respected: 1) High- and low-frequency counterparts (e.g., *lug* [lʏx] – *bin* [bɪn]; *lur* [lʏr] – *bing* [bɪŋ]) should be presented with a maximum distance of trials between them. 2) Items or sets which had an overlapping onset with other items or sets should not be presented as trials in direct succession (e.g., the item set *lug* [lʏx]– *bin* [bɪn]) should not be presented in direct succession to the other item set *luk* [lʏk]– *teg* [tɛx]). High- and low-frequency item sets alternated across the 16 item sets. Taking these constraints into account, two separate Latin Squares were designed that contained only those 4 item sets which did not overlap with each other. These two Latin Squares were then combined in such a way that the transitions between them were also controlled for initial overlap. The order of the first 8 trials was reflected by the remaining 8 item sets. Using this Latin Square procedure, sixteen experimental versions resulted: Every item set occurred at each position across all experimental versions. The position of the production cue (left vs. right) was also counterbalanced across participants.

RESULTS

Test items leading to wrong or invalid responses (mispronunciations, voice key errors, and hesitations) were not included in the reaction time analysis. They were coded as errors. Reaction times above 800 ms and below 200 ms were considered as invalid and did not enter the reaction time analysis. Observations deviating from a participant's and an item's mean by more than two standard deviations were considered as outliers and also discarded from the reaction time analysis. 161 (3.9%) trials were treated as errors and 67 (1.6%) as outliers.

The mean of the high-frequency items and the mean of the low-frequency items were submitted to *t*-tests. Two complementary analyses were computed, one treating participants (t_1) and one treating quadruples (t_2) as random factor (Clark, 1973).

The mean voice onset latencies, standard deviations and error rates for Experiment 3.2 are summarized in Table 3-6.

Table 3-6. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) in Experiment 3.2.*

Frequency	M	(SD)	% Err	(SD)
high	436	(41)	3.2	(2.2)
low	445	(48)	4.6	(3.5)
Difference scores	-9		-1.4	

The analysis of reaction times showed that high-frequency monosyllabic pseudo-words were produced significantly faster than low-frequency monosyllabic pseudo-words ($t_1(15) = 2.651$, $p < .05$; $t_2(7) = 2.904$, $p < .05$). The analysis of errors revealed a tendency towards more errors in the low-frequency condition than in the high-frequency condition ($t_1(15) = 1.689$, $p = .112$; $t_2(7) = 2.418$, $p < .05$).

DISCUSSION

The results of Experiment 3.2 show a significant frequency effect for monosyllabic Dutch pseudo-words. High-frequency syllables are faster produced than low-frequency syllables. This strongly supports the notion of the mental syllabary. As potential confounds were carefully controlled for, the present effect might be interpreted as a small but significant syllable frequency effect. The fact that the present effect is not replicating the size of the syllable frequency effect for the practice phases of Experiment 3.1 which amounted to 32 ms is not surprising considering that this result was due to an incomparable number of learning and practice trials by different subjects.

After having demonstrated a syllable frequency effect on mono-syllables, we now turn to disyllabic targets. This can provide a follow-up to another finding of Levelt and Wheeldon's (1994) results. By using disyllabic Dutch words to investigate effects of syllable frequency they factorially manipulated first versus second syllable frequency. The overall word-form frequency was controlled for. An interesting outcome of their experiments was that there was only a frequency effect for the second syllable but not for the first one. They argued that this result reflects an underlying principle namely that "[…] the speaker cannot or will not begin to articulate the word before its phonetic encoding is complete." (Levelt & Wheeldon, 1994, p. 254).

This conclusion depends crucially on the assumption that the retrieval of a phonetic syllable from the mental syllabary is independent of the processing status of the preceding syllable. For example, as soon as the second syllable of a disyllabic word has been constructed at the level of phonological encoding, the retrieval of its gestural score will be initiated, even if the retrieval of the gestural score for the first syllable is not yet complete. As a consequence there can be some temporal overlap between the retrieval of the second and the first syllable. But even if the first syllable's gestural score has been retrieved before phonetic encoding of the second syllable has been completed, articulation will wait until both syllabic programs of the target word have entered the output buffer. Any speed benefit gained from a high-frequency first syllable will thus not become apparent.[13]

In order to replicate the frequency effect obtained with monosyllabic pseudo-words and to test the prediction that (only) the second syllable in a disyllabic word will exhibit frequency effects we investigated disyllabic (Dutch) pseudo-words having the frequency-manipulation on their second syllables.

EXPERIMENT 3.3: DISYLLABIC PSEUDO-WORDS
FREQUENCY-MANIPULATION ON THE SECOND SYLLABLE

In Experiment 3.3, disyllabic Dutch pseudo-words were used to investigate effects of syllable frequency. High- and low-frequency syllables were embedded in pseudo-words, the second syllable in these pseudo-words was frequency-manipulated. All of the used pseudo-words were phonotactically possible strings of Dutch. Under the assumption that articulation waits for the retrieval of all gestural scores of the phonological word, the syllable frequency effect should be replicated.

---

[13] Notice that this assumption presupposes that the processing duration of the second syllable is independent of the processing of the preceding syllable(s). Moreover – according to Levelt and Wheeldon – the first syllable could only have an effect when the retrieval of the first syllable is completed only after the second syllable is retrieved, which is very unlikely. Levelt and Wheeldon state that it takes approximately 200 ms to articulate a syllable; whereas Wheeldon and Levelt (1995) showed that the generation of *phonological* syllables only needs half of that time, some 100 ms. Thus, they argue – as their reported syllable frequency effect amounts to only 15 ms – "[…] it seems implausible that phonetic encoding of the second syllable can 'overtake' encoding of the first one due the advantageous frequency conditions." (Levelt & Wheeldon, 1994, p. 254).

METHOD

Participants. Thirty-two participants were tested.

Procedure and Apparatus. The procedure and apparatus were the same as in Experiment 3.2.

Materials. All 32 items were disyllabic pseudo-words obeying Dutch phonotactics. In this experiment, the same syllables as in Experiment 3.2 were used which served as second syllables. In total, there were again 16 high-frequency and 16 low-frequency disyllabic items. As first syllables, eight high-frequency CV-syllables were selected. For the count *number of occurrences*, the CV-syllables ranged in frequency from 54.43 to 137.74 with an average of 97.34 (SD =33.01). For the count *summed frequency of occurrence*, the CV-syllables ranged in frequency from 1,101.76 to 3,475.79 with an average of 2,900.53 (SD = 924.34). One high-frequency syllable was assigned to one quadruple to serve as first syllable for all four members of this quadruple in order to keep the comparability within the quadruples (e.g., li.kem [li.kɛm], high-frequency; li.kes [li.kɛs], low-frequency; li.wes [li.ʋɛs], high-frequency; li.wem [li.ʋɛm], low-frequency). It was taken care of that none of the resulting disyllabic pseudo-words would form any existing word form in Dutch. A full list of items used in this experiment can be found in the Appendix 3-E. The assignment of items to experimental pairs or sets remained the same, i.e., the pairing of two items taken from different quadruples, which were presented as one experimental set was the same as in Experiment 3.2. A list showing these pairs is also given in the Appendix 3-F. The same numbers as in Experiment 3.2 were used which were presented in between two subsequent test items.

Acoustic versions of the disyllabic pseudo-words were spoken by a female native speaker of Dutch. The spoken pseudo-words were digitized at a sampling rate of 22 kHz, to be used during the learning phase of the experiment. They varied in duration from 567 ms to 799 ms with an average of 644.78 ms (SD = 57.42). There was no difference in duration between high-frequency and low-frequency items (both *t*s < 1). The disyllabic pseudo-words were spoken with stress on the second syllable. It has been suggested that articulatory routines for stressed and unstressed syllables are independently represented in the repository (Crompton, 1981; Levelt, 1989). Thus, in order to keep the basic syllable material between Experiments 3.2 and 3.3 as consistent as possible, we opted for the non-default stress-pattern in Dutch, that is, the second syllable carries the stress.

Design. Thirty-two different experimental versions were constructed applying the same principles as in Experiment 3.2.

RESULTS

The raw data were treated in the same way as in the previous experiments. 410 (5 %) trials were treated as errors and 156 (1.9 %) as outliers.

The mean voice onset latencies, standard deviations, and error rates of Experiment 3.3 are summarized in Table 3-7.

Table 3-7. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) in Experiment 3.3.*

Frequency	M	(SD)	% Err	(SD)
high	435	(51)	5.2	(3.5)
low	435	(51)	4.8	(4.0)
Difference scores	0		-0.4	

The frequency of the second syllable in a disyllabic pseudo-word did not affect naming latencies in this experiment as demonstrated by *t*-tests (both *t*s < 1). The comparison of error rates did not show a significant difference either ($t_1 < 1$; $t_2(7) = 1.049$, $p = .329$).

DISCUSSION

The results of Experiment 3.3 did not show any second syllable frequency-effect for disyllabic Dutch pseudo-words. However, before concluding that disyllabic pseudo-words containing high- and low-frequency syllables are not sensitive to frequency-effects, let us carefully consider what could have possibly caused the present null-effect.

By comparing Levelt and Wheeldon's (1994) results to the present experiments, we can distinguish a few differences that might be crucial in explaining the (divergent) data pattern. The fact that their experiment used existing disyllabic Dutch words that had to be learned in the context of four other targets, whereas participants in the present experiments had to memorize only two disyllabic pseudo-words could be responsible for the different results. The theoretical assumption that led to the hypothesis that only the second syllable should exhibit frequency effects was that speakers do not initiate articulation before all phonetic syllables have been retrieved. Could it be the case that this assumption is more plausible for existing words where linguistic integrity plays a role than for pseudo-words where integrity does not play such a prominent role? Do speakers set another response criterion? Under the present high-speed conditions participants may have started articulation as soon as the gestural score for the initial syllable was available for execution.

Results that point to the assumption that not all of the phonological word's components have to be phonetically encoded before articulation is initiated stem from a study by Schriefers and Teruel (1999). They investigated the production of German adjective noun phrases while hearing priming syllables that were identical to one of the syllables in the target phrase (as in "lila Säge" [purple saw]). Overall, they obtained priming effects of the first syllable of the first word, weak priming effects for the second syllable of that word but no priming effects for the second word. However, based on the number of restarts and hesitations, participants were divided into two groups for post-hoc analyses. Participants in the "careful" group (mean of 3 restarts and hesitation errors) showed priming for both, the first and the second syllable of the first word, whereas participants in the "hasty" group (mean of 8 restarts and hesitation errors) only showed priming effects for the first syllable. Schriefers and Teruel argued that this reflects speakers' ability to adjust the size of the planning unit and that planning of the first syllable might suffice to initiate articulation.[14]

The assumption that speakers in general "wait" until all components of a polysyllabic phonological word is phonetically encoded as proposed by Levelt and Wheeldon (1994) is

---

[14] The fact that the first syllable for all items in the present experiment consisted of only eight syllables which were thus highly repeated might also have contributed to the initiation of articulation after the retrieval of the first syllable.

apparently incorrect. It rather seems likely - as argued in Meyer et al. (2003) - that articulation cannot start before completion of *phonological word encoding*, but it can start after *phonetic encoding* of the first syllable of a disyllabic word. Then, if this is indeed what happens here, one should predict a first-syllable frequency effect. The next experiment was designed to determine just that.

EXPERIMENT 3.4: DISYLLABIC PSEUDO-WORDS
FREQUENCY-MANIPULATION ON THE FIRST SYLLABLE

In Experiment 3.4, the *first* syllable of the disyllabic Dutch pseudo-words was frequency-manipulated. Since there was no frequency effect for the second syllable in a disyllabic pseudo-word, the aim of this experiment is to test whether the first syllable in this pseudo-word context is sensitive to frequency effects. According to the explanation that the absence of the syllable frequency effect for the second syllable is due to the initiation of articulation upon phonetic completion of the first syllable of the disyllabic targets, we expect – contrary to the original hypothesis – a syllable frequency effect for the first syllable.

METHOD

Participants. Sixteen participants were paid for their participation. None of them took part in any of the previous experiments.

Procedure and Apparatus. The procedure and apparatus were the same as in Experiment 3.2.

Materials. Materials in this experiment consisted of disyllabic (Dutch) pseudo-words. All pseudo-words were possible strings of Dutch. Most of the syllables used for building the second syllables were the same high-frequency CV-syllables used in Experiment 3.3. In three cases, these second syllables were exchanged by other high-frequency syllables as they fitted the base material better than the original ones. In these cases the regrouping of syllable position resulted in either existing words or in words very similar to existing words. Thus, we exchanged two of the previously taken CV-syllables (‘li’ and ‘ra’) to appear as part of another quadruple and introduced one new CV-syllable (‘mo’ to replace ‘si’) (see item list in the Appendix 3-G; for the pairing of items into frequency-homogeneous sets see Appendix 3-H). For the count *number of occurrences*, the CV-syllables ranged from a frequency of 54.43 to 135.38 per one million words with an average of 88.89 (SD =29.68). For the count *summed frequency of occurrence*, the CV-syllables ranged from a frequency of 1,101.76 to 3,475.79 per one million words with an average of 2,845.28 (SD = 910.09).

Acoustic versions of the pseudo-words were spoken by a female native speaker of Dutch. The spoken targets were digitized at a sampling rate of 22 kHz, to be used during the learning phase of the experiment. They varied in duration from 491 ms to 722 ms with an average of 622.69 ms (SD = 55.86). There was no difference in duration between high-frequency and low-frequency syllables (both *t*s < 1).

Design. The design of Experiment 3.4 was the same as in Experiment 3.2.

RESULTS

The raw data were treated in the same way as in the previous experiments. There were 148 (3.6%) errors and 82 (2.0%) outliers.

The mean voice onset latencies, standard deviations, and error rates for Experiment 3.4 are summarized in Table 3-8.

Table 3-8. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) in Experiment 3.4.*

Frequency	M	(SD)	% Err	(SD)
high	417	(50)	3.9	(4.2)
low	427	(55)	3.3	(2.7)
Difference scores	-10		0.6	

Disyllabic pseudo-words were produced significantly faster when the first syllable was high-frequency than when the first syllable was low-frequency as shown by *t*-tests

($t_1(15) = 3.508$, $p < .01$; $t_2(7) = 7.320$, $p < .001$). The analysis of error rates did not show a significant difference ($t_1 < 1$; $t_2(7) = 1.008$, $p = .347$).

DISCUSSION

The results of Experiment 3.4 revealed a significant syllable frequency effect for disyllabic Dutch pseudo-words in which the first syllable is frequency-manipulated whereas the second syllable is consistently high-frequency. The magnitude of the effect (10 ms) is comparable to the size of the effect in Experiment 3.2 (9 ms). Thus, this experiment replicates the results of the experiment that investigated the monosyllabic pseudo-words. Again, the results of this experiment support the notion of a mental syllabary.

Initially, it was predicted that (only) the second syllable in a disyllabic (pseudo-) word would be frequency-sensitive, which we did not find in Experiment 3.3. We cannot attribute this difference to differences in materials between the experiments, because the relevant syllables were essentially the same. The obvious conclusion is that under the present experimental conditions subjects were able to initiate articulation of the first syllable right upon completion of its phonetic encoding. In that case frequency properties of the first, but not the second syllable will affect onset latencies.

GENERAL DISCUSSION

The aim of the present study was to test the notion of a mental syllabary. The assumption that speakers access a mental storage of precompiled articulatory gestures at the interface of phonological and phonetic encoding was described as an inherent part of the speech production model by Levelt et al. (1999). Such a mental storage implies different retrieval times for high- and low-frequency syllables. In four experiments using a symbol-position association learning task we investigated effects of syllable frequency by contrasting the production of high- and low-frequency syllables in mono- and disyllabic pseudo-words. The syllabic material used in these experiments was carefully controlled for any potential confounds by applying specific search routines. Two high- and low-frequency syllables were each assigned to one quadruple thereby guaranteeing that onsets, offsets, phoneme and bigram frequency and also the transitional probabilities between the single phonemes were controlled for within each of these quadruples.

Experiment 3.1 investigating monosyllabic pseudo-words did not yield a significant syllable frequency effect, at least not in the test phases. However, the fact that high-frequency syllables in practice phases were produced faster than low-frequency syllables motivated the replication with a modified experimental set up and a slightly changed design. The crucial differences between these two experiments were the reduced number of target items presented in one set (only two items instead of four) and that there was no overt articulation before the actual test phase (to avoid articulatory training effects). In Experiment 3.2, a significant syllable frequency effect was obtained. We therefore attributed the observed syllable frequency effect to the retrieval of stored, pre-compiled syllable gestural scores.

In all experiments reported here we exclusively used existing Dutch syllables. Speakers are, however, able to produce new syllables as well. Participants will, for instance, rather fluently read materials containing pseudo-words consisting of phonotactically legal but non-existing syllables. Apparently, we can assemble syllables on-line. This means that there are two possible operation routes in the phonetic encoding of syllables: retrieving pre-compiled syllabic programs and on-line assembly of phonetic programs. Although the present experiments provide support for the first route, retrieval of precompiled gestural programs, they are neutral with respect to the existence and ubiquity of the second, on-line assembly route. Assembly may be limited to just new syllables. It may also be the dominant operation for low-frequency syllables. But it cannot be excluded that the assembly route is *always* active, running in parallel to the retrieval route. Speech latencies are then determined by whichever operation is fastest.

