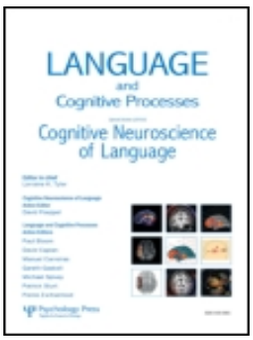
Monitoring metrical stress in polysyllabic words



Routledge
Taylor & Francis Group

Language and Cognitive Processes

ISSN: 0169-0965 (Print) 1464-0732 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp20

Monitoring metrical stress in polysyllabic words

Niels O. Schiller , Bernadette M. Jansma , Judith Peters & Willem J. M. Levelt

To cite this article: Niels O. Schiller , Bernadette M. Jansma , Judith Peters & Willem J. M. Levelt (2006) Monitoring metrical stress in polysyllabic words, Language and Cognitive Processes, 21:1-3, 112-140, DOI: 10.1080/01690960400001861

To link to this article: https://doi.org/10.1080/01690960400001861

LANGUAGE AND COGNITIVE PROCESSES
2006, 21 (1/2/3), 112–140

Psychology Press
Taylor & Francis Group

Monitoring metrical stress in polysyllabic words

Niels O. Schiller
Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands, and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Bernadette M. Jansma and Judith Peters
Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands

Willem J. M. Levelt
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

This study investigated the monitoring of metrical stress information in internally generated speech. In Experiment 1, Dutch participants were asked to judge whether bisyllabic picture names had initial or final stress. Results showed significantly faster decision times for initially stressed targets (e.g., *KAno* ''canoe'') than for targets with final stress (e.g., *kaNON* ''cannon''; capital letters indicate stressed syllables). It was demonstrated that monitoring latencies are not a function of the picture naming or object recognition latencies to the same pictures. Experiments 2 and 3 replicated the outcome of the first experiment with trisyllabic picture names. These results are similar to the findings of Wheeldon and Levelt (1995) in a segment monitoring task. The outcome might be interpreted to demonstrate that phonological encoding in speech production is a rightward incremental process. Alternatively, the data might reflect the sequential nature of a perceptual mechanism used to monitor lexical stress.

Correspondence should be addressed to Niels O. Schiller, Universiteit Maastricht, Faculty of Psychology, Department of Cognitive Neuroscience, P. O. Box 616, 6200 MD Maastricht, The Netherlands. Email: n.schiller@psychology.unimaas.nl

Niels O. Schiller is supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) and by the Dutch Science Foundation (NWO; grant no. 453-02-006). The authors would like to thank Mart Bles and Christine Firk (both Maastricht University), Suzan Kroezen, Janneke van Elferen, and Anne Jacobs (all Radboud University Nijmegen) for their assistance in running the experiments and the members of the Utterance Encoding group of the Max Planck Institute for Psycholinguistics for helpful discussions.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960400001861

INTRODUCTION

Models of speech production (e.g., Caramazza, 1997; Dell, 1986, 1988; Garrett, 1975, 1980; Levelt, 1989, 1992, 2001; Levelt, Roelofs, & Meyer, 1999) assume that the generation of a spoken utterance involves several processes, such as conceptual preparation, lexical access, word form encoding, and articulation. Word form encoding or phonological encoding can be further divided into a number of processes (see recent overview in Meyer, 2000). Levelt et al. (1999) presented one of the most fine-grained models of phonological encoding to date (see also Dell, 1986, 1988). According to this model, phonological encoding can start after the word form (e.g., *banana* /bənæna/) of a lexical item has been accessed in the mental lexicon. First, the phonological encoding system must retrieve the corresponding segments and the metrical frame of a word form. According to Levelt et al. (1999), segmental and metrical retrieval are assumed to run in parallel. During segmental retrieval the ordered set of segments (phonemes) of a word form are retrieved (e.g., /b/, /ə/, /n/, /æ/, /n/, /a/), while during metrical retrieval the metrical frame of a word is retrieved or computed, which consists at least of the number of syllables and the location of the lexical stress (e.g., for *baNAna* this would be a frame consisting of three syllables the second of which is stressed, i.e. /_ ‘_ _/).