Recent neuroimaging data point towards this latter hypothetical state of affairs. In an fMRI study, Mayer, Ackermann, Dogil, Erb, and Grodd (2003) investigated whether they could find neuro-cognitive differences of the two separate routes of phonetic encoding in German. Access to the mental syllabary was represented by reading and repetition of the syllables with the highest frequency, on-line assembly by the reading and repetition of syllables with the lowest frequency. Preliminary results of 14 participants showed enhanced activation of brain areas in the temporal region of the left hemisphere solely during the production of high-frequency syllables. During the production of low-frequency syllables no such activation was observed, supporting the assumption that high-frequency syllables are retrieved from a mental storage.

Disyllabic pseudo-words with frequency manipulated syllables

In Experiments 3.3 and 3.4, disyllabic pseudo-words were tested that had the frequency-manipulation on the second syllable (Exp. 3.3) and first syllable (Exp. 3.4), respectively. Both experiments were carried out in order to replicate the syllable frequency effect previously obtained for monosyllabic pseudo-words and to test the theory's prediction for syllable frequency effects in disyllabic (pseudo-)words.

Experiment 3.3 using disyllabic pseudo-words containing frequency-manipulated syllables on the second position attempted to find a frequency effect as predicted by the outcome of prior experiments (Levelt & Wheeldon, 1994) and by the WEAVER++ model (for an overview see, Roelofs, 1997). As Levelt and Wheeldon (1994) claimed, the initiation of articulation itself is dependent on the completeness of the phonetic encoding of the corresponding planning unit, namely the phonological word. In case of a disyllabic word, phonetic encoding is completed as soon as both syllables are ready for execution. The outcome of Experiment 3.3 contradicts this assumption. No syllable frequency effect was obtained for the second syllable of a disyllabic (pseudo-)word. Apparently, articulation was initiated before the frequency-manipulated second syllable was retrieved from the mental syllabary.

There are other findings in conflict with the assumption that speakers wait with speech onset until both syllables of a disyllabic word are phonetically encoded. For example, Bachoud-Lévi, Dupoux, Cohen, and Mehler (1998) failed to find different production latencies for mono- and disyllabic target words in a series of picture-naming experiments in French and English. The authors concluded that either a) word forms are not generated sequentially or b) that speakers in fact start articulation before completion of the phonological word's phonetic encoding. The assumption that phonological encoding is a sequential process is supported by several studies (Meyer 1990, 1991, Meyer & Schriefers, 1991; Roelofs, 1998), thus, it rather has to be concluded that speakers do start articulation before the whole phonological word is phonetically generated. However, by presenting results from a series of experiments investigating effects of word length, Meyer et al. (2003) offer a convincing explanation for the results of Bachoud-Lévi et al. (1998). In Meyer et al.' experiments, Dutch participants had to name object with short, monosyllabic and long, disyllabic names. Interestingly, they found effects of word length only when long and short target words were presented in separate blocks, not when they were presented in mixed blocks. They argued that speakers used different response criteria in pure and mixed blocks: Participants tried to meet different response deadlines by either generating the motor program for one syllable of monosyllabic target words or the motor programs for both syllables of disyllabic target words before speech onset. Several studies suggest that participants can adopt different time criteria when to respond contingent on task difficulty (Lupker et al., 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Taylor & Lupker, 2001). For example, Lupker et al. (1997) found that participants can adjust their performance in relation to the preceding trial. Therefore, Meyer and colleagues (2003) suggest that the absence of a word length effect reported in the study by Bachoud-Lévi et al. (1998) could be due to their stimuli presentation only in mixed blocks. In mixed blocks, participants might set an intermediate response criterion and start with the first completed syllable also in disyllabic words.

Schriefers and Teruel (1999) also argued for strategic control of response initiation by speakers. They claim that the actual response criterion is apparently influenced by various factors and that the amount of what has to be (phonetically) completed before speech onset is variable.

Taken the results from the latter studies together, speakers seem to be able to adjust their speech coordination to the difficulty and speed of the current experimental task, and furthermore different speakers generally seem to have a preference for faster or slower responses or for smaller or larger planning units. The actual response criterion is apparently influenced by various factors and the amount of what has to be phonetically completed before speech onset seems variable.

We interpreted the results of Experiment 3.3 in that light and concluded that the absence of the second syllable's frequency effect was due to the early initiation of articulation, namely after completion of phonetically preparing the first syllable. This account was confirmed by the results observed in Experiment 3.4. Here, the frequency of the first syllable of a disyllabic pseudo-word was manipulated. We found a significant syllable frequency effect (which was of the same size as the effect observed in Experiment

3.2 and thereby replicated the results with the monosyllabic pseudo-words). Thus, taken the results of Experiments 3.3 and 3.4 together, the assumption that speakers can start articulation as soon as the first syllable is fully encoded is supported.

In line with the assumption that speakers can start articulation upon phonetic completion of the first syllable in a disyllabic pseudo-word are recent results from Spanish and French. Carreiras and Perea (in press) investigated disyllabic pseudo-word naming in Spanish, manipulating first and second syllable frequency. They obtained significant syllable frequency effects for the first but not for the second syllable. This result also suggests that speakers start their articulation before having retrieved or computed all of the word's syllables. That this result is not restricted to the experimental use of pseudo-words[15] is not only shown by the study by Schriefers and Teruel (1999) but also by Brand et al. (2002). In a larger-scale study these latter authors collected the naming latencies of 600 French frequency-manipulated disyllabic words from 100 participants. A facilitative effect for the frequency of the first syllable but not for the second syllable was obtained.

All in all, the moment speakers actually initiate articulation apparently depends on various factors. It does not seem to be the case that there is a fixed planning unit here. With respect to the present theoretical framework, we would like to conclude that the present results as well as the findings reported in the literature can be explained by an advanced and completed *phonological* planning on the one hand and (more importantly) the incremental nature of the subsequent processes on the other hand. The phonological encoding procedures must operate on the (phonological) word as a whole as it needs some "look ahead" to allow for stress and intonation assignment. The further processes, that is the transformation of the single phonological syllables into phonetic representations might take place in a piecemeal fashion, syllable by syllable. This incremental, left-to-right procedure in which successive syllables are generated leads to some temporal overlap of processes: The first syllable's gestural score has been successfully retrieved from storage, while the second and potential later syllables are still under construction. Furthermore, the fact that words with more syllables have longer speech onset latencies than words with fewer syllables when other factors, such as number of segments or word frequency are controlled for does not pose a problem for this proposal. These effects of word length can

---

[15] The experimental use of pseudo-words is unavoidable when we have to control for a great number of potentially confounding variables as in the present experiment.

be explained by assuming that the *phonological* rather than the *phonetic* encoding of a phonological word must be completed before speech onset (see Eriksen, Pollack, & Montague, 1970; Klapp, Anderson, & Berrian, 1973; Meyer et al., 2003; Santiago et al., 2000, but see also Roelofs, 2002 and Santiago et al., 2002; Wheeldon & Lahiri, 1997). Further research is needed to investigate which planning unit is used in a given speech context and what factors release the articulation initiation.

CONCLUSIONS

We found a clear and significant syllable frequency effect in monosyllabic as well as in disyllabic pseudo-words. This was interpreted as evidence for the existence of the mental syllabary; access to pre-compiled gestural scores for high-frequency syllables from storage being faster than access to low-frequency syllables. Such pre-compiling will contribute to the speed and fluency of spoken language production.

Furthermore, only in the disyllabic pseudo-words that had the frequency-manipulation on the *first* syllable we obtained a significant syllable frequency-effect but not when the *second* syllable was manipulated. We interpreted this finding as a result of speakers' flexibility to initiate articulation when the phonetic planning of the first syllable is completed for execution.

EFFECTS OF SYLLABLE PREPARATION AND SYLLABLE FREQUENCY: A COMBINED STUDY

CHAPTER 4

(An adapted version of this chapter has been submitted for publication as Cholin, J. & Levelt, W. J. M., 'Effects of syllable preparation and syllable frequency in speech production')

ABSTRACT

In the current paper, we examined whether the syllable preparation effect (Cholin, Schiller, Levelt, 2004) interacts with the syllable frequency effect (Cholin, Levelt, & Schiller, submitted). A reading variant of the implicit priming paradigm was used to contrast the voice onset latencies for frequency-manipulated disyllabic Dutch pseudowords. By simultaneously manipulating the items' form overlap and initial syllable frequency, we tested whether or not the preparation effect involves the retrieval of stored syllable units from a mental syllabary. As accessing and retrieving of low-frequency syllables takes longer than retrieving high-frequency syllables it was predicted that low-frequency sets should yield a larger gain through preparation compared to the high-frequency sets. The findings reported here confirm this prediction: Low-frequency syllables benefit significantly more from the preparation than high-frequency syllables. This result suggests that the preparation effect includes access to the mental syllabary; it converges with the earlier reported findings of syllable preparation and syllable frequency. This finding adds further evidence to the claim that syllables are core functional units in speech production and extends our understanding of the nature of the preparation effect.

INTRODUCTION

The present chapter aims to investigate whether the effects of *syllable preparation* reported in Cholin, Schiller, & Levelt (2004) and *syllable frequency* reported in Cholin, Levelt, & Schiller (submitted) have distinguishable sources or whether both effects in fact have the same origin. To answer this question, the two methods, which were used to test for the separate effects of syllable preparation and syllable frequency were combined into one single task, namely an Implicit Priming Reading Task with frequency-manipulated items. In both of the former studies, evidence for the role of the syllable at the interface of phonological and phonetic encoding in speech production could be obtained. The first study, investigating whether syllables could be prepared only on the basis of advanced knowledge about the segmental structure or whether additional syllabic structure information would add to the preparation effect, supports the assumption that syllables play in fact an important role at the phonology/ phonetics interface. The second study reported a syllable frequency effect (i.e., high-frequency syllables yielded faster reaction times than low-frequency syllables), supporting the notion of a *Mental Syllabary.*

The mental syllabary is thought of to be a store of precompiled motor programs of syllabic size, which is accessed during the translation of phonological syllables into phonetic syllables. This mental store serves as mediator between the two procedures of phonological and phonetic encoding by supplying gestural scores for syllables which have all phonetic specifications to guide the next encoding step, articulation.

Thus, we have evidence for a) the independent role of syllables at the interface of phonology and phonetics and b) the assumption that these independent units are frequency-sensitive elements in a mental. This interpretation was based on the assumption that only stored units exhibit frequency effects.

The main question which shall be addressed throughout this chapter is as follows: Are the two observed effects, namely those of syllable preparation and of syllable frequency, independent of each other or does the reported preparation effect include (advanced) access to the mental syllabary? This study as well as the former studies takes the Levelt, Roelofs and Meyer (1999) model of speech production with its computer simulation WEAVER++ as referential frame.

In a first step, this model of speech production and its computer simulation WEAVER++ (Roelofs, 1997b; Levelt et al., 1999) will be introduced, the assumptions and predictions of other models will be discussed when contrary to the Levelt et al. model. The present chapter exclusively concerns the stages of phonological and phonetic encoding during the process of speech production. An overview of the empirical evidence for the syllable's role at different levels of word-form encoding is presented and the two different methods which were used to investigate the effects of *syllable preparation* and *syllable frequency* and their specific results are discussed in more detail.

THE INTERFACE OF PHONOLOGICAL AND PHONETIC ENCODING

At the interface of phonological and phonetic encoding, abstract phonological syllables are translated into phonetic representations. A crucial assumption of the Levelt et al. (1999) theory is that speakers have access to a repository of syllabic gestures. This repository, coined the 'mental syllabary' (Levelt, 1992; Levelt & Wheeldon, 1994), contains the articulatory scores for at least the high-frequency syllables of the language.

PHONOLOGICAL ENCODING

Phonological encoding starts with the retrieval of the target word's *phonological code* from the mental lexicon. This code consists of an ordered set of phonemic segments. For stress-timed languages such as English and Dutch the model also assumes the existence of sparse metrical markers in phonological codes. In these languages, the default position is defined as the first full-vowel syllable of a word. Thus, for polysyllabic words that do not have primary stress on the first stressable syllable, the metrical structure is stored as part of the phonological code; but, for monosyllabic words and for all other polysyllabic words, it is not stored but computed (see also Schiller, Fikkert, & Levelt, in press). Metrical structures describe abstract groupings of syllables (σ) into feet (Σ) and feet into phonological words (ω). A phonological word is defined as the minimal unit which is above the metrical foot, to which clitics, such as unstressed words, can attach (as in "she's" where the reduced form of 'is' is attached to 'she', see Levelt, 1989; Wheeldon & Lahiri, 1997). Crucially, at the stage of phonological encoding, phonological segments are not yet assigned to syllabic positions, nor is the C(onsonsant)-V(owel) (hereafter CV-) structure for the word specified. This is in contrast to other models of spoken word production (in particular Dell, 1986, 1988), which assume that the retrieved phonological codes are pre-syllabified. Internal syllabic positions such as onset, rhyme and coda are thus pre-specified in Dell's model. The main argument for *not* pre-determining syllable positions in the phonological codes stored in the lexicon results from the phenomenon of *resyllabification*. In connected speech, syllable boundaries often differ from a word's or morpheme's canonical syllabification. The domain of syllabification is the phonological word, which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticization (Booij, 1995). If, for instance, the stored phonological code for the word *protect* would be syllabified (i.e., as *pro-tect*, then the speaker must 'resyllabify' the word when used in a different context, such as the past tense (*pro-tec-ted*) or cliticization (*protect it* - *pro-tec-tit*). The ubiquity of such 'resyllabifications' in the normal use of English (or Dutch for that matter; see Schiller, Meyer, Baayen, & Levelt, 1996), renders pre-specification of segments to syllable positions highly inefficient.[16]

The alternative assumption, therefore, is that a word's syllabification is not retrieved but computed on-line depending on the context in which the word appears. During this process, called 'prosodification', retrieved segments are incrementally combined to form successive syllables. Also, these successive syllables are incrementally assigned the appropriate metrical properties, either following default stress, or otherwise the retrieved non-default stress marking feature. The incremental composition of syllables follows, on the one hand, universal syllabification constraints (such as maximization of onsets and sonority gradations) and, on the other hand, language-specific rules, e.g., phonotactics. Together, these rules create well- pronounceable syllables. The output of phonological encoding is a phonological word, specified for its metrical, syllabic, and segmental properties.

Dell's (1986, 1988) model makes different assumptions with respect to the phonological encoding process. As already mentioned above, his model includes abstract

---

[16] The claim that phonological codes are not pre-syllabified is, in part, a language-specific claim. For a language like Mandarin Chinese, which has a small set of syllables and limited resyllabification processes, the story might be different (see Chen, Chen, & Dell, 2002).

phonological representations that are specified for internal syllabic positions, i.e., the word form retrieved from the mental lexicon activates not only segmental information but also syllabic frames. These syllabic frames serve as placeholders into which the retrieved segments are inserted during the process of segment-to-frame-association.

ACCESS TO THE MENTAL SYLLABARY AND PHONETIC ENCODING

The fairly abstract, syllabified phonological words are incrementally translated into articulatory-motor programs. The model assumes that as soon as a syllable emerges during incremental syllabification, the corresponding syllabic articulatory gesture will be selected from the repository possibly located in Broca's area or a pre-motor area (Dronkers, 1996; Indefrey & Levelt, 2000; Kerzel & Bekkering, 2000).

The output of the mental syllabary in turn serves as input to phonetic encoding. During this latter step contextually-driven phonetic fine-tuning of retrieved motor programs occurs: The motor programs are still rather abstract representations of the articulatory gestures which have to be performed at different articulatory tiers, a glottal tier, a nasal tier and an oral tier. The gestural scores are abstract in the sense that their execution is highly context-dependent (due to allophonic variation, coarticulation and, as a result of this, assimilation). The actual details of the movements in realizing the scores, such as lip protrusion and jaw lowering, is within the domain of the articulatory system (Goldstein & Fowler, 2003). According to Levelt (1989), the stored syllable can be pronounced with more or less force, with shorter or longer duration, and different kinds of pitch movements. These are free parameters, which have to be set from case to case. For new or very low-frequency syllables it is proposed that articulatory plans are assembled using the segmental and metrical information specified in the phonological syllables. Finally, the articulatory network, a coordinative motor system that includes feedback mechanisms (Goldstein & Fowler, 2003; Saltzman, 1986; Saltzman & Kelso, 1981), transforms these articulatory plans into overt speech.

WEAVER++

The WEAVER++ model provides an account for the absence of a syllable priming effect and for the presence of a syllable frequency effect. So far, the model has been more successful in the former case than in the latter. Here, the computational rationale is discussed in more detail, because it provides the motivation for the present study. *WEAVER* (*W*ord-form *E*ncoding by *A*ctivation and *VER*ification) is the spreading activation based computer network model developed by Roelofs (1992, 1996, 1997a, 1997b, 1998, 1999), which is based on Levelt's (1989, 1992) theory of speech production. WEAVER++ adopts Dell’s (1986) assumption of word form retrieval by the spread of activation and Levelt’s (1992) on-line syllabification and access to a syllabary (Levelt & Wheeldon, 1994).

In accordance with Levelt and Wheeldon (1994), WEAVER++ (Roelofs, 1997a, 1997b) assumes that the syllabification of a word is computed on-line during the speech production process. In the WEAVER++ model, segments in the retrieved phonological code are not specified for their syllable position, but only for their serial order within a word. The actual syllabic position of a segment is determined by the syllabification process. Each retrieved segment in the phonological code spreads activation to all syllabic gestures in which it partakes. Hence, upon retrieval of a phonological code, there are always multiple phonetic syllable programs in a state of activation. How is the appropriate syllable program selected? There are, first, selection conditions. The crucial one is that the syllable matches the phonological syllable that is incrementally composed; this involves a procedure of verification. Second, each syllable in the syllabary has a frequency dependent selection threshold. This causes the predicted syllable frequency effect on naming latencies. Note, however, that the threshold assumption is a ‘modular’ one. Removing it does not affect the architecture of the system. Third, selection is subject to Luce’s (1959) choice rule. During any smallest interval, the probability of selecting the (verified) target syllable equals the ratio of its activation to the summed activation of all syllable nodes. Given this choice ratio, the expected selection latency can be computed.

SYLLABLE PRIMING STUDIES

Explicit syllable priming studies

A series of cross-linguistic studies used a syllable priming task to identify the syllable as a relevant unit in speech production (for Dutch: Baumann, 1995; Schiller, 1997, 1998; for Mandarin Chinese: Chen, Lin, & Ferrand, 2003; for French: Brand et al., 2003; Evinck, 1997; Ferrand, Segui, & Grainger, 1996; Schiller et al., 2002; for English: Ferrand, Segui, & Humphreys, 1997; Schiller, 1999, 2000; Schiller & Costa, submitted; for Spanish and an overview see Schiller et al., 2002). However, rather than demonstrating a syllable priming effect, each of these studies found a segmental overlap effect, namely that phonologically related primes, whatever their syllabic relation to the target word, facilitated the response relative to unrelated control primes, with longer CVC-primes being more effective than CV-primes. Moreover, no specific syllable priming effects were obtained. Several subsequent studies have failed to find a syllable priming effect but rather confirmed the finding of a segmental overlap effect[17].

The converging outcome of all of these syllable priming studies points to the conclusion that, at the stage where the (explicit) priming taps into the speech production process, there is only segmental but no syllabic representation available. If one considers that the primes in the syllable-priming task never get overtly articulated, one can imagine that these syllabic primes may not reach the late stage where syllables are computed on-line. Instead, the primes are assumed to speed up the segmental spell out, i.e., the moment when segments become available as part of the stored word form which is retrieved from the mental lexicon. At the stage of segmental spell-out there is only segmental, but no syllabic, information available according to the Levelt et al. theory (1999; but see Dell, 1986, 1988). It should be mentioned that Dell's model cannot account for the absence of syllable priming effects reported repeatedly in the literature (Schiller, 1997, 1998, 1999, 2000; Schiller et al., 2002). Rather, his model would predict a syllable priming effect:

---

[17] Much discussion has been given to the results of the apparent syllable priming effect in French (Ferrand et al., 1997, Experiment 5). However, Brand et al. (2003)'s failure to replicate the Ferrand effects suggests that this should not be taken as strong evidence for the syllable effect (see also Evinck, 1997; and for a review Schiller et al., 2002).

Lexical entries are already syllabified and should therefore be sensitive to the syllable match or mismatch between prime and target words, respectively.

Implicit syllable priming studies and effects of syllable preparation

A paradigm which is known to be sensitive also to later stages of word form encoding is the “*implicit* priming paradigm”: Whereas the explicit priming is sensitive only to early stages of phonological encoding, the implicit priming paradigm exhibits effects that emerge at these early stages but also comprise later stages at the interface of phonological and phonetic encoding, i.e., on-line syllabification, possibly including access to the mental syllabary (Meyer, 1990, 1991; Roelofs, 1996, 1998; Roelofs & Meyer, 1998).

In the implicit priming paradigm, participants learn a small set of prompt-response pairs. The response words are either phonologically related or not: In so-called *homogeneous sets*, they share the first two or three segments, the shared string between items in the response set is referred to as *implicit prime*. In *heterogeneous sets*, there is either no such shared phonological property among the items in the set. The heterogeneous sets are created by regrouping the prompt-response pairs from the homogeneous sets, e.g., *loner*, *beacon*, *major* (etc.). Production latency (the time between onset of prompt and speech onset, measured by voice key) is the dependent variable. The standard effect in the implicit priming paradigm is faster production latencies for homogeneous than for heterogeneous blocks. Meyer (1990, 1991) reported such an effect only when the response words in the homogeneous sets shared one or more word-initial segments. No effect was found for shared word-final segments demonstrating the incrementality of the process of syllabification. The preparation effect was found to increase with the length of the shared initial stretch[18].

---

[18] For one experiment, Meyer (1991) reports that sets with open initial syllables (CV) that share only those two initial segments produced preparation effects that were equivalent to effects produced for sets with closed syllables (CVC) that shared three initial segments. This result was surprising because a pure segmental length effect would predict larger greater preparation effects in the CVC sets since they comprise one more shared segment. This finding supports the possibility of syllabic effects that are independent of segmental length. However, contrary to this result, Roelofs (1996, Experiment 6) showed that the size of the preparation effect depends on the length of the shared syllable in terms of number of segments.

In a recent study, Cholin, Schiller and Levelt (2004) used an *odd man out* variant (Janssen, Roelofs, & Levelt, 2002) of the implicit priming technique to demonstrate that speakers can benefit from a shared syllable structure. In this study, participants learned sets of prompt-response pairs. All the words from a given response set shared their initial segmental information. Response sets had - in a addition to the overlap of initial segments - either a constant or a variable syllable structure of the first syllable. (See Table 4-1 for an example of a constant and a variable set).

Table 4-1. *Example for a constant and variable item set.*

Constant set	Variable set
spui.en ([to] drain)	*spoe*.len ([to] rinse)
spui.de (drained)	*spoel*.de (rinsed)
spui.er (person who drains)	*spoe*.ler (person who rinses)
spui.end (draining)	*spoe*.lend (rinsing)

Under the assumption that the syllable is a relevant processing unit in speech production, it was predicted that speakers can phonetically prepare their utterance only when they know what the form of the initial syllable will be (as they do in constant sets) but not when the initial syllable varies within the set of words (as they do in the variable sets). Two different CV-structures (CVV, Exp.2.1; CCVV, Exp.2.2) were investigated. A significantly larger preparation effect for constant sets compared to variable sets was found. The constant sets were on average 64 ms faster produced than the variable sets. Control studies showed that variable sets also yielded a preparation effect (compared to a baseline homogeneous condition where no phonological property was shared between members in a response set). Thus, in variable sets, responses can also be prepared but only up to on-line syllabification. In these sets, all stages *preceding* on-line syllabification contribute to the preparation effect. In constant sets, where the implicit prime includes initial segments and also the syllabic structure of the first syllable, the first syllable can be fully prepared for articulation. All stages prior to articulation, including segmental spell-out, on-line syllabification and possibly access to the mental syllabary, contribute to the preparation effect.

In the odd-man-out variant of the implicit priming task, syllabic information is relevant in the sense that any deviating syllabic structure, i.e., an item with a different syllable structure, reduces the preparation effect. The authors argued that this preparation effect, which cannot be attributed to a segmental overlap effects as the segmental overlap is identical, offers strong evidence for the relevance of syllables in word form encoding. These results testify to the relevance of syllabic units at the interface of phonological and phonetic encoding, however, they do not represent indisputable evidence for the notion of separately stored syllabic units. The implicit priming paradigm taps into later stages of word-form encoding than explicit priming paradigms do, but the preparation effect is likely to arise at any one of these stages or, more likely, at a combination of these stages. It cannot be concluded with certainty that access to the mental syllabary is involved. Effects that represent strong evidence for the existence of the mental syllabary are effects of syllable frequency because only stored units are thought to exhibit frequency effects.

EFFECTS OF SYLLABLE FREQUENCY

In a recent study, Cholin, Levelt, & Schiller (submitted) investigated effects of syllable frequency. If the mental syllabary consists of retrievable representations corresponding to syllables, than the retrieval process should be sensitive to frequency differences. A Symbol - Position Association Learning task was used to contrast the production of high- and low-frequency syllables in mono- and disyllabic pseudo-words. The basic syllable material was carefully controlled for any potential confounds such as onsets, offsets, phoneme and bigram-frequency and also the transitional probabilities between the single phonemes of the syllables. (The materials used in that study are partly reused in the present study; the detailed description of the construction of the specific syllabic material is therefore given in the method section of this chapter.) The participants' task was to respond as fast as possible with a previously learned associated target-word when a production cue was presented on the screen. In learning phases, participants were presented a symbol on one of two potential positions (the left or right position on a computer screen) and were simultaneously presented the to-be-associated word via headphones. In the test phase the same icon was shown on either the right or the left side of the computer screen to prompt the previously associated target. The auditory presentation of the target ensured that potential confounds deriving from orthographic factors could be excluded. The results can be summarized as follows: A small but highly significant syllable frequency effect was obtained in testing monosyllabic pseudo-words. This effect was replicated investigating disyllabic pseudo-words bearing the frequency-manipulation on the first syllable. No effect was obtained for disyllabic pseudo-words bearing the frequency-manipulation on the second syllable, suggesting that speakers can start articulation as soon as the first syllable is retrieved.

Evidence for the assumption that syllables are retrieved as whole entities also comes from recent studies in Spanish and French. Carreiras & Perea (in press) obtained significant syllable frequency effects for speech production in Spanish (see also Perea & Carreiras, 1996, 1998). Brand, Rey, Peereman, & Spieler, (2002) also found significant syllable frequency effects in a larger-scale study. Laganaro (2003) investigated effects of syllable frequency in French from a psycho- and neurolinguistic perspective. In her psycholinguistic studies she found that syllable frequency affected naming latencies in a picture naming experiment (see also Alario, Ferrand, Laganaro, New, Frauenfelder & Segui, in press) and the reaction times in non-word reading. In her neurolinguistic study she reported a significant frequency effect on segmental substitution errors and a syllable frequency effect on non-word repetition performance in French Aphasics. Similarly, Aichert and Ziegler (2004) found a syllable frequency effect in the error performance of patients with apraxia of speech. These studies are supportive of the notion of a mental syllabary.

To summarize, there is evidence that syllables play a functional role at the interface of phonological and phonetic encoding during speech production: The absence of syllable priming effects in the explicit priming studies on the one hand and the positive syllable-preparation effects in the implicit priming studies on the other hand suggest that (phonologically abstract) syllables are not stored in the mental lexicon (as argued by Dell, 1986, 1988) but rather generated by an on-line syllabification process[19]. The syllable

---

[19] Please note that the idea of a mental syllabary is also comparable with Dell's syllabified phonological code, the existence of stored units does not deny a store for pre-compiled syllabic programs.

frequency effects confirm the assumption that speakers have access to a mental syllabary located at this interface. In explaining the preparation effects for the constant versus the variable sets in the odd-man-out version of the implicit priming task, the authors (Cholin, Schiller, & Levelt, 2004) proposed that preparation for the *constant* sets involves access to the mental syllabary as all relevant syllabic information is given in advance. It was argued that the preparation effect in constant sets is eventually determined by the advanced retrieval of stored syllabic gestures. (In variable sets, this was not possible due to the lack of information about the relevant syllabic structure.) Even though this seems to be the most likely account for the current data pattern, it could also be assumed that the preparation effect arises at an earlier stage not including access to the mental syllabary. The purpose of the present study is to test the assumption that the preparation effect involves (advanced) retrieval of stored syllabic representations.

EXPERIMENT 4.1: THE IMPLICIT PRIMING TASK AS A READING STUDY WITH FREQUENCY-MANIPULATED DISYLLABIC PSEUDO-WORDS

This specific reading variant of the implicit priming paradigm was first introduced by Roelofs (in press). In three experiments, he investigated whether object naming and oral reading share the same phonological encoding procedures and moreover, whether these processes operate serially or in parallel. The results suggest that the underlying mechanisms of these processes - due to the very similar results in the corresponding tasks - are in fact shared at the level of segmental retrieval. It could been shown that the effects are comparable to the serial effects observed with the classic implicit priming paradigm (Meyer, 1990, 1991).

On the basis of these results we decided to use this specific variant of the implicit priming task for our purpose: We want to investigate whether the preparation effects which were found in preliminary research (Cholin, Schiller, & Levelt, 2004) include advanced access to the mental syllabary, i.e., whether the retrieval of stored syllable scores contributes to the preparation effects.

In the present study, we will combine the (previously tested) factors of *preparation* and *frequency*. We presented disyllabic (Dutch) pseudo-words that contained frequency-manipulated first syllables; the second syllables were always high frequent. These pseudo-words were arranged in sets of four items each. These sets were either homogeneous or heterogeneous, and consisted either of high- or of low-frequency items. More precisely, a homogeneous set consisted of disyllabic pseudo-words that had an identical first syllable as in *westa*, *wesli*, *weswa*, and *wesjo*. The first syllable *wes* is high-frequent. The homogeneous low-frequency counterpart to this set consisted of *wemta*, *wemli*, *wemwa*, and *wemjo* (with *wem* being low-frequent). The corresponding heterogeneous (high- or low-frequency) sets, consisted of four pseudo-words with different first syllables, that were created by regrouping the items used in the homogeneous sets.

Before each test-phase participants were familiarized with the respective set by (silently) looking at the items. For homogeneous sets participants could therefore prepare the first syllable of the upcoming target word before it was displayed on the screen. For heterogeneous sets, no such preparation was possible as the four target words contained four different first syllables. Due to the preparation in homogeneous blocks, faster production times are predicted in comparison to heterogeneous blocks (i.e., the *preparation effect*; Meyer, 1990, 1991; Cholin, Schiller, & Levelt, 2004).

The size of this preparation effect should, however, vary with respect to the frequency of the to be prepared syllable. In a previous study it was shown that low-frequency syllables take more time to be produced than high-frequency ones (Cholin, Levelt, & Schiller, submitted). If, however, a low-frequency syllable can be prepared for in advance and access to the mental syllabary, is actually part of the preparation effect, then the resulting gain for the low-frequency syllables should be larger than for high-frequency syllables. In fact, if the preparation effect actually includes access of the mental syllabary no frequency effect should remain for homogenous blocks. Thus, the crucial prediction is an interaction between *preparation* and *frequency*. If, however, the preparation effect does not include the retrieval of syllables from the mental syllabary no interaction should be observed, but rather comparable frequency effect for heterogeneous and homogeneous sets (see Chapter 3).

METHOD

Participants. Thirty-two native speakers of Dutch participated in the Experiment. They were randomly taken from the pool of participants of the Max Planck Institute in

Nijmegen, The Netherlands and were paid for their participation. They had no known hearing deficit, and they had normal or corrected-to-normal vision.

Material. 8 high- and 8 low-frequency syllables served as base for constructing the experimental materials. Two high- and two low-frequency syllables each formed a *quadruple*. High- and low-frequency syllables within one quadruple are controlled for phoneme length, phoneme frequency, CV structure and bigram frequency by a specific way of pairing the high- and low-frequency counterparts into one quadruple: The corresponding onsets and offsets are represented each in the high- and in the low-frequency context. Four such quadruples were chosen. A full list of experimental quadruples and their frequency counts is given in the Appendix 4-A[20]. See Table 4-2 for a schematic depiction of such a quadruple.

---

[20] All frequencies counts were obtained from the computer database CELEX (CEntre for LEXical Information), which has a Dutch lexicon based on 42 million word tokens. Syllable frequency was 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. The syllable frequency ranges from 0 to approximately 90,000 per million words, with a mean frequency of 121. Two syllable frequency counts were calculated: The number of occurrences of each syllable (independent of the frequency of occurrence of the syllable in a particular word position, i.e., first or second syllable position within a word) and the number of the summed frequency of occurrence of each syllable (within words). Only instances that had in both scores comparable values were taken. The 16 *high-frequency* items ranged in the count for the *number of occurrence* (per one million words) from a value of 1.48 to 23.50 with an average of 6.72 (SD = 7.15) and for the count of the *summed frequency of occurrence* (per one million words) from a value from 62 to 1192.57 with an average of 311.18 (SD = 369.99). For the low-frequency-items the values in both counts were a follows: For the count number of occurrence (per one million words), low-frequency-items ranged from 0.02 to 1.19 with an average of 0.25 (SD = 0.39). For the count of *summed frequency of occurrence* (per million words), low-frequency items ranged from zero to 3.1 with an average of 1.1 (SD = 1.3).

Table 4-2. *Example of an experimental quadruple consisting of two high- and two low-frequency syllables, onsets and offsets are frequency controlled.*

Syllables within one quadruple	
high-frequency	low-frequency
kem [kɛm]	kes [kɛs]
wes [ʋɛs]	wem [ʋɛm]

To each of the 8 high- and 8 low-frequency syllables four endings (*ta* [ta], *li* [li], *wa* [ʋa], and *jo* [jo]) were added in such a way that out of one syllable 4 different disyllabic pseudo-words were created. As a result, 16 pseudo-words were built that had all the same syllable structure, namely CVC.CV (e.g., kem.ta [kɛm.ta]). 8 of the pseudo-words consisted of high-frequency CVC-syllables (e.g., kem - kem.ta [kɛm.ta]); the other 8 pseudo-words had as first syllabic constituents the low-frequency CVC-syllables (e.g., kes.ta [kɛs.ta]). High-frequency CV-syllables were used to serve as the second syllable[21]. All of the used pseudo-words were phonotactically possible strings of Dutch.

Design. The 16 pseudo-words were grouped into homogeneous (high- and low-frequency) and heterogeneous (high- and low-frequency) sets. Homogeneous sets consisted of four pseudo words sharing the first syllable (e.g., kemta, kemli, kemwa, and kemjo). There were homogeneous high-frequency sets as well as homogeneous low-frequency sets, the four pseudo-words again 'derived' from one base syllable (as in kesta, kesli, keswa, and kesjo). Following this procedure, 8 homogeneous high-frequency sets were created and 8 homogeneous low-frequency sets, each of them containing 4 items. Heterogeneous sets were created by regrouping the items from the homogeneous sets; resulting in another 16 item sets. More precisely, the heterogeneous sets consisted of four pseudo-words with different first syllables of the same frequency and the same second syllables as in the homogeneous set (e.g., kemta, binli, merwa, and supjo). Thus, the two

---

[21] For the count *number of occurrence* CV-syllables had an average of 84.8 (SD = 34.19) per million words and for the count *summed frequency of occurrence* an average of 2,821.59 (SD = 1,150.92).

high- and the two low-frequency syllables of each quadruple appeared in a homogeneous set as well as in a heterogeneous context; every item served as its own control. Thus, the experiment consisted of two crossed within-subject factors, namely the factor *Preparation* (two levels, homogeneous versus heterogeneous) and the factor *Frequency* (two levels, high versus low). In total, the experiment consisted of 32 different item sets, containing 64 different items. For a full list of items and their distribution over the four conditions see the Appendix 4-B.