During segment-to-frame association previously retrieved segments are combined with their metrical frame. The retrieved ordering of segments prevents them from being scrambled (/b/$_1$, /ə/$_2$, /n/$_3$, /æ/$_4$, /n/$_5$, /a/$_6$). They are inserted incrementally into slots made available by the metrical frame to build a phonological word, i.e. a sequence of one or more well-formed syllables. The phonological or prosodic word forms the domain of phonotactic constraints and syllabification (Booij, 1995). This incremental syllabification process respects universal and language-specific syllabification rules, e.g., *ba.NA.na* (dots mark syllable boundaries).[1] Roelofs (1997, 2000) provided a computational model of this theory including a suspense/resume mechanism making initiation of encoding in the absence of complete information possible. For instance, segment-to-frame association can start before all segments have been selected, then be suspended until the remaining segments become available, and then the process can be resumed. Evidence for the incremental ordering during segmental encoding comes from a number of studies using different experimental

[1] A phonological (or prosodic) word is not necessarily identical to the syntactic (or grammatical) word because some syntactic words such as pronouns or prepositions, which cannot bear stress themselves, cliticize onto other words forming one phonological word together, e.g., *gave* + *it* –> /gɛɪ.vɪt/.

paradigms (e.g., Meyer, 1990, 1991; Van Turennout, Hagoort, & Brown, 1997; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). Segment-to-frame association is the process that lends the necessary flexibility to the system depending on the speech context (Levelt et al., 1999). After the segments have been associated with the metrical frame, the resulting phonological syllables may be used to activate the corresponding phonetic syllables in a mental syllabary (Cholin, Levelt, & Schiller, in press; Cholin, Schiller, & Levelt, 2004; Crompton, 1981; Levelt, 1989, 1992; Levelt & Wheeldon, 1994; Schiller, Meyer, Baayen, & Levelt, 1996; Schiller, Meyer, & Levelt, 1997). Once the syllabic gestural scores are made available, they can be translated into neuro-motor programs, which are used to control the movements of the articulators, and then be executed resulting in overt speech (Goldstein & Fowler, 2003; Guenther, 2003; Schiller, van Lieshout, Meyer, & Levelt, 1999). In this study, we will focus on metrical encoding, i.e., the processes involved in producing the correct lexical stress of words.

As stated above, a number of studies showed that phonological words are encoded incrementally. Meyer (1990, 1991) used a preparation paradigm to show that participants are faster in naming a word if they can prepare segmental material of the target. For instance, participants are faster to name *banana* if they know beforehand that the target started with a *b* (/b/). They are even faster if they know that the target started with *ba* (/bə/), etc. That is, the preparation effect increases with the size of the known word initial stretch. However, no preparation effect is obtained when participants can prepare segmental material from the final part (e.g., *na* /na/) of the word (Meyer, 1990, 1991). This was taken as evidence that segmental encoding proceeds in an incremental fashion from beginning to end of words during phonological encoding.

More on-line data about the time course of segmental encoding during speech production comes from a study by Van Turennout et al. (1997). They used lateralised readiness potentials, i.e., a derivative of the human electroencephalogram to demonstrate that semantic information about a word is available to the speech production system at an earlier point in time than phonological information. However, they also showed that the first segment of a word is encoded approximately 80 ms earlier than the last segment. The words in their study were on average 1.5 syllables long. Van Turennout et al.'s result demonstrates not only the temporal ordering of segments during phonological encoding but it also gives an indication of the speed of this process, i.e., 50–55 ms from syllable onset to syllable offset.

Additional evidence for the incremental nature of phonological encoding comes from a study by Wheeldon and Levelt (1995). They asked participants to monitor for pre-specified segments when generating