The presentation of blocks of homogeneous and heterogeneous sets alternated in blocks of four sets within the same frequency condition. For example, four high-frequency homogeneous sets were followed by four high-frequency heterogeneous sets; then four low-frequency homogeneous sets and again four low-frequency heterogeneous sets were presented. The remaining 16 sets were presented following the same procedure. The succession of the high- and low-frequency as well as the homogeneous and heterogeneous sets were fully counterbalanced over 32 participants by applying a Latin-Square design.

Procedure and Apparatus. The participants were tested individually in all experiments. The participants were seated in a quiet room in front of a computer screen (iiyamaLM704UT). The distance between participants and screen was approximately 50 cm. Before the experiment started, participants received written instructions about the task. The experiment consisted of alternating learning and test phases. In the learning or rather 'warm-up'-phase, participants were shown sets of four words on the computer screen. They participants were instructed to look silently over the four target words. The four words were presented for 9 seconds with a vertical spacing in 24-points lowercase Arial, white one a black computer screen.

In the test phase, the words were presented one by one on the screen and participants were instructed to read aloud the target words as fast and as accurately as possible. The structure of a trial was as follows: First, the participants saw a warning signal (an asterisk) for 500 ms. Next, the screen was cleared for 500 ms, followed by the display of the target for 1500 ms. Stimuli were presented in white on a black background. Finally, before the start of the next trial there was a blank interval of 500 ms. Thus, the total duration of a trial was 3 seconds. Simultaneously with target onset the voice key was activated for 1500 ms. The prompt disappeared after the response with a delay of 500 ms. The asterisk of the next trial appeared after 100 ms. All the items within each set were repeated five times in a random order, resulting in 20 reading responses per set. On average, an experimental session lasted 30 minutes.

The presentation of the stimuli and the measuring of the reaction times were controlled by the NESU2000 software package. The spoken reactions were registered by a Sennheiser MD211N microphone, which fed into a NESU-Box2 voice key device and a DAT recorder (Sony DTC-55ES). The experimenter sat in the same room as the participant and took note of hesitations, voice key errors, wrong naming responses, and time outs.

RESULTS AND DISCUSSION

Observations leading to wrong or invalid responses (mispronunciations, voice key errors, and hesitations) were not included in the reaction time analysis. Observations deviating from a participant's and an item's mean by more than two standard deviations were considered as outliers and also discarded from the reaction time analysis. 347 (1.7%) trials were treated as errors and 390 (1.9%) as outliers.

Analyses of variance were run with *Preparation* (homogeneous versus heterogeneous) and *Frequency* (high versus low) as independent variables. Two complementary analyses were computed, one treating participants (F_1) and one treating item-quadruples (F_2) as random factor (Clark, 1973)[22].

The mean voice onset latencies, standard deviations and error rates for Experiment 4.1 are summarized in Table 4-3.

---

[22] Since the selection of one item determined the selection of three remaining items within one quadruple, items cannot be considered a random variable in this design.

Table 4-3. *Mean voice onset latencies (in ms), percentage errors, and standard deviations (in parentheses) in Experiment 4.1.*

	Frequency							
	High – frequent				Low - frequent			
Preparation	M	(SD)	% Err	(SD)	M	(SD)	% Err	(SD)
Homogeneous sets	386	(40)	2.0	(1.5)	386	(44)	1.7	(1.5)
Heterogeneous sets	439	(37)	1.5	(1.6)	447	(38)	1.5	(1.5)
Difference scores	-53		0.5		-61		0.2	

There is a significant effect of *Preparation* in the analysis of reaction times ($F_1(1,31) = 122.99$, $MS_e = 857.98$, $p < .001$; $F_2(1,3) = 480.67$, $MS_e = 27.48$, $p < .001$), reflecting faster naming latencies for homogeneous than for heterogeneous sets. *Frequency* did not show a significant main effect ($F_1(1,31) = 1.69$, $MS_e = 317.90$, $p = .204$; $F_2(1,3) = 5.22$, $MS_e = 11.16$, $p = .106$). Most interestingly a significant interaction between *Preparation* and *Frequency* was obtained. By comparing the difference scores for the high-frequency condition (heterogeneous versus homogeneous sets: 439 ms – 386 ms = 53 ms) to the difference scores for the low-frequency condition (447 ms – 386 ms = 61 ms), it can be seen that the preparation effect is in fact larger for the low-frequency sets. Thus, the low-frequency condition gains 8 ms due to the advanced preparation compared to their high-frequency counterparts. The interaction between the factors *Preparation* and *Frequency* is significant ($F_1(1,31) = 5.21$, $MS_e = 101.47$, $p = .05$; $F_2(1,3) = 14.34$, $MS_e = 4.66$, $p = .05$). Post-hoc analyses revealed a significant difference between the high- and the low-frequency sets in the heterogeneous context ($t_1(31) = 2.701$, $p = .05$; $t_2(3) = 4.210$, $p = .05$). Here, no advanced preparation was possible as the four items within these sets had no initial syllable overlap. Thus, the effect of 8 ms reflects a clear (syllable-) frequency effect. Reaction times for the high- and low-frequency sets in the homogeneous context had identical values, the preparation effect wiped out the frequency effect (both ts < 1). The analysis of error rates yielded no significant effects.

SUMMARY AND CONCLUSIONS

The main goal of Experiment 4.1 was to investigate whether or not the *preparation effect* found in an earlier study (Cholin et al., 2004) includes access to the mental syllabary. Preparation effects have been reported in implicit priming studies, when speakers – in an experimental situation - know in advance about the initial part of the to-be-produced utterance (e.g., Meyer, 1990, 1991; and see also Janssen et al., 2002; and Roelofs & Meyer, 1998). Depending on the structure of the previously given information, speakers can prepare more or less of the target utterance. In the implicit priming study using the odd-man-out variant (Cholin et al., 2004), a syllable preparation effect was observed: Sets allowing to prepare for the initial segments and the syllable-structure were produced faster than sets allowing for the preparation of (the same number of) segments only. This syllable preparation effect provides strong evidence for the relevance of syllabic units during speech production, this effect, however, does not ascertain the existence of a mental syllabary where these units are stored and retrieved during production. The remaining question is whether or not the observed syllable preparation effect reflects access to the mental syllabary. The significant syllable frequency effects found in a subsequent study (Cholin et al., submitted) provide strong evidence for the retrieval of syllabic gestural scores from such a mental store.

The present study was conducted to provide an answer to the question sketched above, namely whether or not access to the mental syllabary is involved and is thereby contributing to the syllable preparation effect. The combination of the implicit priming paradigm with the syllable-frequency approach allowed us to address this question. By using a reading variant of the classic implicit priming paradigm, we investigated the voice onset latencies for frequency-manipulated disyllabic pseudo-words in Dutch. Participants were required to read aloud pseudo-words presented on the screen.

Participants were familiarized with each experimental set (containing four target words, with either high- or low-frequency first syllables) before each test phase, thus for items in homogeneous sets participants could prepare for the first syllable. They could, however, not do so for items in heterogeneous sets. The classic preparation effect for shared versus non-shared phonological properties (Meyer, 1990, 1991) was replicated, the homogeneous sets were on average 57 ms faster produced than the heterogeneous sets. The size of this preparation effect differed significantly for low- and high-frequency items. The larger preparation effect for low- than for high-frequency items supports the assumption that the observed preparation effect includes access to the mental syllabary. Otherwise the preparation effect should not have been affected by the manipulation of syllable frequency. In homogeneous sets the first syllable could be prepared for articulation as all relevant information was given beforehand. That was the case for high- as well as for low-frequency sets. Thus, both of the first syllables, could be prepared equally efficiently. In other words, the head start that high-frequency syllables have by a faster syllabary access compared to low-frequency syllables was annihilated by the advanced preparation in the present experiment. This is reflected in the data: In homogeneous sets no syllable frequency effect was obtained. Voice onset latencies for homogeneous high- and low-frequency sets were identical.

A syllable frequency effect was only obtained in the heterogeneous sets where no implicit prime was given. As all of the four pseudo-words in both the high- and the low-frequency heterogeneous sets were different, no preparation was possible. As a consequence, word-form encoding processes could first start with target presentation. Here, the items containing the high-frequency first syllables should be faster produced than low-frequency syllables. From the former syllable-frequency studies (Cholin et al., 2004) we knew that we could expect a rather small but significant effect. And in fact, this effect amounted to 8 ms and is thereby replicating the syllable frequency effect from the former (syllable-frequency) study (9 ms in Exp. 3.2 testing monosyllabic pseudo-words, and 10 ms in Exp. 3.4 testing disyllabic pseudo-words with the frequency-manipulation on the first syllable position).

Thus, we found the predicted interaction between the factors *preparation* and *frequency*; the gain or benefit due to the preparation of the homogeneous sets is significantly larger for the low-frequency sets compared to the high-frequency sets. One might call this the “reversed syllable frequency effect”.

One could object that the effect of syllable frequency has to be attributed to perception rather than to production processes. However, in a recent study, Roelofs (in press) introduced this specific reading variant of the implicit priming paradigm. The performance of reading target-words aloud and object naming of the corresponding picture in an integrated task (Exp. 3 of that study) yielded the same results as previously observed with the implicit priming paradigm (e.g., Meyer, 1990, 1991). Moreover, the results from the above mentioned syllable-frequency study using exactly the same disyllabic pseudo-words as Chapter 3 (Exp. 3.4) cannot be explained by attributing the effects to perception processes. In this experiment, participants learned the to-be-produced target words in a separate learning phase. In the actual test phase participants produced the target words depending on the position of the retrieval cue. Taking both data patterns into account, the attribution of the obtained effect to perceptual processes is highly unlikely.

To conclude, we found an interaction between the two factors *syllable preparation* and *syllable frequency*. This finding was interpreted as evidence for the assumption that – depending on the nature of the implicit prime – access to the mental syllabary is involved in advanced preparation. The results of this study also replicate the syllable frequency effect found in an earlier study. Evidence that stored gestural scores are retrieved from the mental syllabary at the interface of phonological and phonetic encoding is thus provided using another paradigm.

GENERAL DISCUSSION

CHAPTER 5

The fluent production of speech is a very complex human skill. It requires the coordination of several articulatory subsystems. The instructions that lead articulatory movements to execution are the result of the interplay of speech production levels that operate above the articulatory network. During the process of word-form encoding, in particular during *phonological* and *phonetic encoding* processes, the groundwork for the articulatory programs is prepared which then serves the articulators as basic units. This thesis investigated whether or not *syllables* form the basis for the articulatory programs and in particular whether or not these syllable programs are stored, separate from the store of the lexical word-forms, and retrieved to ease articulation.

In this final chapter, first the empirical results of this thesis are summarized and then their implications for theories of speech production are discussed. Finally, potential further research is proposed.

The results of this thesis provide strong evidence that the syllable constitutes a functional unit in speech production. Furthermore, the specific claims for the exact nature of the role of the syllable in speech production as proposed by Levelt, Roelofs and Meyer (1999) were confirmed. The crucial claims tested in this thesis were as follows: Syllables are assumed to play an important role during the process of speech production. More specifically, the theory proposes that syllables are not stored within the word form in the mental lexicon but are created during a late, sub-lexical on-line syllabification process. Furthermore, the existence of a mental syllabary is hypothesized which contains stored syllabic motor programs. This syllabary functions as a mediator between phonological and phonetic encoding.

The first claim of the Levelt et al. model, namely that the word-form that is retrieved from the mental lexicon during phonological encoding, is not (pre-) syllabified was confirmed by prior syllable priming studies (Baumann, 1995; Brand et al., 2003; Evinck, 1997; Schiller, 1997, 1998, 1999, 2000; Schiller & Costa, submitted; Schiller, Costa & Colomé, 2002). If syllable frames were stored, syllable priming effects would have been expected. As discussed, this result speaks against models that assume syllabified word-forms (see for example Dell, 1986, 1988). Rather, it is assumed that syllabification is a late on-line process that can flexibly adjust to the given contextual demands. However, interpreting these data as supporting late on-line syllabification presupposes that the syllable is in fact a relevant functional unit in speech production, an assumption for which strong support has not been previously reported.

Thus, this thesis has two primary aims. First, to demonstrate that the syllable is in fact represented and used as a functional unit during the process of speech production. Second, to specify the *stages* of word form encoding where the syllable comes into play.

SUMMARY OF RESULTS

The experiments in Chapter 2 tested whether evidence for the functional role of the syllable during speech production could be found. The *odd-man-out* variant of the *implicit priming paradigm* was used to investigate this question. This paradigm is known to be sensitive to later stages of word-form encoding and was therefore chosen to tap into the late process of on-line syllabification and possibly syllabary access. In the two experiments, testing two different syllable structures (CVV, Exp. 2.1 and CCVV, Exp. 2.2), participants produced previously learned target words repeatedly within an experimental block. Two kinds of blocks were presented, constant and variable blocks. In constant blocks, items within response sets shared initial segments as well as the syllable structure of the first syllable. In variable sets, only initial segments between items in response sets were identical but not the syllable structure. The homogeneity of segmental overlap in both the constant and the variable sets was a crucial requirement to investigate whether, in addition to segmental information, speakers use information about the syllabic structure in order to prepare the response. The prediction was that speakers – under the assumption that the syllable is a relevant processing unit in speech production and is stored and accessed independently during word form encoding – need knowledge about the current syllabic structure in order to prepare for a target utterance. The results confirmed the prediction that constant sets allow for more preparation than the variable sets; the preparation effect was significantly larger for constant than for variable sets.

In the preparation task, syllabic information is relevant in the sense that any deviating CV-structure, i.e., an item with a different syllable, reduces the preparation effect. Thus, the results of Chapter 2 offer strong evidence for the relevance of syllables in word-form encoding. Furthermore, taking the results of prior syllable priming studies into account, it was concluded that this method taps into the right level of processing where syllables are, in fact, encoded, namely, during the sub-lexical on-line syllabification process. However, the conclusion that the preparation effect (in constant sets) includes access to the mental syllabary remains in the end speculative; clear evidence for the existence of a mental syllabary that releases stored phonetic syllable representations still had to be revealed. It had to be established that speakers access such a mental store during word-form encoding. Only after demonstrating the existence of such a store can we re-evaluate the underlying mechanisms involved in the observed preparation effects.

Effects of syllable frequency would certainly provide clear evidence for the existence of a mental syllabary. A mental storage implies different retrieval times for high- and low-frequency syllables, as is the case for high- and low-frequency words (Jescheniak & Levelt, 1994). In a series of experiments, the production of high- and low-frequency syllables was contrasted in Chapter 3. A newly developed paradigm, a *Symbol-Position Association Learning Task* and a very carefully controlled syllable material set was used to test for syllable frequency effects. Significant syllable frequency effects were found by investigating monosyllabic (Exp. 3.2) and disyllabic Dutch pseudo-words bearing the frequency-manipulation on the *first* syllable (Exp. 3.4). No effect of syllable frequency was found for disyllabic Dutch pseudo-words with high- or low-frequency *second* syllables (Exp. 3.3). Taken together, the results of Chapter 3 provide clear evidence for the existence of a mental syllabary. The claim that the syllable is an important unit of speech production is again confirmed. These results also increase the plausibility that access to the mental syllabary may in fact have been involved in the preparation effects reported in Chapter 2.

The study reported in Chapter 4 was carried out in order to verify the interpretation that the advanced preparation on the basis of segmental and syllabic structure information includes the retrieval of pre-compiled gestural scores from memory. Furthermore, this study allows for the replication of the syllable frequency effects with another paradigm. A reading variant of the classic implicit priming paradigm was used to contrast the voice onset latencies for frequency-manipulated disyllabic Dutch pseudo-words. Homogeneous versus heterogeneous high- and low-frequency item sets were presented; thus we crossed two factors, *syllable preparation* (homogeneous versus heterogeneous) and *syllable frequency* (high- versus low-frequency). The crucial prediction was for an interaction between these two factors: The preparation effect for items with low-frequency initial syllables was predicted to be larger than for items with high-frequency first syllables. The difference score between the low-frequency homogeneous and heterogeneous sets should be larger than the difference score between the high-frequency homogeneous and heterogeneous sets. This interaction is predicted as accessing and retrieving gestural scores for low-frequency syllables from the mental syllabary takes longer than the retrieval of high-frequency syllables, therefore the preparation effect for low-frequency (heterogeneous – homogeneous) sets should yield a larger gain through the preparation. This prediction was confirmed.

An additional prediction is that the frequency effects should only be observed in the heterogeneous sets. As homogeneous sets allow full preparation of the first syllable, up to and including the retrieval of the stored syllable motor programs, the benefit of retrieving a high- compared to a low-frequency syllable should be lost in the homogeneous sets. The prediction that the advanced preparation wipes out the frequency effect is confirmed; there is no difference between the homogeneous high- and low-frequency sets. Only by contrasting the heterogeneous high- and low-frequency sets, a significant syllable-frequency was obtained. To conclude, the predicted interaction was found; the benefit from the preparation for low-frequency (homogeneous versus heterogeneous) sets is larger compared to high-frequency sets. These results suggest that advanced preparation involves access to the mental syllabary when all relevant information, namely segmental and syllabic information, is given.