the Dutch translation of an English word. This task can be seen as a production equivalent of the phoneme-monitoring task employed in speech comprehension research (for an overview see Connine & Titone, 1996). Wheeldon and Levelt found that participants were faster in monitoring for the first consonant in a $C_1VC_2.C_3VC_4$ word (where C stands for consonant and V for vowel), such as *lifter* (''hitchhiker''), than for the second consonant (e.g., $C_1 \ll C_2$). Furthermore, they were faster in monitoring for C_2 than for C_3 ($C_2 \ll C_3$) and C_3 was faster than C_4 ($C_3 < C_4$), although this last difference did not reach significance. Wheeldon and Levelt (1995) took their results to confirm the incremental encoding of segments during phonological encoding in speech production. They argued that their monitoring effect occurred at the phonological word level, i.e., when a fully syllabified phonological representation of a word was generated. Interestingly, the monitoring difference between C_1 and C_2 (55 ms) corresponds nicely to the data found by Van Turennout et al. (1997) with another monitoring task (50 to 55 ms; see above). Recently, Wheeldon and Morgan (2002) replicated this result for English using a slightly different methodology (see also Morgan & Wheeldon, 2003) and Schiller (in press) replicated and extended the results for Dutch. Importantly for this study, if Wheeldon and Levelt (1995) were correct in assuming that the phonological word level is being monitored in such a task, speakers should also be able to monitor metrical stress in self-generated words. Furthermore, if a comparable incremental pattern is obtained for monitoring metrical stress as for monitoring segments, such a pattern may give us information about the time course of metrical encoding.

Before we will describe in more detail the processes involved in self-monitoring, we will briefly turn to the metrical stress system in Dutch and summarise the psycholinguistic evidence that is available at the moment.

METRICAL STRESS IN DUTCH

Although the intricacies of the Dutch metrical stress system are still under debate (for an overview see Kager, 1989), we will provide a brief summary here. In the theory of Trommelen and Zonneveld (1989, 1990) and Zonneveld, Trommelen, Jessen, Bruce, and Árnason (1999) bisyllabic words receive stress on the initial syllable, except when the final syllable is a so-called super-heavy syllable, i.e., a syllable with a rhyme of the type VVC or VCC (where V stands for a vowel, VV for a long vowel or a diphthong, and C for a consonant). In that case, stress falls on the super-heavy final syllable. According to this account, only words carrying stress on a final syllable that is not super-heavy are exceptional (e.g., *fo.REL* ''trout'' in Dutch). The stress patterns of those words are assumed to be

stored in the lexicon, whereas the remaining stress patterns could be generated by rules.

The psycholinguistic account of metrical stress representation put forward in Levelt's theory is less complicated (see Roelofs & Meyer, 1998). Levelt et al.'s (1999) position is that the metrical structure of regular words is derived by a simple default rule (i.e., ''stress the first syllable containing a full vowel''). A full vowel is any vowel except for schwa, which can never be stressed in Dutch (as in English or German; Kager, 1989). Only for irregular words (less than 10% of the word tokens) the metrical frame must be stored in the lexicon. Note that some words that are regular according to linguistic accounts, are irregular according to Levelt et al.'s (1999) position (e.g., *ci.TROEN* /sitrun/ ''lemon'', which has a super-heavy final syllable).

Few psycholinguistic studies have investigated the representation of metrical stress. Nickels and Howard (1999) found that lexical stress location affected word production in a group of seven English aphasic patients. All seven patients were significantly worse at repeating bisyllabic words with primary stress on the second syllable relative to words with primary stress on the first syllable. According to Howard and Smith (2002), errors of metrical stress result from a difficulty in phonological assembly: phonological errors are more likely to occur when the number of segments in a phrase increases and when the metrical stress cannot be assigned by default but has to be assembled instead. Cappa, Nespor, Ielasi, and Miozzo (1997) described an Italian aphasic patient who produced more errors on irregular than on regular words. Assuming that this patient had an impairment of lexical stress representations of irregular words, this result would support Levelt et al.'s theory (see also Laganaro, Vacheresse, & Frauenfelder, 2002; Miceli & Caramazza, 1993). However, Schiller, Fikkert, and Levelt (2004) did not obtain a metrical priming effect in Dutch, not even for irregular words, i.e., words that should be stored in the lexicon according to Levelt et al. (1999).

In summary, the evidence about whether or not metrical stress is stored in the lexicon is inconclusive at the moment. A distinction between regular words, for which stress can be derived by rule, and irregular words, for which stress has to be stored in the lexicon, has proven descriptively valuable. However, it is not entirely clear which words should be considered as irregular (see also Howard & Smith, 2002). Possibly, metrical stress is computed for the majority of the words as long as their stress pattern can be derived by some linguistic rule. This might also include words that are irregular according to psycholinguistic definitions, but regular in terms of certain linguistic theories (Trommelen & Zonneveld, 1989, 1990; see also Schiller et al., 2004). In this study, we will not be concerned with whether metrical stress is stored or computed.

The internal self-monitoring task used in the experiments reported below is assumed to have access to the phonological word level, i.e., a fully prosodified representation (see below).

SELF-MONITORING DURING SPEECH PRODUCTION

In the experiments described in this study, we required our participants to monitor for lexical stress in certain target words. However, how does verbal self-monitoring proceed? When we are engaged in speaking, we constantly monitor the coordination of processes such as the selection of meanings, retrieval of words, syntactic and phonological encoding, and articulation. When we produce a speech error, we can interrupt ourselves and self-correct the error because we are able to listen to our own speech via auditory-sensory feedback while we speak. This is called *external monitoring*. However, we can even self-correct an error before the unintended word has been completely uttered. For instance, in a task involving the description of visual patterns, Levelt (1983) found self-repairs such as ''[...] is a *v* – a horizontal line'' (Levelt 1983, p. 64). In this example, too little of the word *vertical* was pronounced to make recognition via the external monitoring system possible. In order to interrupt oneself after the articulation of only the first segment of an intended word, the error must have been detected before the onset of articulation, suggesting the existence of *internal monitoring*. Maybe the most impressive evidence for an internal monitor is that when speech errors are induced in the laboratory, errors resulting in taboo words (e.g., *tool kits* becoming *cool tits*) occur significantly less often than other errors. However, elevated Galvanic skin responses recorded simultaneously suggest that participants actually generate the taboo word errors internally but detect them before they are overtly uttered, supporting the existence of a pre-articulatory self-monitor system for speaking (Motley, Camden, & Baars, 1982).

In Levelt's perceptual loop theory of self-monitoring (Levelt, 1989), the external monitor is used when we self-perceive our own acoustic speech signals. Presumably, listening to our own overt speech or to speech generated by somebody else is processed through the same perceptual system, as shown, for instance, by recent neuroimaging studies (e.g., Price, Wise, Warburton, Moore, Howard, Patterson, Frackowiak, & Friston, 1996). Levelt (1989; Levelt et al., 1999) assumes that an internal monitor also proceeds through the general comprehension system (but see also Postma, 2000). A central perception-based monitor would be economical since two different types of monitoring (internal and external) could be processed by the capabilities of one single perceptual system. Originally,

the internal monitoring system in Levelt's theory could only access the phonetic plan, i.e., the output of the speech planning process immediately prior to articulation (Levelt, 1989). However, Wheeldon and Levelt (1995) found no correlation between the monitoring latencies and the acoustic intervals between the target segments in the carrier words used in their experiment (see also Wheeldon & Morgan, 2002). Furthermore, the generation of internal speech was found to occur at a significantly faster rate than overt articulation. Therefore, Levelt et al. (1999) suggested in the most recent version of their theory that the internal monitoring system has access to a more abstract code of the planning process, i.e., the phonological planning level. At this level, the speech planning system provides a fully prosodified, syllabified word form.

This leads to the theoretical motivation of the present study. If internal monitoring has indeed access to the phonological word level, it should not only be possible to monitor for segments but also for metrical stress. Furthermore, if metrical stress is monitored from beginning to end of a phonological word, this should be reflected in the reaction times for initial vs. final stress. If the monitor has access to earlier levels of representation or processing stages, however, the whole metrical pattern might be available to the monitor at once, and consequently no difference in reaction times should be visible for initial vs. final stress.

THE EXPERIMENTS

We employ the methodology of implicit picture naming to investigate self-monitoring of internal speech (Van Turennout et al., 1997, 1998). In picture naming, presumably all stages of the speech production process have to be completed, e.g., conceptualisation, lexical access, word form encoding, and articulation (see Glaser, 1992 for a review). In the present study, native speakers of Dutch were presented with pictures that all had polysyllabic names. Participants were required to generate internally the corresponding phonological word form for each picture and press a button when the word fulfilled a certain phonological criterion and withhold the button press when the word did not fulfill the criterion. By using tacit naming plus a minimal push-button response, we were able to investigate phonological and/or phonetic encoding in a direct way. The correctness of push-button responses suggested that participants came up with the correct and intended names of the pictures.