IMPLICATIONS FOR A MODEL OF SPEECH PRODUCTION

The results reported in this thesis make the relevance of the syllable as speech production unit clear. Moreover, these results provide insight into the processes underlying word-form encoding. The Levelt et al. model's claims, particularly with respect to the specific "when" and "where" the syllable comes into play, were confirmed.

On the basis of the results of Chapter 4, the implications for a model of speech production can be specified in further detail. The experiment reported in this Chapter was carried out to get a complete picture of the processes at the interplay of phonological and phonetic encoding. The combination of the two previously used methods was motivated by the question of whether or not advanced preparation includes access to the mental syllabary. The prediction that the frequency effect would be annihilated by the preparation effect was based on the interpretation of the results of Chapter 2. It was argued that the larger preparation effect for sets where the segmental and the syllabic structure was shared across target items in response sets was due to advanced preparation. In sets where only segmental information was given, the preparation effect was significantly smaller. Only in the constant sets, where syllabic information is given, can the syllabification process begin to incrementally put the segments together and create the first syllable. Whether or not these abstract phonological syllables then feed into the mental syllabary and activate their motor programs so that phonetic encoding could also be completed was assumed but could not be verified on the basis of the results of Chapter 2. The syllable frequency effects observed in Chapter 3 made this assumption likely, as now there was strong evidence for the existence of a mental syllabary.

Coming back to where I started, the prediction of Chapter 4 was the interaction of the factors *syllable preparation* and *syllable frequency* under the assumption that preparation involves access to the mental syllabary and will thereby annihilate the frequency effect: In homogeneous sets, both in the high and the low-frequency condition, preparation could start before the target was displayed. All relevant information, that is segmental and syllabic information, is given as the first syllable across targets is identical in each set. Thus, the respective segments can be retrieved from memory. As information about the corresponding syllable is also given, the first syllable can be assembled incrementally and can then access its gestural score in the mental syllabary. In a final step, phonetic encoding of the first syllable is completed and the syllable can be transmitted to the articulatory network where it is either stored in an articulatory buffer until the second syllable is also available or is immediately uttered. This advanced preparation can be done for the high- and for the low- frequency sets with equal efficiency. As a consequence, the frequency-difference that normally manifests itself in faster retrieval procedures for high-frequency syllables does not apply here. Only in heterogeneous sets did the syllable frequency effect appear as here the “race” between high- and low-frequency syllables is first opened with target presentation. Hence, the high-frequency syllables have again an advantage compared to low-frequency syllables. If the retrieval of stored units were *not* included in the preparation effect, than the preparation effect should not have been affected by the syllable-frequency manipulation. The syllable frequency effect observed in Chapter 3 is replicated by these results with the same small but highly significant effect using another paradigm.

Taken together, the results of Chapter 4 strongly suggest that the preparation effect involves access to the mental syllabary when syllabic information is also given beforehand. The claim that the syllable constitutes a functional unit during speech production is confirmed and also the following claims about where and when the syllable comes into play. As prior syllable priming studies could not find evidence for the *early* relevance of syllable units in the process of word-form encoding, the results from the first preparation studies (Chapter 2) were interpreted as syllable structure effects and were taken as evidence that syllables emerge first at the rather late stage of on-line syllabification. The results of Chapter 3 demonstrate significant syllable frequency effects that clearly confirm the assumption of retrieved gestural scores. Furthermore, the results of the two studies investigating disyllabic pseudo-words provide even further insight into the process of phonological and phonetic encoding. The finding that only a syllable frequency effect for the first but not for the second syllable in disyllabic pseudo-word could be obtained was interpreted with respect to the point in time when speakers start their articulation. Two possible ways were pointed out: (i) Speakers may wait until both syllables are retrieved from storage and the phonetic encoding for both units is completed before articulation is initiated. (ii) Speakers start articulation as soon as the first syllable is phonetically encoded and start articulation while the phonetic encoding of the second one is still to be completed. Importantly, the phonological encoding - in both cases - must be completed. The phonological encoding procedures must operate on the phonological word as a whole; some preview is required in order to allow for stress and intonation assignment. The subsequent process of on-line syllabification is an incremental, left-to-right procedure in which successive syllables are created. This can lead to temporal overlap of processes: The first syllable's gestural score can be successfully retrieved from storage, while the second and potential later syllables are still under construction. On the basis of the current data pattern, it was concluded that speakers in fact – at least in this experimental situation – started articulation right after the retrieval of the first syllable. Accordingly, the frequency of the second syllable has no influence on the reaction times of the target word's onset. It was concluded that speaker's are flexible in their decision of how much of the phonological word must also be phonetically encoded before speech onset.

To summarize, the results of this thesis provide strong evidence for the assumption that the syllable plays a crucial role in speech production and supports the notion that the syllable comes into play at the interface of phonological and phonetic encoding and involves access to the mental syllabary. The crucial claims concerning syllabification and the existence of a mental store for syllables proposed by Levelt et al. (1999) are confirmed. These results speak against other theories of speech production that do not assume two separate stores for word(-form)s and syllables (see for example Dell, 1986, 1988).

FURTHER RESEARCH

Having demonstrated syllable frequency effects, subsequent studies investigating the interaction between different sources of frequency could provide further support for the proposed model of word-form encoding. For example, further evidence for the claim that word-forms and syllables are stored independently would come from the finding of additivity between word and syllable frequency effects. This must be addressed by further research.

Phonological restrictions on syllable formation in Dutch only allowed for the investigation of two syllable types (e.g., [lei], Consonant-Vowel (CVV) and [spui], CCVV). Since speakers of Dutch tend to insert an epenthetic schwa in CVCC syllables (e.g., melk → meløk; Van Donselaar, Kuijpers, Cutler, 1999), this syllable structure could not be investigated. An extension to English and German (which do not 'allow' for schwa epenthesis) provides the opportunity to investigate additional syllable types.

The extension to other languages might also include the extension to non-Indo-European languages, e.g., Mandarin Chinese. Recent findings from Mandarin Chinese (Chen, Chen, Dell, 2002) offer insight into a tone language that has far fewer syllable types compared to languages such as Dutch and English (Dutch has more than 12,000 different syllables whereas Mandarin Chinese has a syllable inventory of 400 different syllables). Additionally, syllables in Mandarin Chinese are not resyllabified in connected speech, thereby making the storage of phonological syllables an attractive option. In support of stored phonological syllables is the observation of syllable exchange errors in Mandarin Chinese but not in Dutch (Chen, 2000). These errors are likely to arise at the level of phonological encoding when phonological syllables as whole units are retrieved and misplaced. Also, unlike what has been found for Dutch and other Indo-European languages (see above), in Mandarin Chinese syllable priming effects have been obtained (Chen, Lin, & Ferrand, 2003). The different results for Dutch and Mandarin Chinese could be due to the contrasting properties of the respective languages. In Mandarin Chinese, syllabic units may be represented earlier in the process of speech production. However, this would not deny the existence of a mental store for *phonetic* syllables at a later stage.

The results of Chapters 3 and 4 suggest that both the high- and the low-frequency syllables are retrieved from the mental syllabary. However, on the basis of these results it cannot be decided whether or not there are in fact two operating routes in phonetic encoding of syllables, i.e., (i) retrieving pre-compiled syllabic programs and (ii) on-line assembly of phonetic programs. As discussed, the on-line assembly may be the dominant operation for the creation of new and low-frequency syllables. However, it cannot be excluded that the on-line assembly route is in fact *always* active, that is, running in parallel to the retrieval route. (Whichever route is then fastest will win the race.) Potentially, functional neuroimaging techniques, such as fMRI, will provide a tool to distinguish between these two possibilities. Also neurolinguistic approaches, such as the investigation of *speech apraxia* (the deficits exhibited by apraxics seem to be tied to the articulation of syllabic units) may help to gain insights into potentially different routes.

To conclude, throughout this thesis it could be shown that syllables are in fact relevant units in speech production. Levelt and colleagues (1999, see also Levelt, 1989) explicitly incorporate syllables in their model of speech production. They made detailed claims about the functional role the syllable is assumed to play during word-form encoding. Theoretical arguments ("resyllabification") as well as the absence of a syllable-priming effect in prior on-line studies (see above) support the claim that syllables are not stored within the word-form in the mental lexicon. The claims that syllabification is a late on-line process and that syllables are stored in a mental syllabary were confirmed by this thesis. It provides evidence for a repository of pre-compiled gestural scores for syllables that contribute to the fluency and accuracy of spoken language. Future research has to further specify the nature of word-form encoding, for example, whether all, new, low and high-frequency syllables are stored in the mental syllabary and if and how other operating routes are involved in the process of phonetic encoding.

REFERENCES

Abercrombie, D. (1967). Elements of General Phonetics. Edinburgh: Edinburgh University Press.

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, *88*, 148-159.

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

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (in press). Predictors of picture naming speed. *Behavior Research Methods, Instruments, and Computers.*

Ashby, J., & Rayner, K. (in press). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes*.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical data base (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language, 39*, 331-346.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. (Revised ed.). Hove, UK: Psychology Press.

Bagemihl B. (1995). Language games and related areas. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 697-712). Cambridge, MA: Blackwell.

Baumann, M. (1995). *The production of syllables in connected speech*. Ph.D. dissertation, Nijmegen University.

Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell. Untersuchungen an englischen und deutschen Versprechern* [The representation of the speech production process in a spreading activation model: Studies of German and English speech errors]. Tübingen: Niemeyer.

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206-244). Cambridge, MA: Blackwell.

Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.

Brand, M., Rey, A., & Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language*, *48*, 435-443.

Brand, M., Rey, A., Peereman, R., & Spieler, D. (2002). Naming bisyllabic words: A large scale study. *Abstracts of Psychonomic Society, 7*, 94.

Carreiras, M., & Perea, M. (in press). Naming pseudowords in Spanish: Effects of syllable frequency in production. *Brain and Language*.

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia, 43*, 15-26.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word form encoding in Mandarin Chinese as assessed by the implicit priming paradigm. *Journal of Memory and Language*, *46*, 751-781.

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology*, *45*, 107-120.

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (submitted). Effects of syllable frequency in speech production.

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*, 47-61.

Costa, A. & Sebastián-Gallés, N. (1998). Abstract phonological structure in language production: Evidence from Spanish. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 886-903.

Crompton, A. (1981). Syllables and segments in speech production. *Linguistics, 19*, 663-716.

Cutler, A, McQueen, J., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. In E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in honor of Jacques Mehler* (pp.181-194). Cambridge, MA: MIT Press.

Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes, 12*, 839-845.

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

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

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124-42.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*, 159-161.

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

Evinck, S. (1997). *Production de la parole en français: Investigation des unités impliquées dans l'encodage phonologique des mots* [Speech production in French: Investigation of the units implied during the phonological encoding of words]. Unpublished Ph.D. dissertation, Bruxelles University.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllable units. *Journal of Memory and Language*, *35*, 708-723.

Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, *25*, 458-470.

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

Fujimura, O. & Lovins, J. B. (1978). Syllables as concatenative phonetic units. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 107-130). Amsterdam: North-Holland.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation. Advances in research and theory* (pp. 133-177). New York, NY: Academic Press.

Garrett, M. F. (1982). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Speech and talk*, vol. 1 (pp. 177-220). New York, NY: Academic Press.

Goldstein, L. & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). Berlin: Mouton de Gruyter.

Hendriks, H. & McQueen, J. (Eds.) (1996). *Annual Report 1995*. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Hooper, J. B. (1972). The syllable in phonological theory. *Language*, *48*, 525-540.

Howard, D., & Smith, K. (2002). The effects of lexical stress in aphasic word production. *Aphasiology*, *16*, 198-237.

Indefrey P., & Levelt, W. J. M. (2000). The neural correlates of languages production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845-865). Cambridge, MA: MIT Press.

Janssen, P. D., Roelofs, A., & Levelt, W. J. M. (2002). Inflectional frames in language production. *Language and Cognitive Processes*, *17*, 209-236.

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

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kerzel, D. & Bekkering H. (2000). Motor activation from visible speech: Evidence from stimulus-response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 634-647.

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

Ladefoged, P. (1982). *A course in phonetics*. Second edition. San Diego, CA: Harcourt Brace Jovanovich.

Laganaro, M. (2002). Are syllables represented and retrieved during phonological encoding? An integration of psycholinguistic and neurolinguistic studies. Poster presented at Euresco conference on *The Science of Aphasia: Functional Neuroimaging Studies of Language and its Impairments*. Acquafredda di Maratea, Italy.

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. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239-269.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201-212.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217-245). New York: Springer.

Luce, R. A. (1959). *Individual choice behavior*. New York: Wiley.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 570-590.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*, 323-350.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*, 499-546.

Mayer, J., Ackermann, H., Dogil, G., Erb, M., & Grodd, W. (2003). Syllable retrieval versus on-line assembly: fMRI examination of the syllabary. In M. J. Solé, D. Recasens, & J. Romero (Eds.), Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003), Barcelona, 3-9 August, 2003 (pp. 2541-2543).

Meijer, P. J. A. (1996). Suprasegmental structures in phonological encoding: The CV-structure. *Journal of Memory and Language*, *35*, 840-853.

Meyer, A. & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 1146-1160.

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

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

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 452-467.

Nooteboom, S. (1969). The tongue slips into pattern. In A. G. Sciarone, A. J. von Essen, & A. A. van Raad (Eds.), *Nomen: Leyden studies in linguistics and phonetics* (pp. 114-132). The Hague: Mouton.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281.

Perea, M. & Carreiras, M. (1996). Efectos de frecuencia silábica en la tarea de pronunciación mixta ['Effects of syllable frequency in the mixed pronunciation task']. *Psicológica, 17*, 425-440.

Perea, M. & Carreiras, M. (1998). Effects of syllable frequency and syllable neighbourhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 134-144.

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

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*, 854-876.

Roelofs, A. (1997a). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12*, 657-693.

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

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904-921.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173-200.

Roelofs, A. (in press). Seriality of phonological encoding in naming objects and reading their names. *Memory and Cognition*.

Roelofs, A. & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 922-939.

Roelofs, A. (2002). Syllable structure effects turn out to be word length effects; Comment on Santiago et al. (2000). *Language and Cognitive Processes*, *17*, 1-13.

Santiago, J., MacKay, D. G., & Palma, A. (2002). Length effects turn out to be syllable structure effects: Response to Roelofs (2002). *Language and Cognitive Processes, 17*, 15-29.

Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognitive Processes, 15*, 1-44.

Schiller, N. O. (1997). *The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography*. PhD dissertation, Nijmegen University (MPI series; 2).

Schiller, N. O. (1998). The effect of visually masked primes on the naming latencies of words and pictures. *Journal of Memory and Language*, *39*, 484-507.

Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language*, *68*, 300-305.

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 512-528.

Schiller, N. O. (in press). The onset effect in word naming. *Journal of Memory and Language*.

Schiller, N. O. & Costa, A. (submitted). The role of the syllable in phonological encoding: Evidence from masked priming?

Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 35-59). Berlin: Mouton de Gruyter.

Schiller, N. O., Fikkert, P., & Levelt, C. C. (in press). Stress priming in picture naming: An SOA study. *Brain and Language*.

Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence form the syllabification of intervocalic consonants. *Language and Speech*, *40*,103-140.

Schiller, N. O., Meyer, A. S., Baayen, R. H., & Levelt, W. J. M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, *3*, 8-28.

Schriefers, H. & Teruel, E. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology, 11*, 17-50.

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

Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations. Part II* (pp. 337-383). Dordrecht: Foris.

Sevald, C. A., Dell, G., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, *34*, 807-820.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 295-342). New York: Halsted Press.

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processing in language* (pp. 17-51). Hillsdale, NJ: Erlbaum.

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

Stemberger J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, 56, 235-259.

Taylor, T.E. & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 117-138.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, *15*, 49-74.

Treiman, R. & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, *27*, 87-104.

Van Donselaar, W. A., Kuijpers, C. T. L., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language, 41*, 59-77.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.

Vroomen, J. & de Gelder, B. (1994). Speech segmentation in Dutch: no role for the syllable. In *Proceedings of the 1994 International Conference on Spoken Language Processing, September 18-22, 1994, Yokohama* (pp. 1135-1138).

Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37*, 356-381.

Wilshire, C. E. & Nespoulous, J.-L. (2003). Syllables as units in speech production: data from aphasia. *Brain and Language*, *84*, 424-447.

Zwitserlood, P., Schriefers, H., Lahiri, A., & van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 260-271.

APPENDICES

Appendix 2-A

Materials for Experiment 2.1

Four-item sets		Three-item sets	
Constant sets	Variable sets	Constant sets	Variable sets
lei.den ([to] lead)	hui.len (to cry)	lei.den	hui.len
lei.dde (led)	huil.de (cried)		
lei.der (leader)	hui.ler (person who cries)	lei.der	hui.ler
lei.dend (leading)	hui.lend (crying)	lei.dend	hui.lend
ha.ten ([to] hate)	boe.nen ([to] polish)	ha.ten	boe.nen
haa.tte (hated)	boen.de (polished)		
ha.ter (hater)	boe.ner (person who polishes)	ha.ter	boe.ner
ha.tend (hating)	boe.nend (polishing)	ha.tend	boe.nend
po.ten ([to] plant)	ro.ken ([to] smoke)	po.ten	ro.ken
poo.tte (planted)	rook.te (smoked)		
po.ter (person who plants)	ro.ker (smoker)	po.ter	ro.ker
po.tend (planting)	ro.kend (smoking)	po.tend	ro.kend
wa.den ([to] wade)	ha.ken ([to] crochet)	wa.den	ha.ken
waa.dde (waded)	haak.te (crocheted)		
wa.der (person who wades)	ha.ker (person who crochets)	wa.der	ha.ker
wa.dend (wading)	ha.kend (croching)	wa.dend	ha.kend

Note. Vowels in open syllables, e.g., *ro.ken* and vowels marked twice in orthography, e.g., *roo.kte*, both have a long pronunciation in Dutch.