EXPERIMENT 1: MONITORING FOR METRICAL STRESS IN BISYLLABIC TARGETS

In Experiment 1, we asked participants to silently generate the names of pictures one at a time and press a button when the corresponding picture

name had initial/final stress. If metrical stress is monitored from beginning to end, just like segments (Wheeldon & Levelt, 1995), decision latencies should be faster for picture names with initial stress than for picture names with final stress. In combination with the monitoring experiment, we carried out a couple of control studies.

Method

Participants. Thirty-one participants (all undergraduate students from the University of Nijmegen) took part in exchange for pay. They all had normal or corrected-to-normal vision. All participants were right-handed and native speakers of Dutch.

Materials. The materials consisted of 64 bisyllabic, monomorphemic Dutch nouns. Line drawings of the corresponding objects were either taken from the picture database of the Max Planck Institute for Psycholinguistics or drawn by a professional artist. Items could be divided into four groups of equal size depending on the consonant-vowel structure of their first syllable (CV vs. CVC) and the location of their lexical stress (initial vs. final). All items were between four and seven segments (phonemes) long and the different item categories had mean frequencies of occurrence between 17 and 25 per million as determined by CELEX (see Baayen, Piepenbrock, & Gulikers, 1995), i.e., all item categories were of moderate frequency. Picture-name agreement was also matched across item categories (for details see Table 1). A complete list of all items can be found in Appendix A.

Design. The experiment started with a familiarisation and a practice block including the entire set of pictures. Then one naming block was

TABLE 1
Lexico-statistical characteristics of the target words in Experiment 1

Stress location	*CV structure of the first syllable*	*Example*	*Mean CELEX frequency (per one million words)*	*Mean picture-name agreement (on a 1–7 scale)*	*Mean length in segments*
Initial	CV	*boter*	24.7	6.21	5.0
Initial	CVC	*banjo*	23.1	5.78	6.2
Final	CV	*banaan*	17.2	6.06	5.1
Final	CVC	*balkon*	19.8	5.89	6.2

Note: The picture-name agreement is based on a sample of $n = 20$ native Dutch participants who rated pictures and their corresponding labels on a scale from 1 (low agreement) to 7 (high agreement).

presented, followed by two monitoring blocks with reversed instructions, and two object decision blocks. After each block there was a short break. For the naming block, all 64 pictures were presented in a single block, which was randomised individually for each participant. For the monitoring, half of the participants started with a block in which they had to actively respond to picture names with stress on the first syllable and withhold responses for names with final stress. Then they received a second block with the same material in which the response contingencies were reversed. The other half of the subjects was presented with the reversed block order. The order of trials was randomised for each block and each participant individually. For the object/non-object decision, each block contained four pictures of existing objects from each of the four experimental categories indicated in Table 1 (resulting in 16 initial and 16 final stress picture names) plus the 32 pictures of nonsense objects (see below). The same nonsense objects were presented in both blocks. The order of trials was randomised individually for each block and participant.

Procedure. Participants were tested individually. They were seated behind a computer screen and asked to place their right index finger on the right button of a button-box that was placed in front of them. The experiment started with a picture naming part in which participants were asked to name all 64 pictures. Pictures were of approximately equal size. They all fitted into a 7×7 cm square. Pictures appeared one at a time on a computer screen and the participants' task was to name them as fast and as accurately as possible. Each trial started with a fixation point that was visible for 500 ms in the centre of the screen, followed by a blank screen for 300 ms. Then the picture appeared in the centre of the screen and remained in view until a verbal response was given. At picture onset, a clock was started. Verbal responses were registered with a microphone in front of participants. The microphone was connected to a voice key, which stopped the clock when it was triggered. After 1000 ms the next trial started. The Nijmegen Experimental Set-Up (NESU) controlled the presentation of the trials. Before the picture naming trials started, participants were familiarised with the pictures. Each picture was shown individually with the picture name underneath until the participant pressed the button and the next picture appeared. After picture familiarisation, each picture was shown again to the participants who were asked to name the pictures aloud as fast and as accurately as possible. The practice block served the purpose of demonstrating whether or not participants knew the name for each picture.

The picture-naming task was followed by a self-monitoring task using the same pictures. In the monitoring part, the same participants were asked to suppress overt naming of the pictures and to press the button as fast and

as accurately as possible in case the picture name had initial stress (e.g., *KAno* ''canoe''). When the picture name had final stress, they were required to withhold the button press (e.g., *kaNON* ''cannon''). All 64 pictures were shown one at a time. In a second block, instructions were switched (i.e., press the button for final stress, but withhold button-press for initial stress) and the same pictures were shown again in order to get a button-press response for every item. An experimental trial consisted of the following events: First, a fixation-cross appeared for 500 ms in the centre of the screen, which participants were asked to fixate. Then, after 300 ms, a picture appeared around the same location on the screen. As soon as possible after picture onset, participants had to give their response. Reaction times (RTs) were registered automatically. The picture disappeared from the screen when participants responded or after 2000 ms. The following trial began after an inter-trial interval of 1000 ms. The trial sequencing was also controlled by NESU.

Finally, there was an object/non-object decision part. In this last part of the experiment, participants were required to make an identification judgement about each target picture to control for potential visual differences between the pictures that selectively affect either initial or final stress targets. For this part, 32 pictures of nonsense objects (i.e., objects without a meaning; taken from Kroll & Potter, 1984) were used. They were selected from a larger set of nonsense objects that was pre-tested before. Pictures of existing objects (e.g., persons, animals, natural and artificial objects) used in the picture naming and self-monitoring part were presented on the computer screen intermixed with non-existing objects (pseudo-objects). Participants were required to press with their right hand side as fast and as accurately as possible the *YES* button on a button box if they thought the picture was denoting an existing object and the *NO* button with their left hand side otherwise. An experimental trial consisted of the following events: First, a fixation-cross appeared for 500 ms in the middle of the screen. Then, after the screen was blank for 300 ms, a picture appeared in the same location and remained on the screen until a response was given. After another 1000 ms the next trial started. Button press responses were registered automatically. NESU controlled the presentation of trials. Participants visually inspected all the pictures of nonsense objects before the experiment started.

Results

One participant was excluded from the naming part due to voice-key failure. Errors (wrong responses, voice-key failures, etc.) and time-outs were discarded from the RT analysis (4.1%). Furthermore, we only took into account RTs between 300 ms and 1500 ms. The mean naming latencies

for picture names with initial stress was 823 ms ($SD = 56$) while it was 787 ms ($SD = 69$) for picture names with final stress. This 36 ms advantage for picture names with final stress over picture names with initial stress was significant by participants but not by items, $t_1(29) = 5.33$, $p < .01$; $t_2(62) = 1.74$, n.s. Error rates showed a similar pattern. More errors were made on pictures with initial stress (4.9%) than on pictures with final stress (3.2%). This effect was not significant, however, $t_1(29) = 1.89$, $p = .07$; $t_2(62) = 1.32$, n.s.

As far as the monitoring part is concerned, wrong button presses and time-outs were counted as errors (19.4%) and discarded from the RT analysis. Furthermore, for the RTs only latencies above 350 ms and below 1500 ms were taken into account. The mean RTs were 937 ms ($SD = 150$) for picture names with initial stress and 1007 ms ($SD = 130$) for picture names stressed on the second syllable. One-tailed *t*-tests revealed that the 70 ms advantage of the initial stress condition over the final stress condition was significant, $t_1(30) = 3.66$, $p < .01$; $t_2(62) = 3.47$, $p < .01$. A similar result was obtained from the error analysis. Participants made more errors in the final stress condition (15.5%) than in the initial stress condition (12.1%), showing that there was no speed-accuracy trade-off at hand. However, the one-tailed analysis based on arc-sin transformed error proportions did not reveal a significant difference, $t_1(30) = 1.86$, $p = .07$; $t_2(62) = 1.39$, n.s. Nevertheless, both in the RTs and in the error patterns a clear advantage for initial over final stress became apparent.

For the object decision part, only YES-responses were taken into account. Wrong button presses were counted as errors (2.5%) and discarded from the RT analysis. Furthermore, for the RTs only latencies above 200 ms and below 1000 ms were taken into account. The mean RTs for the two stress conditions (initial vs. final stress) were 464 ms ($SD = 81$) for picture names with initial stress and 458 ms ($SD = 81$) for picture names stressed on the second syllable. That is, pictures with final stress names were recognised slightly faster than pictures with initial stress names. *T*-tests revealed no difference between the final stress and the initial stress condition, $t_1(30) = 1.35$, n.s.; $t_2(62) = 1.07$, n.s. A similar result was obtained from the error analysis: There were slightly fewer errors in the final stress condition (2.1%) than in the initial stress condition (2.8%). This 0.7% difference did not reach significance, however, $t_1(30) < 1$; $t_2(62) < 1$.