Appendix 2-B

Materials for Experiment 2.2

Four-item sets		Three-item sets	
Constant sets	Variable sets	Constant sets	Variable sets
kne.den ([to] knead)	kla.gen ([to] complain)	kne.den	kla.gen
knee.dde (kneaded)	klaag.de (complained)		
kne.der (kneader)	kla.ger (complainer)	kne.der	kla.ger
kne.dend (kneading)	kla.gend (complaining)	kne.dend	kla.gend
spui.en ([to] drain)	spoe.len ([to] rinse)	spui.en	spoe.len
spui.de (drained)	spoel.de (rinsed)		
spui.er (person who drains)	spoe.ler (person who rinses)	spui.er	spoe.ler
spui.end (draining)	spoe.lend (rinsing)	spui.end	spoe.lend
pra.ten ([to] speak)	dwa.len ([to] wander)	pra.ten	dwa.len
praa.tte (spoke)	dwaal.de (wandered)		
pra.ter (speaker)	dwa.ler (wanderer)	pra.ter	dwa.ler
pra.tend (speaking)	dwa.lend (wandering)	pra.tend	dwa.lend
plei.ten ([to] argue)	sle.pen ([to] drag)	plei.ten	sle.pen
plei.tte (argued)	sleep.te (dragged)		
plei.ter (arguer)	sle.per (person who drags)	plei.ter	sle.per
plei.tend (arguing)	sle.pend (draging)	plei.tend	sle.pend

Note. Vowels in open syllables, e.g., *sle.pen* and vowels marked twice in orthography, e.g., *slee.pte*, both have a long pronunciation in Dutch.

Appendix 3-A

Experimental materials for the Pre-test

Set numbers	*High-frequency words*	*Word-form freq.*	*Low-frequency words*	*Word-form freq.*
1	wa.ter (water)	368	wa.sem (steam)	0.6
1	ra.pport (report)	68	ra.dijs (radish)	1.1
1	ta.fel (table)	250	tan.dem (tandem)	0.7
1	fo.to (photo)	108	fa.kir (fakir)	0.7
2	sol.daat (soldier)	114	spa.gaat (splits)	0
2	re.gel (line)	140	ras.ter (fence)	0.6
2	ein.de (end)	157	e.land (elk)	1.1
2	ka.mer (room)	368	ka.jak (kayak)	0.1
3	sui.ker (sugar)	40	sal.to (somersault)	1.1
3	moe.der (mother)	602	mij.ter (mitre)	0.5
3	bo.dem (bottom)	49	ba.tik (batik)	0.4
3	ho.tel (hotel)	93	ho.bo (oboe)	0.5
4	vi.nger (finger)	158	vij.zel (screw jack)	1
4	a.vond (evening)	287	as.ter (aster)	1.2
4	schou.der (shoulder)	193	scho.ffel (grubber)	0.9
4	ko.ffie (coffee)	112	ko.kos (coconut)	0.2

Appendix 3-B

Experimental syllable-quadruples

Quad. Nr.	*High-freq. sets*	*Frequency counts*		*Low-freq. sets*	*Frequency counts*	
		No. of occurrence	*No. of summed frequency*		*No. of occurrence*	*No. of summed frequency*
1	bin [bɪn]	6.4	127.26	bing [bɪŋ]	0.17	0.48
1	ning [nɪŋ]	23.5	1,192.57	nin [nɪn]	0.02	0.00
2	bur [bʏr]	4.76	214.79	bug [bʏx]	0.02	0.38
2	lug [lʏx]	2.36	49.71	lur [lʏr]	0.21	0.67
3	ket [kɛt]	4.86	60.81	keg [kɛx]	0.07	0.24
3	teg [tɛx]	6.71	190.57	tet [tɛt]	0.67	2.43
4	kem [kɛm]	1.48	62.24	kes [kɛs]	1.19	3.1
4	wes [ʋɛs]	3.24	162.60	wem [ʋɛm]	0.02	0.1
5	luk [lʏk]	1.74	209.14	lup [lʏp]	0.19	0.67
5	sup [sʏp]	4.17	82.55	suk [sʏk]	0.26	3.02
6	mer [mɛr]	5	313.12	meg [mɛx]	0.07	1.4
6	reg [rɛx]	8.19	339.86	rer [rɛr]	0.05	0.00
7	sjam [ʃɑm]	1.67	34.12	sjag [ʃɑx]	0.05	0.07
7	wag [ʋɑx]	4.31	546.24	wam [ʋɑm]	0.86	3.67
8	tur [tʏr]	4.95	227.81	tug [tʏx]	0.71	1.98
8	zug [zʏx]	1.86	63.81	zur [zʏr]	0.33	2.33

Appendix 3-C

Experimental item-sets for Experiment 3.1

Set number (Quadruple Nr. in parenthesis)	*High-frequency sets*	*Low-frequency sets*
1 (1)	ning [nɪŋ]	nin [nɪn]
1 (2)	lug [lʏx]	lur [lʏr]
1 (6)	mer [mɛr]	meg [mɛx]
1 (7)	sjam [ʃɑm]	sjag [ʃɑx]
2 (1)	bin [bɪn]	bing [bɪŋ]
2 (2)	bur [bʏr]	bug [bʏx]
2 (6)	reg [rɛx]	rer [rɛr]
2 (7)	wag [ʋɑx]	wam [ʋɑm]
3 (3)	ket [kɛt]	keg [kɛx]
3 (4)	wes [ʋɛs]	wem [ʋɛm]
3 (5)	luk [lʏk]	lup [lʏp]
3 (8)	tur [tʏr]	tug [tʏx]
4 (3)	teg [tɛx]	tet [tɛt]
4 (4)	kem [kɛm]	kes [kɛs]
4 (5)	sup [sʏp]	suk [sʏk]
4 (8)	zug [zʏx]	zur [zʏr]

Appendix 3-D

Pairing of syllables into frequency-homogeneous sets in Experiment 3.2

Set number	*High-frequency sets*	*Low-frequency sets*
1	lug – bin	lur – bing
2	mer – sjam	meg – sjag
3	ning – reg	nin – rer
4	wag – bur	wam – bug
5	luk – teg	lup – tet
6	wes – sup	wem – suk
7	ket – zug	keg – zur
8	tur – kem	tug- kes

Appendix 3-E

Materials for Experiment 3.3

Quadruple Nr.	*High-frequency sets*	*Low-frequency sets*
1	ta.bin [ta.bɪn]	ta.bing [ta.bɪŋ]
1	ta.ning [ta.nɪŋ]	ta.nin [ta.nɪn]
2	ko.bur [ko.bʏr]	ko.bug [ko.bʏx]
2	ko.lug [ko.lʏx]	ko.lur [ko.lʏr]
3	si.ket [si.kɛt]	si.keg [si.kɛx]
3	si.teg [si.tɛx]	si.tet [si.tɛt]
4	li.kem [li.kɛm]	li.kes [li.kɛs]
4	li.wes [li.ʋɛs]	li.wem [li.ʋɛm]
5	ra.luk [ra.lʏk]	ra.lup [ra.lʏp]
5	ra.sup [ra.sʏp]	ra.suk [ra.sʏk]
6	wa.mer [ʋa.mɛr]	wa.meg [ʋa.mɛx]
6	wa.reg [ʋa.rɛx]	wa.rer [ʋa.rɛr]
7	ti.sjam [ti.ʃɑm]	ti.sjag [ti.ʃɑx]
7	ti.wag [ti.ʋɑx]	ti.wam [ti.ʋɑm]
8	jo.tur [jo.tʏr]	jo.tug [jo.tʏx]
8	jo.zug [jo.zʏx]	jo.zur [jo.zʏr]

Appendix 3-F

Pairing of items into frequency-homogeneous sets in Experiment 3.3

Set number	*High-frequency sets*	*Low-frequency sets*
1	kolug – tabin	kolur – tabing
2	wamer – tisjam	wameg – tisjag
3	taning – wareg	tanin – warer
4	tiwag – kobur	tiwam – kobug
5	raluk – siteg	ralup – sitet
6	liwes – rasup	liwem – rasuk
7	siket – jozug	sikeg – jozur
8	jotur – likem	jotug- likes

Appendix 3-G

Materials for Experiment 3.4

Quadruple Nr.	*High-frequency sets*	*Low-frequency sets*
1	bin.ta [bɪn.ta]	bing.ta [bɪŋ.ta]
1	ning.ta [nɪŋ.ta]	nin.ta [nɪn.ta]
2	bur.ko [bʏr.ko]	bug.ko [bʏx.ko]
2	lug.ko [lʏx.ko]	lur.ko [lʏr.ko]
3	ket.li [kɛt.li]	keg.li [kɛx.li]
3	teg.li [tɛx.li]	tet.li [tɛt.li]
4	kem.ra [kɛm.ra]	kes.ra [kɛs.ra]
4	wes.ra [ʋɛs.ra]	wem.ra [ʋɛm.ra]
5	luk.mo [lʏk.mo]	lup.mo [lʏp.mo]
5	sup.mo [sʏp.mo]	suk.mo [sʏk.mo]
6	mer.wa [mɛr.ʋa]	meg.wa [mɛx.ʋa]
6	reg.wa [rɛx.ʋa]	rer.wa [rɛr.ʋa]
7	sjam.ti [ʃɑm.ti]	sjag.ti [ʃɑx.ti]
7	wag.ti [ʋɑx.ti]	wam.ti [ʋɑm.ti]
8	tur.jo [tʏr.jo]	tug.jo [tʏx.jo]
8	zug.jo [zʏx.jo]	zur.jo [zʏr.jo]

Appendix 3-H

Pairing of items into frequency-homogeneous sets in Experiment 3.4

Set number	*High-frequency sets*	*Low-frequency sets*
1	lugko – binta	lurko – bingta
2	merwa – sjamti	megwa – sjagti
3	ningta – regwa	ninta – rerwa
4	wagti – burko	wamti – bugko
5	lukmo– tegli	lupmo – tetli
6	wesra – supmo	wemra – sukmo
7	ketli – zugjo	kegli – zurjo
8	turjo – kemra	tugjo- kesra

Appendix 4-A

Experimental item-quadruples for Experiment 4.1

Quad. Nr.	*High-freq. sets*	*Frequency counts*		*Low-freq. sets*	*Frequency counts*	
		No. of occurrence	*No. of summed frequency*		*No. of occurrence*	*No. of summed frequency*
1	bin [bɪn]	6.4	127.26	bing [bɪŋ]	0.17	0.48
1	ning [nɪŋ]	23.5	1,192.57	nin [nɪn]	0.02	0.00
2	kem [kɛm]	1.48	62.24	kes [kɛs]	1.19	3.1
2	wes [ʋɛs]	3.24	162.60	wem [ʋɛm]	0.02	0.1
3	luk [lʏk]	1.74	209.14	lup [lʏp]	0.19	0.67
3	sup [sʏp]	4.17	82.55	suk [sʏk]	0.26	3.02
4	mer [mɛr]	5	313.12	meg [mɛx]	0.07	1.4
4	reg [rɛx]	8.19	339.86	rer [rɛr]	0.05	0.00

Appendix 4-B

Materials for Experiment 4.1

Set	*Homogeneous sets*		*Heterogeneous sets*	
nr.	*High-frequency*	*Low-frequency*	*High-frequency*	*Low-frequency*
1	bin.ta [bɪn.ta]	bing.ta [bɪŋ.ta]	bin.ta [bɪn.ta]	bing.ta [bɪŋ.ta]
1	bin.li [bɪn.li]	bing.li [bɪŋ.li]	kem.li [kɛm.li]	kes.li [kɛs.li]
1	bin.wa [bɪn.ʋa]	bing.wa [bɪŋ.ʋa]	sup.wa [sʏp.ʋa]	suk.wa [sʏk.ʋa]
1	bin.jo [bɪn.jo]	bing.jo [bɪŋ.jo]	mer.jo [mɛr.jo]	meg.jo [mɛx.jo]
2	kem.ta [kɛm.ta]	kes.ta [kɛs.ta]	kem.ta [kɛm.ta]	kes.ta [kɛs.ta]
2	kem.li [kɛm.li]	kes.li [kɛs.li]	bin.li [bɪn.li]	bing.li [bɪŋ.li]
2	kem.wa [kɛm.ʋa]	kes.wa [kɛs.ʋa]	mer.wa [mɛr.ʋa]	meg.wa [mɛx.ʋa]
2	kem.jo [kɛm.jo]	kes.jo [kɛs.jo]	sup.jo [sʏp.jo]	suk.jo [sʏk.jo]
3	sup.ta [sʏp.ta]	suk.ta [sʏk.ta]	sup.ta [sʏp.ta]	suk.ta [sʏk.ta]
3	sup.li [sʏp.li]	suk.li [sʏk.li]	mer.li [mɛr.li]	meg.li [mɛx.li]
3	sup.wa [sʏp.ʋa]	suk.wa [sʏk.ʋa]	kem.wa [kɛm.ʋa]	kes.wa [kɛs.ʋa]
3	sup.jo [sʏp.jo]	suk.jo [sʏk.jo]	bin.jo [bɪn.jo]	bing.jo [bɪŋ.jo]
4	mer.ta [mɛr.ta]	meg.ta [mɛx.ta]	mer.ta [mɛr.ta]	meg.ta [mɛx.ta]
4	mer.li [mɛr.li]	meg.li [mɛx.li]	sup.li [sʏp.li]	suk.li [sʏk.li]
4	mer.wa [mɛr.ʋa]	meg.wa [mɛx.ʋa]	bin.wa [bɪn.ʋa]	bing.wa [bɪŋ.ʋa]
4	mer.jo [mɛr.jo]	meg.jo [mɛx.jo]	kem.jo [kɛm.jo]	kes.jo [kɛs.jo]
5	ning.ta [nɪŋ.ta]	nin.ta [nɪn.ta]	ning.ta [nɪŋ.ta]	nin.ta [nɪn.ta]
5	ning.li [nɪŋ.li]	nin.li [nɪn.li]	wes.li [ʋɛs.li]	wem.li [ʋɛm.li]
5	ning.wa [nɪŋ.ʋa]	nin.wa [nɪn.ʋa]	luk.wa [lʏk.ʋa]	lup.wa [lʏp.ʋa]
5	ning.jo [nɪŋ.jo]	nin.jo [nɪn.jo]	reg.jo [rɛx.jo]	rer.jo [rɛr.jo]
6	wes.ta [ʋɛs.ta]	wem.ta [ʋɛm.ta]	wes.ta [ʋɛs.ta]	wem.ta [ʋɛm.ta]
6	wes.li [ʋɛs.li]	wem.li [ʋɛm.li]	ning.li [nɪŋ.li]	nin.li [nɪn.li]
6	wes.wa [ʋɛs.ʋa]	wem.wa [ʋɛm.ʋa]	reg.wa [rɛx.ʋa]	rer.wa [rɛr.ʋa]
6	wes.jo [ʋɛs.jo]	wem.jo [ʋɛm.jo]	luk.jo [lʏk.jo]	lup.jo [lʏp.jo]
7	luk.ta [lʏk.ta]	lup.ta [lʏp.ta]	luk.ta [lʏk.ta]	lup.ta [lʏp.ta]
7	luk.li [lʏk.li]	lup.li [lʏp.li]	reg.li [rɛx.li]	rer.li [rɛr.li]
7	luk.wa [lʏk.ʋa]	lup.wa [lʏp.ʋa]	wes.wa [ʋɛs.ʋa]	wem.wa [ʋɛm.ʋa]
7	luk.jo [lʏk.jo]	lup.jo [lʏp.jo]	ning.jo [nɪŋ.jo]	nin.jo [nɪn.jo]
8	reg.ta [rɛx.ta]	rer.ta [rɛr.ta]	reg.ta [rɛx.ta]	rer.ta [rɛr.ta]
8	reg.li [rɛx.li]	rer.li [rɛr.li]	luk.li [lʏk.li]	lup.li [lʏp.li]
8	reg.wa [rɛx.ʋa]	rer.wa [rɛr.ʋa]	ning.wa [nɪŋ.ʋa]	nin.wa [nɪn.ʋa]
8	reg.jo [rɛx.jo]	rer.jo [rɛr.jo]	wes.jo [ʋɛs.jo]	wem.jo [ʋɛm.jo]

SUMMARY

The production of fluent speech is a very complex human skill. It involves the highly automated coordination of several articulatory. The research reported in this thesis focuses on the question which speech production elements serve as basic units for articulation. Potential candidates are words, morphemes, phonemes, and *syllables*. The speech production model by Levelt, Roelofs and Meyer (1999) served as the framework for the present thesis. This model assumes that, in addition to other units such as words, morphemes and phonemes that are crucial to the process of speech production, syllables constitute another important unit. Specifically, it is proposed that syllables form the basic units for articulation.