Discussion

Native speakers of Dutch are faster and more accurate in deciding that a bisyllabic word has initial stress than in deciding that it has final stress. This result supports the prediction made on the basis of the outcome for

segmental monitoring (Wheeldon & Levelt, 1995). Wheeldon and Levelt (1995) interpreted their data as reflecting the time course of phonological encoding in speech production. Our data might also be interpreted as reflecting genuine production processes. When the phonological word is built, i.e., during segment-to-frame association, a monitoring device monitors for lexical stress, and as soon as a stressed syllable is found, a button-press response is initiated. However, there is also an alternative account for the data of Experiment 1. It might be that the phonological word is built and only after phonological encoding has been completed, is the result monitored for lexical stress. If such monitoring proceeds from the beginning towards the end of a phonological word, the same pattern of monitoring latencies would be expected, and the outcome does not necessarily have anything to do with production processes. We will come back to this discussion at a later point.

The differences in monitoring latencies found in Experiment 1 between the two conditions might potentially be criticised for several reasons. First, initial stress might be monitored faster and more accurately than final stress because the Dutch language has a strong preference for initial stress. A lexico-statistical analysis of the Dutch lexicon revealed that 75.1% of the monomorphemic bisyllabic nouns have stress on the initial syllable, while 24.9% have it on the final syllable (type count). If one takes frequency of occurrence into account (token count), the distribution changes only slightly: 66.5% of the bisyllabic noun tokens have initial stress, 33.5% have final stress. That is, the vast majority of the bisyllabic nouns have initial stress in Dutch, but among the final stress nouns some are of relatively high frequency.[2] If one assumes that initial stress is retrieved or computed faster than final stress because it occurs much more often than final stress, an incremental monitoring effect might be due to frequency of occurrence of the corresponding stress patterns.

However, if this were the case, one would also expect—ceteris paribus—that initial stress pictures were named faster than final stress pictures because on average the metrical frame would be available earlier for the former category than for the latter and consequently processing could

[2] The picture is even more extreme if all bisyllabic noun items are included in the analysis, i.e., also compounds and derivations. In that case, there are 85.8% words with initial stress and 14.2% with final stress (type count). Compounds usually have initial stress in Dutch (e.g., *DAK.pan* ''roof tile'') and suffixes are usually unstressed (e.g., *WAAR.heid* ''truth'', consisting of the adjective morpheme *waar* ''true'' and the nominal suffix *-heid*) such that derived nouns generally also have stress on the first syllable. Again, taking frequency into account, a token count revealed that 66.8% of the bisyllabic nouns in Dutch have initial stress, while 33.2% have final stress showing that some final stress words have a relatively high frequency of occurrence.

proceed faster. As mentioned in the introduction, picture naming involves several steps. Assuming that image recognition and lexical access occur on average at approximately the same speed for both categories of picture names, word form encoding might be faster for words with regular stress than for words with irregular stress because regular stress might be assembled (i.e., retrieved or derived by rule) faster than irregular stress. Even if segmental retrieval occurs equally fast for words with regular and irregular stress, the metrical pattern might be available faster for words with regular stress than for words with irregular stress. If this were the case, associating the segments with the metrical frames might proceed faster for words with regular stress than for words with irregular stress, and consequently the former might be named faster than the latter. To test this prediction, the picture-naming task was included in the experiment. The (non-significant) naming advantage of final over initial stress pictures showed that monitoring latencies and picture naming latencies were not confounded. This demonstrates that the monitoring of stress in tacit naming at the level of phonological word planning is independent of more peripheral processes during phonological encoding such as neuromuscular preparation or articulatory execution.