The speech production system consists of several different processing levels, which can be categorized as *conceptualization*, *formulation* and *articulation*. During conceptualization a pre-verbal message is generated. This message, which corresponds to what the speaker wants to express, contains an ordered, hierarchical pattern of semantic information. In two subsequent steps, the single message components are syntactically and phonologically *formulated*. During *syntactic encoding*, the syntactic structure is generated. During *phonological encoding*, the corresponding word-form is encoded. For this, the morphemes of a word are spelled out phoneme by phoneme. At the end of the phonological encoding process, the single phonemes are combined, following universal and language-specific syllabification rules and (possibly) acknowledging stored metrical information, to form optimally pronounceable units, namely *syllables*. These abstract phonological syllables then have to be translated into phonetic syllabic programs before they can be articulated. The transformation of phonological syllables into phonetic syllables is assumed to be mediated by the so-called *mental syllabary*, a store for pre-compiled motor programs of syllabic size. Here, at least the high-frequency syllables of a language are stored and accessed. This assumption is supported by calculations showing that 500 high-frequency syllables suffice to produce approximately 80% of all speech in

Dutch, English or German. For new or (very) low-frequency syllables, articulatory plans are assembled using the segmental and metrical information specified in the phonological syllables. In a final step, the articulatory network transforms these articulatory plans into overt speech.

The research reported in this thesis aimed to find evidence for the functional role of the syllable during speech production. In particular, verifying the existence of a mental syllabary containing syllabic units which are independently retrieved from the stored word-forms, was a central goal.

The experiments in Chapter 2 approached these questions by using a variant of the *implicit priming paradigm*. Prior syllable priming studies suggest that syllabification is a late sub-lexical on-line process. The current paradigm, namely *the odd-man-out variant*, is sensitive to later stages of word-form encoding and was therefore chosen to tap into the late process of on-line syllabification and possibly syllabary access. In two experiments, testing two different syllable structures (CVV, Exp. 1 and CCVV, Exp. 2), participants produced previously learned target words repeatedly within an experimental block. Two types of response sets were compared, namely *constant* and *variable* response sets: Items in constant sets had overlapping initial segments *and* a constant syllable structure (as in **spui**.en, *to drain*; **spui**.de*, drained;* **spui**.er*, person who drains*; **spui**.end*, draining).* Items in variable sets had an overlap of the first segments but did not have a constant syllable structure (e.g., **spoe**.len, *to rinse*; **spoel**.de, *rinsed*; **spoe**.ler, *person who rinses*; **spoe**.lend, *rinsing*). Note that the second item of this latter set shares the same initial segments but has a different initial syllable structure; it is the *odd-man-out* of the set. The homogeneity of segmental overlap in both the constant and the variable sets was a crucial requirement to investigate whether, in addition to segmental information, speakers use information about the syllabic structure in order to prepare their response. Under the assumption that the syllable is a relevant processing unit in speech production, it was predicted that speakers can phonetically prepare their utterance only when they know what the form of the initial syllable will be (as they do in constant sets) but not when the initial syllable varies within the set of words (as they do in the variable sets). The results confirmed the assumption that constant sets allowed for more preparation than the variable sets; the preparation effect was significantly larger for constant than for variable sets.

Thus, the results of Chapter 2 provide strong evidence for the relevance of syllables in word-form encoding. Taking the results of prior syllable priming studies into account, it was concluded that this method taps into the right level of processing where syllables are encoded, namely during the sub-lexical on-line syllabification process. Furthermore, it was proposed that the larger preparation effect for sets that share initial segments and syllables is partly the result of access to the mental syllabary. This proposal had to remain speculative; clear evidence for the existence of a mental syllabary containing stored phonetic syllable representations still had to be revealed. It had to be established that speakers access such a mental store during word-form encoding. Only after demonstrating the existence of such a store can the underlying mechanisms involved in the observed preparation effects be reevaluated. Effects of syllable frequency would provide clear evidence for the existence of a mental syllabary.

The experiments in Chapter 3 were conducted to contrast the production of high- and low-frequency syllables. A newly developed paradigm, a *Symbol-Position Association Learning Task* and a carefully controlled set of syllable materials were used to test for syllable frequency effects. The participants' task was to respond as fast as possible with a previously learned associated target-word when a production cue was presented on the screen. In learning phases, participants were presented a symbol on one of two potential positions (the left or right position on a computer screen) and were simultaneously presented with the to-be-associated word via headphones. In the test phase the same icon was shown on either the right or the left side of the computer screen to prompt speakers to produce the previously associated target word. Significant syllable frequency effects were found by investigating Dutch mono-syllabic pseudo-words (Exp. 2) and disyllabic pseudo-words bearing the frequency-manipulation on the *first* syllable (Exp. 4). No effect of syllable frequency was found for disyllabic Dutch pseudo-words with high- versus low-frequency *second* syllables (Exp. 3). Taken together, the results of Chapter 3 provide clear evidence for the existence of a mental syllabary. The claim that the syllable is an important unit of speech production is again confirmed. These results also increase the plausibility that access to the mental syllabary may in fact have been involved in the preparation effects reported in Chapter 2.

The study reported in Chapter 4 was carried out in order to verify the interpretation that the advanced preparation on the basis of segmental and syllabic structure information includes the retrieval of pre-compiled gestural scores from memory. Furthermore, this study allows for the replication of the syllable frequency effects with another paradigm. A reading variant of the classic implicit priming paradigm was used to contrast the voice onset latencies for frequency-manipulated disyllabic Dutch pseudo-words. Homogeneous versus heterogeneous high- and low-frequency item sets were presented; thus two factors were crossed, *syllable preparation* (homogeneous versus heterogeneous) and *syllable frequency* (high- versus low-frequency). The crucial prediction was that of an interaction between these two factors. The preparation effect for items with low-frequency initial syllables was predicted to be larger than for items with high-frequency first syllables: The difference score between the low-frequency homogeneous and heterogeneous sets should be larger than the difference score between the high-frequency homogeneous and heterogeneous sets. The homogeneous, high- and low-frequency sets, both consist of items with identical first syllables (e.g., **kem**ta, **kem**li, **kem**wa, **kem**jo (high-frequency) versus **kes**ta, **kes**li, **kes**wa, **kes**jo (low-frequency)). The sets only differed in the frequency of the first syllables. Thus, both (homogeneous) sets allowed the complete (phonetic) preparation of the first syllable including access to the stored syllable program. It was assumed that accessing and retrieving low-frequency syllables from the mental syllabary should take longer than the retrieval of high-frequency syllables. Therefore, it was expected that the benefit due to the preparation would be larger for the low-frequency item sets compared to the high-frequency item sets, in each case in relation to their heterogeneous sets. A frequency effect was only expected in the heterogeneous sets, where no preparation was possible. These predictions were confirmed by the data: A significant interaction between the factors syllable preparation and syllable frequency was found. The benefit from the preparation for low-frequency (homogeneous versus heterogeneous) sets was larger compared to high-frequency sets. Furthermore, a significant syllable frequency effect for the heterogeneous sets was found, thereby replicating the syllable frequency effect observed in Chapter 3 with another paradigm. These results suggest that advanced preparation involves access to the mental syllabary when all relevant information, namely segmental and syllabic information, is given.

Taken together, the results of this thesis provide strong evidence for the assumption that the syllable plays a crucial role in speech production and supports the notion that the syllable comes into play at the interface of phonological and phonetic encoding. The existence of the mental syllabary that contains pre-compiled syllabic motor-programs is strongly supported by the results of this thesis. The crucial claims concerning syllabification and the existence of a mental store for syllables proposed by Levelt et al. (1999) are confirmed.

SAMENVATTING

De productie van vloeiende spraak is een complexe menselijke vaardigheid. Daarbij speelt de automatische coördinatie van verschillende articulatorische subsystemen een grote rol. Het onderzoek beschreven in deze dissertatie concentreert zich op de vraag welke elementen in de spraakproductie als basiseenheden voor articulatie dienen. Potentiële kandidaten zijn woorden, morfemen, fonemen en syllaben. Het spraakproductiemodel van Levelt, Roelofs en Meyer (1999) diende als raamwerk voor deze dissertatie. Dit model gaat ervan uit dat naast andere eenheden zoals woorden, morfemen en fonemen, die cruciaal zijn voor het proces van spraakproductie, syllaben een speciale functie vervullen als basiseenheid voor de articulatie.

Het spraakproductiesysteem bestaat uit verschillende verwerkingsniveaus, die kunnen worden ingedeeld in *conceptualisatie, formulering en articulatie.* Tijdens de conceptualisatie wordt een preverbale boodschap gegenereerd. Deze boodschap, die overeenkomt met wat de spreker wil uitdrukken, bestaat uit een geordende hierarchisch patroon van semantische informatie. In de twee volgende stappen worden de afzonderlijke componenten syntactisch en fonologisch *geformuleerd*. Tijdens de *syntactische codering* wordt de syntactische structuur gegenereerd. Tijdens de *fonologische codering* wordt de bijbehorende woordvorm gecodeerd. Hiervoor worden de morfemen van een woord foneem voor foneem uitgespeld. Aan het eind van het fonologische coderingsproces worden de afzonderlijke fonemen in overeenstemming met universele en taalspecifieke syllabificatieregels samengevoegd tot optimaal produceerbare eenheden, namelijk *syllaben*. Deze abstracte fonologische syllaben moeten vervolgens worden vertaald in fonetische syllabische programmas' voordat ze kunnen worden uitgesproken. Er wordt verondersteld dat de omzetting van fonologische syllaben naar fonetische syllaben tot stand komt door middel van de zogenaamde *mentale syllabeverzameling*, een opslagplaats voor kant-en-klare motorprogramma's ter grootte van een syllabe. Hier zijn de hoog-frequente syllaben van een taal opgeslagen. Deze aanname wordt ondersteund door berekeningen die aantonen dat met zo’n 500 hoog-frequente syllaben ongeveer 80% van alle spraak in het Nederlands, Engels of Duits kan worden geproduceerd. Voor nieuwe of (zeer) laag- frequente syllaben worden er articulatorische plannen samengesteld, door gebruik te maken van de segmentele en metrische informatie die in de fonologische syllaben zijn gespecificeerd. Tijdens de laatste stap zet het articulatorische netwerk, een coördinatief motorsysteem, deze articulatorische plannen om in gesproken spraak.

Het onderzoek dat wordt beschreven in deze dissertatie probeerde bewijs te vinden voor de functionele rol van de syllabe tijdens spraakproductie. In het bijzonder was het doelom het bestaan van een mentale syllabeverzameling te verifiëren, waaruit syllabische eenheden kunnen worden opgehaald, onafhankelijk van de opgeslagen woordvormen.

De experimenten beschreven in Hoofdstuk 2 onderzoeken of er evidentie is voor de functionele rol van de syllabe gedurende de productie van spraak, door gebruik te maken van een variant van het impliciete priming paradigma. Eerdere syllabe priming experimenten suggereren dat syllabificatie een laat sublexicaal online proces is. Het paradigma gebruikt in Hoofdstuk 2, de *odd-man-out variant*, is gevoelig is voor de latere fases van woordvormcodering en werd daarom gekozen om de late processen van online syllabificatie en mogelijk de toegang tot de syllabeverzameling te onderzoeken. In twee experimenten, waarin twee verschillende syllabestructuren werden onderzocht (CVV in Experiment 1 and CCVV in Experiment 2), produceerden proefpersonen meerdere malen eerder geleerde doelwoorden binnen een experimenteel blok. Er werden twee soorten respons-verzamelingen met elkaar vergeleken, namelijk constante en variabele. De woorden in de constante verzamelingen hadden overlappende initiële segmenten en een constante syllabestructuur, zoals **spui**.en, **spui**.de, **spui**.er en **spui**.end. De woorden in de variabele verzamelingen hadden overlappende initiële segmenten, maar geen constante syllabe structuur, zoals **spoe**.len, **spoel**.de, **spoe**.ler en **spoe**.lend. Het tweede item van deze verzameling deelt dezelfde initiële segmenten, maar heeft een andere initiële syllabestructuur: dit is de *odd-man-out* in de verzameling. De homogeniteit in de segmentele overlap in zowel de constante als de variabele verzamelingen was een cruciale vereiste om te onderzoeken of sprekers naast segmentele informatie ook gebruik maken van informatie over de syllabestructuur bij de voorbereiding van een respons. Onder de aanname dat de syllabe een relevante verwerkingseenheid in spraakproductie is, was de voorspelling dat sprekers hun uiting fonetisch kunnen voorbereiden als ze weten wat de vorm van de eerste syllabe zal zijn (zoals het geval is in de constante verzamelingen), maar niet als de eerste syllabe varieert binnen de set van woorden (zoals in de variabele verzamelingen). De resultaten bevestigde de aanname dat de constante verzamelingen meer voorbereiding toelieten dan de variabele; het voorbereidingseffect was significant groter voor de constante dan voor de variabele verzamelingen.

De resultaten van Hoofdstuk 2 leveren dus sterk bewijs voor het belang van syllaben in woordvormcodering. Op grond van eerdere syllabe priming studies werd geconcludeerd dat deze methode gevoelig is voor processen die plaats vinden op het juiste verwerkingsniveau, namelijk het niveau waarop syllaben worden gecodeerd tijdens het sublexicale online syllabificatieproces. Daarnaast zou men kunnen concluderen dat het meer effectieve voorbereidingsproces voor verzamelingen met dezelfde initiële segmenten en syllaben deels het resultaat is van een geavanceerde toegang tot de mentale syllabeverzameling. Dit voorstel was echter speculatief; duidelijk bewijs voor het bestaan van een mentale syllabeverzameling waarin fonetische syllaberepresentaties liggen opgeslagen moest nog worden gevonden. Om de mechanismen die ten grondslag liggen aan het waargenomen voorbereidingseffect nader te kunnen onderzoeken, moest eerst vastgesteld worden dat sprekers tijdens de codering van woordvormen daadwerkelijk gebruik maken van een dergelijke mentale syllabeverzameling. Een effect van syllabefrequentie zou duidelijk bewijs leveren voor het bestaan van een mentale syllabeverzameling.

De experimenten in Hoofdstuk 3 werden uitgevoerd om de productie van hoog- en laagfrequente syllaben te vergelijken. Er werd gebruikt gemaakt van een nieuw ontwikkeld paradigma, namelijk een *Symbool Positie Associatie Leertaak*, en van nauwkeurig gecontroleerd syllabemateriaal om een syllabefrequentie-effect te meten. De taak van de proefpersonen was om zo snel mogelijk met een van te voren geleerd doelwoord te reageren zodra het bijbehorende symbool op het scherm verscheen. Tijdens de leerfase kregen proefpersonen een symbool (een kleine luidspreker) aangeboden dat zich op een van de twee mogelijke posities (links of rechts op het computerscherm) bevond. Tegelijkertijd kregen ze via de hoofdtelefoon het woord, dat hieraan gekoppeld moest worden, te horen. Tijdens de testfase verscheen hetzelfde symbool links of rechts op het scherm. De proefpersoon moest vervolgens zo snel mogelijk het doelwoord, dat in de leerfase aan deze positie gekoppeld was, produceren. De resultaten lieten significante syllabefrequentie-effecten zien voor Nederlandse monosyllabische pseudo-woorden (Exp. 2) en voor bisyllabische pseudo-woorden waarbij de frequentie van de *eerste* syllabe was gemanipuleerd (Exp. 4). Er werd geen effect van syllabefrequentie gevonden voor bisyllabische pseudo-woorden met hoog- of laagfrequente *tweede* syllaben (Exp. 3). Samengevat leveren de resultaten van Hoofdstuk 3 duidelijk bewijs voor het bestaan van een mentale syllabeverzameling. De aanname dat de syllabe een belangrijke eenheid vormt voor spraakproductie werd opnieuw bevestigd. De resultaten maken het bovendien aannemelijker dat de voorbereidingseffecten beschreven in Hoofdstuk 2 het resultaat waren van het gebruik van een mentale syllabeverzameling.

Het onderzoek beschreven in Hoofdstuk 4 was uitgevoerd om steun te vinden voor de interpretatie dat de voorbereiding, waarbij gebruik wordt gemaakt van informatie over de segmentele en syllabische structuur, bestaat uit het ophalen van kant-en-klare gebaren voor syllaben uit het geheugen. Daarnaast gaf deze studie de mogelijkheid het eerder gevonden syllabefrequentie-effect te repliceren met gebruik van een ander paradigma. Ditmaal werd er gebruik gemaakt van een leesvariant van het klassieke impliciete priming paradigma om de spraak-onset-latenties voor frequentiegemanipuleerde bisyllabische pseudo-woorden te vergelijken. In dit experiment werden twee factoren gekruist: de syllabevoorbereiding (homogene versus heterogene verzamelingen) en de syllabefrequentie (hoog- versus laagfrequent). Dit resulteerde in vier verschillende item-verzamelingen. De voorspelling was dat het voorbereidingseffect groter was voor items met laagfrequente initiële syllaben dan voor items met hoogfrequente syllaben. Het verschil tussen de laagfrequente homogene en heterogene verzamelingen zou dus groter moeten zijn dan het verschil tussen de hoogfrequente homogene en heterogene verzamelingen. Beide homogene item-verzamelingen, laag- en hoogfrequent, bestonden uit items waarvan de eerste syllaben identiek warer, bijvoorbeeld **kem**ta, **kem**li, **kem**wa, **kem**jo (hoogfrequent) versus **kes**ta, **kes**li, **kes**wa, **kes**jo (laagfrequent). Het enige verschil tussen deze twee verzamelingen was de frequentie van de eerste syllabe. Dus beide (homogene) sets lieten toe dat de productie van de eerste syllabe volledig kon worden voorbereid, inclusief de toegang tot de opgeslagen syllabeprogramma’s. De aanname was dat het ophalen van laagfrequente syllaben uit de mentale syllabeverzameling meer tijd zou kosten dan het ophalen van hoogfrequente syllaben. Daarom was de verwachting dat het voordeel ten gevolge van de voorbereiding (in vergelijking met de heterogene sets) groter zou zijn voor de laagfrequente item-verzamelingen dan voor de hoogfrequente. Alleen in de heterogene verzamelingen werd een frequentie-effect verwacht, aangezien hier geen voorbereiding mogelijk was. De voorspellingen werden door de data bevestigd: er werd een significante interactie gevonden tussen de factor syllabevoorbereiding en de factor syllabefrequentie. Het voordeel als gevolg van het voorbereidingseffect (homogene versus heterogene verzamelingen) was groter voor de laagfrequente dan voor de hoogfrequente verzamelingen. Bovendien werd er een significant frequentie-effect gevonden voor de heterogene verzamelingen, waardoor het eerder gevonden syllabefrequentie-effect (zie Hoofdstuk 3) werd gerepliceerd met een ander paradigma. De resultaten suggereren dat de voorbereiding bestaat uit toegang tot de mentale syllabeverzameling wanneer alle relevante informatie, namelijk de segmentele en syllabische informatie, gegeven is.

Samengevat leveren de resultaten van deze dissertatie sterk bewijs voor de aanname dat de syllabe een cruciale rol speelt in spraakproductie en ondersteunen het idee dat de syllabe een belangrijke schakel vormt tussen de fonologische en fonetische codering. Het bestaan van een mentale syllabeverzameling die kant-en-klare syllabische motorprogramma’s bevat, wordt sterk ondersteund door de resultaten van deze dissertatie. De theoretische voorstellen betreffendesyllabificatie en het bestaan van een mentale opslagplaats voor syllaben, zoals geformuleerd door Levelt et al. (1999) worden bevestigd.

ZUSAMMENFASSUNG

Die Produktion von gesprochener Sprache ist eine der komplexesten menschlichen Fähigkeiten. Dieser hochautomatisierte Vorgang erfordert die Koordination verschiedener artikulatorischer Untersysteme. Die in dieser Dissertation vorgestellte Forschung konzentriert sich auf die Frage, welche Sprachproduktionselemente die zugrundeliegenden Einheiten für die Artikulation bilden. Potentielle Kandidaten sind Wörter, Morpheme, Phoneme und *Silben*. Das Sprachproduktionsmodell von Levelt, Roelofs und Meyer (1999) diente als theoretischer Rahmen für die vorliegende Arbeit. Dieses Modell nimmt an, dass neben den anderen für die Sprachproduktion relevanten Einheiten wie den Wörtern, Morphemen und Phonemen, die *Silbe* ein weiteres wichtiges Element darstellt.

Das Sprachproduktionssystem besteht aus verschiedenen Prozessebenen, die sich in *Konzeptualisierung*, *Formulierung* und *Artikulation* unterteilen lassen. Während der Konzeptualisierung wird eine präverbale Botschaft generiert. Diese Botschaft – also das, was der Sprecher ausdrücken will - enthält in einer geordneten, linearisierten Abfolge die semantische Information. In den nächsten beiden, aufeinander folgenden Produktionsschritten werden die einzelnen Bestandteile dieser Botschaft syntaktisch und phonologisch *formuliert*. Während der *syntaktischen Enkodierung* wird die syntaktische Struktur der Äußerung generiert. Während der *phonologischen Enkodierung* wird die entsprechende Wortform enkodiert. Hierzu werden die Morpheme eines Wortes Phonem für Phonem ausformuliert. Am Ende des phonologischen Enkodierungsprozesses werden die einzelnen Phoneme zusammen mit sprachspezifischen Silbifizierungsregeln und möglicherweise mit gespeicherter metrischer Information kombiniert, um optimal aussprechbare Einheiten - nämlich die Silben - zu bilden. Diese abstrakten phonologischen Silben müssen dann in phonetische Silbenrepräsentationen übersetzt werden, bevor sie artikuliert werden. Es wird angenommen, dass die Umwandlung von phonologischen in phonetische Silben mittels eines sogenannten *mentalen Silbenspeichers* erfolgt, einem Speicher, der vorgefertigte Motor-Programme für Silben enthält. Hier sind die hochfrequenten Silben einer

Sprache gespeichert. Diese Annahme wird durch Analysen gestützt, die zeigen, dass 500 Silben ausreichen, um ungefähr 80% der gesprochenen Sprache im Niederländischen, Deutschen und Englischen abzudecken. Für neue und niedrigfrequente Silben werden artikulatorische Pläne zusammengestellt, die die segmentale und metrische Information benutzen, die in den phonologischen Silben spezifiziert ist. In einem letzten Schritt übersetzt das artikulatorische Netzwerk die artikulatorischen Pläne in gesprochene Sprache.

Ziel der in dieser Dissertation beschriebenen Forschung war es, Evidenz für die funktionale Rolle der Silbe während der Sprachproduktion zu finden. Insbesondere sollte die Existenz des mentalen Silbenspeichers empirisch gestützt werden, der silbische Einheiten enthält, die unabhängig von den gespeicherten Wortformen abgerufen werden.
Die Experimente in Kapitel 2 haben diese Fragestellung mit einer Variante des *Implicit-Priming-Paradigmas* getestet. Frühere Silben-Priming Studien deuten darauf hin, dass die Silbifizierung ein später, sub-lexikaler On-line-Prozess ist. Die hier verwendete sogenannte *Odd-man-out-Variante* ist sensitiv für die späteren Stufen der Wortform-Enkodierung und wurde deshalb ausgewählt, um den späten Prozess der On-line-Silbifizierung und möglicherweise den Zugriff auf den Silbenspeicher zu untersuchen. In zwei Experimenten, in denen zwei unterschiedliche Silbenstrukturen getestet wurden (CVV, Exp. 1 and CCVV, Exp. 2), produzierten die Probanden wiederholt zuvor gelernten Zielwörter eines Antwortsets. Zwei Typen von Antwortsets wurden miteinander verglichen: konstante and variable Antwortsets. Die Items der konstanten Sets wiesen überlappende Anfangssegmente sowie eine konstante Silbenstruktur auf (z. B.: **spui**.en, ablaufen; **spui**.de*, lief ab;* **spui**.er*, jemand, der etwas ablaufen lässt*; **spui**.end*, ablaufend).* Die Items der variablen Sets wiesen ebenfalls überlappende Anfangssegmente auf, enthielten jedoch jeweils ein Item, dessen Silbenstruktur von der der anderen Items abwich (z.B., **spoe**.len, *spülen*; **spoel**.de, *spülte*; **spoe**.ler, *jemand, der spült*; **spoe**.lend, *spülend*). Das Item mit abweichender Silbenstruktur war der Odd-man-out des variablen Sets.

Die Homogenität der segmentalen Überlappung sowohl in dem konstanten wie auch in dem variablen Set war eine wichtige Voraussetzung, um untersuchen zu können, ob Sprecher – zusätzlich zu segmentaler Information - die Information über die Silbenstruktur nutzen, um eine Äußerung vorzubereiten. Unter der Annahme, dass die

Silbe eine relevante Verarbeitungseinheit darstellt, wurde vorhergesagt, dass Sprecher ihre Äußerung phonetisch nur dann vorbereiten können, wenn ihnen die Form der initialen Silben bekannt ist (wie in den konstanten Antwort-Sets). Sie sollten die Äusserung phonetisch nicht vorbereiten können, wenn die initiale Silbe zwischen den Items innerhalb eines Sets variiert (wie in den variablen Sets). Die Ergebnisse bestätigten diese Annahme. Der Vorbereitungseffekt war signifikant größer für die konstanten als für die variablen Sets. Folglich liefern die Resultate von Kapitel 2 starke Evidenz für die Relevanz der Silbe für die Wortformenkodierung. Aufgrund dieser neuen Erkenntnisse und Ergebnissen aus früheren Silben-Priming Studien kann man davon ausgehen, dass die hier verwendete Methode des Implicit-Priming tatsächlich auf der Verarbeitungsebene misst, auf der Silben enkodiert werden, d.h. während der sub-lexikalen On-line-Silbifizierung. Darüber hinaus kann man schließen, dass der größere Vorbereitungseffekt für Sets, in denen die initialen Segmente und die erste Silbe übereinstimmten, teilweise durch den Zugriff auf das mentale Silbenlexikon bedingt war. Diese Annahme musste jedoch zunächst spekulativ bleiben, da klare Evidenz für die Existenz eines mentalen Silbenspeichers, der gespeicherte phonetische Silbenrepräsentationen enthält, noch nicht erbracht war.

Um die Mechanismen, die dem beobachteten Vorbereitungseffekt zugrunde liegen, näher betrachten zu können, musste nachgewiesen werden, dass Sprecher tatsächlich während der Wortformenkodierung auf ein solches Silbenlexikon zugreifen. Silbenfrequenzeffekte würden klare Evidenz für die Existenz eines solchen mentalen Silbenspeichers liefern.

Die Experimente in Kapitel 3 wurden durchgeführt, um die Produktion von hoch- und niedrigfrequenten Silben zu kontrastieren. Ein neu entwickeltes Paradigma, eine Symbol-Positions-Assoziations-Lernaufgabe, und sehr sorgfältig kontrolliertes Silbenmaterial wurden verwendet, um Silbenfrequenzeffekte zu messen. Die Aufgabe der Versuchspersonen bestand darin, so schnell und akkurat wie möglich mit einem zuvor gelernten Zielwort zu antworten, sobald ein Symbol an der entsprechenden Position auf dem Bildschirm erschien. In den Lernphasen wurde den Versuchspersonen ein Symbol (ein kleiner Lautsprecher) präsentiert, das entweder auf der rechten oder linken Bildschirmhälfte des Computerbildschirms positioniert war. Gleichzeitig wurde ihnen das zu assoziierende Kunstwort über Kopfhörer präsentiert. In den Testphasen erschien der kleine Lautsprecher entweder auf der rechten oder der linken Bildschirmseite. Die

Versuchspersonen mussten entsprechend der Position das zuvor assoziierte Zielwort produzierten. Bei der Untersuchung von einsilbigen niederländischen Kunstwörtern (Exp. 2) und bei zweisilbigen niederländischen Kunstwörtern (Exp. 4), bei denen die Frequenz der *ersten* Silbe manipuliert wurde, konnten signifikante Silbenfrequenzeffekte gefunden werden. Es wurden keine Effekte gefunden für zweisilbige niederländische Kunstwörter mit hoch- oder niedrigfrequenten *zweiten* Silben (Exp. 3). Zusammengenommen liefern die Ergebnisse in Kapitel 3 klare Evidenz für die Existenz eines mentalen Silbenspeichers. Die Annahme, dass die Silbe eine wichtige Sprachproduktionseinheit ist, wurde erneut bestätigt. Diese Ergebnisse legen nahe, dass der Zugriff auf das mentale Silbenlexikon tatsächlich Teil des in Kapitel 2 berichteten Vorbereitungseffektes ist.

Die Studie in Kapitel 4 wurde durchgeführt, um die Interpretation zu stützen, dass die Vorbereitung auf Basis der segmentalen und silbischen Information den Zugriff auf vorgefertigte Silbengesten aus dem Gedächtnis beinhaltet. Darüber hinaus sollte diese Studie den zuvor gefundenen Silbenfrequenzeffekt mit einem anderen Paradigma replizieren. Eine „Lese-Variante“ des klassischen Implicit-Priming-Paradigmas wurde benutzt, um die Sprech-Onset-Latenzen von frequenzmanipulierten zweisilbigen niederländischen Kunstwörtern zu kontrastieren. Es wurden homogene hoch- und niedrigfrequente und heterogene hoch- und niedrigfrequente Itemsets präsentiert, d.h. es wurden zwei Faktoren gleichzeitig getestet: Silbenvorbereitung (homogen versus heterogen) und Silbenfrequenz (hoch- versus niedrigfrequent). Die entscheidende Vorhersage war, dass es eine Interaktion zwischen diesen beiden Faktoren geben sollte: der Vorbereitungseffekt für Items mit niedrigfrequenten ersten Silben sollte größer sein als der für Items mit hochfrequenten ersten Silben. Mit anderen Worten, die Differenz zwischen den niedrigfrequenten homogenen und heterogenen Sets sollte größer sein als die Differenz zwischen den hochfrequenten homogenen und heterogenen Sets. Die homogenen Sets, hoch- wie niedrigfrequent, bestanden beide aus Items mit jeweils identischen erste Silben (z.B.: **kem**ta, **kem**li, **kem**wa, **kem**jo (hochfrequent) vs. **kes**ta, **kes**li, **kes**wa, **kes**jo (niedrigfrequent)). Die Sets unterschieden sich lediglich in ihrer Frequenz. Beide (homogene) Sets erlaubten folglich eine komplette (phonetische) Vorbereitung der ersten Silbe einschließlich des Zugriffs auf gespeicherte Silbenprogramme. Es wurde davon ausgegangen, dass der Zugriff auf niedrigfrequente Silben im mentalen Silbenlexikon langsamer erfolgt als der auf hochfrequente Silben. Es wurde nun erwartet, dass der Vorteil durch die Vorbereitung für niedrigfrequente Silben grösser ausfallen sollte als der für hochfrequente Silben in Relation zu den jeweiligen heterogenen Sets, in den keine Vorbereitung möglich war. Ein Frequenzeffekt wurde folglich nur für die heterogenen Sets erwartet. Die Ergebnisse bestätigten die Vorhersagen: Es wurde eine signifikante Interaktion zwischen den Faktoren Silbenvorbereitung und Silbenfrequenz gefunden. Der zeitliche Gewinn durch die Vorbereitung war für die niedrigfrequenten (homogenen versus heterogenen) Sets größer als für die hochfrequenten Sets. Darüber hinaus wurde ein signifikanter Frequenzeffekt für die heterogenen (hoch- versus niedrigfrequent) Sets gefunden, der den Silbenfrequenzeffekt repliziert (der in Kapitel 3 beobachtet wurde) mit einem anderen Paradigma. Diese Ergebnisse legen nahe, dass der Vorbereitungseffekt Zugriff auf das mentale Silbenlexikon beinhaltet, wenn alle relevanten Informationen, d.h. segmentale und silbische Informationen, abrufbar sind.

Alles in allem liefern die Resultate dieser Arbeit starke Evidenz für die Annahme, dass die Silbe eine tragende Rolle in der Sprachproduktion spielt. Darüber hinaus wird auch die Auffassung gestützt, dass die Silbe an der Schnittstelle zwischen phonologischer und phonetischer Enkodierung „ins Spiel kommt“. Die Existenz eines mentalen Silbenspeichers, der vorgefertigte silbische Motorprogramme beinhaltet, wird durch die Ergebnisse dieser Arbeit gestützt. Die kritischen Thesen, die Levelt und Kollegen (1999) bezüglich der Silbifizierung und der Existenz eines mentalen Silbenlexikons aufstellen, konnten bestätigt werden.

CURRICULUM VITAE

Joana Cholin was born in Cagliari, Italy, on July 4, 1972. After graduation from the Maximilian-Kolbe Gymnasium in 1992 in Wegberg, she studied General Linguistics, Neurology and Psychology at the Heinrich-Heine-University (HHU) in Düsseldorf, Germany, where she received her MA in 1999. From 1994 to 1999, she worked as a student assistant in the research project “Representation and Processing of Inflectional Elements” at the Department of General Linguistics at the HHU in Düsseldorf. From 1995 to 1996, she was employed as a clinical linguist at the Neurological Clinic of the HHU and continued working as a speech therapist from 1996 to 1998. From 1999 to 2000, she worked as a university assistant at the Humboldt-University in Berlin, Germany, and as a research associate in the Daimler-Benz-Project “Group Interaction in High Risk Environments”. In 2000, Joana Cholin was awarded a scholarship from the German Max-Planck-Gesellschaft to prepare her Ph.D. thesis at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. In 2003, she received a TALENT-scholarship from the Netherlands Organisation for Scientific Research (NWO) for a postdoctoral year at the Beckman Institute at the University of Illinois in Urbana-Champaign and at the Moss Rehabilitation Research Institute in Philadelphia, USA.

MPI SERIES IN PSYCHOLINGUISTICS

1. The electrophysiology of speaking. Investigations on the time course of semantic, syntactic and phonological processing
   Miranda van Turennout

2. The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsaggital articulography
   Niels O. Schiller

3. Lexical access in the production of ellipsis and pronouns
   Bernadette M. Schmitt

4. The open-/closed-class distinction in spoken-word recognition
   Alette Haveman

5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach
   Kay Behnke

6. Gesture and speech production
   Jan-Peter de Ruiter

7. Comparative intonational phonology: English and German
   Esther Grabe

8. Finiteness in adult and child German
   Ingeborg Lasser

9. Language input for word discovery
   Joost van de Weijer

10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe
    James Essegbey

11. Producing past and plural inflections
    Dirk Janssen

12. Valence and transitivity in Saliba, an Oceanic language of Papua New Guinea
    Anna Margetts

13. From speech to words
    Arie van der Lugt

14. Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language
    Eva Schultze-Berndt

15. Interpreting indefinites: An experimental study of children's language comprehension
    Irene Krämer

16. Language specific listening: The case of phonetic sequences
    Andrea Weber

17. Moving eyes and naming objects
    Femke van der Meulen

18. Analogy in morphology: The selection of linking elements in Dutch compounds
    Andrea Krott

19. Morphology in speech comprehension
    Kerstin Mauth

20. Morphological families in the mental lexicon
    Nivja H. de Jong

21. Fixed expressions and the production of idioms
    Simone A. Sprenger

22. The grammatical coding of postural semantics in Goemai
    Birgit Hellwig

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies
    Fermin Moscoso del Prado Martin

24. Contextual influences on spoken-word processing
    Daniëlle van den Brink

25. Perceptual relevance of prevoicing in Dutch
    Petra M. van Alphen

26. Syllables in speech production: Effects of syllable preparation and syllable frequency
    Joana Cholin