Unfortunately, it was not possible to control for factors such as articulatory difficulty (e.g., initial phoneme) and voice-key sensitivity (e.g., factors causing the voice-key to trigger). These factors are known to potentially affect naming latencies (e.g., Kawamoto, Kello, Jones, & Bame, 1998; Kessler, Treiman, & Mullennix, 2002; Pechmann, Reetz, & Zerbst, 1989; Rastle & Davies, 2002). Therefore, we carried out an additional experiment including a delayed naming task to control for potential differences of these two factors. Fourteen new participants, all native speakers of Dutch, were familiarised with the same pictures as used in the picture-naming task and practiced their names once. After that, the pictures were presented again one at a time for 1000 ms followed by a clear screen for an interval between 1000 and 1800 ms. Then a visual cue (+) was presented on the screen and the voice-key was activated. The visual cue was the sign for participants to respond as fast as possible. The voice-key remained activated for 2000 ms and after another 1500 ms the next trial started. Naming latencies did not reveal any differences between picture names with first syllable stress (408 ms) and picture names with second syllable stress, 413 ms; $t_1(13) = 1.30$, n.s.; $t_2(62) = 1.11$, n.s., rejecting the possibility that articulatory or voice-key related factors might have distorted the naming latencies.

Second, potential monitoring effects might possibly be due to a visual input effect, i.e., the fact that different pictures were used to monitor initial and final stress. Suppose, for instance, that the majority of the pictures denoting final-stress words (e.g., *kaNON*) was for some reason harder to

recognise than the set of pictures denoting words with initial stress (e.g., *KAno*). If this were the case, this could cause longer RTs in the former than in the latter condition. To exclude such an explanation, the object/non-object decision experiment was carried out. Neither RTs nor error patterns support the hypothesis that final stress pictures were harder to recognise than pictures with initial stress.[3] Therefore, a visual ''input'' effect can be excluded as explanation of the monitoring effects. More likely, the temporal availability of the crucial phonological information was responsible for the differences. Following Wheeldon and Levelt (1995) as well as Wheeldon and Morgan (2002), this result might be interpreted as evidence for the incremental nature of metrical encoding in speech production. So far, we only knew that segments were planned sequentially, and it was still an open question when metrical information is encoded. Experiment 1 showed that the direction of metrical planning—or the monitoring of metrical information—is also rightward incremental. Alternatively, a phonological word might be monitored for lexical stress after it has been encoded. The present data cannot clearly distinguish between these two possibilities. However, we will come back to this issue in the General Discussion.

Although the naming latencies showed that picture names with initial stress were not named faster than picture names with final stress, one might still argue that the monitoring results are at least partly due to the fact the initial stress constitutes the default stress pattern in Dutch. In order to exclude such a potential confound in monitoring latencies from higher frequency of occurrence of initial over final stress words, in the next experiment none of the targets had default, i.e., initial stress. This was achieved with trisyllabic words. Trisyllabic picture names allow having participants monitor for metrical stress on the second or third syllable. The

[3] One may argue that extended practice with the experimental stimuli in the naming and self-monitoring parts might be responsible to diminish any differences in identification times for the existing objects across conditions, whereas participants only had very limited experience with the nonsense objects. Therefore, participants might have pressed the YES-button to all pictures they were familiar with and the NO-button to everything else, whether real object or not (i.e., there was no need to identify the objects). To show that this was not the case we refer to an object/non-object identification experiment that was done as a control experiment in another study (Jansma & Schiller, 2004). In that study, participants were once exposed to a set of existing objects and non-objects. (The pictures of objects and non-objects used in the current experiment formed a subset of the materials used in the Jansma and Schiller study.) Participants were then required to make the object/non-object decision, and it turned out that even under circumstances in which participants did not have prior practice with the pictures there was no difference in RTs between pictures corresponding to picture names with first syllable stress and those with second syllable stress (see Jansma & Schiller, 2004, for details).

purpose of Experiments 2 and 3 is to tackle the potential problem of initial (default) vs. non-initial stress (non-default stress).

EXPERIMENT 2: MONITORING NON-DEFAULT METRICAL STRESS IN TRISYLLABIC TARGETS

Let us start with the lexico-statistical facts about trisyllabic words in the Dutch lexicon (again based on CELEX; see Baayen et al., 1995). Our analysis revealed that 24.1% of the monomorphemic nouns have stress on the initial syllable, while 45.0% have it on the pre-final syllable, and 30.7% have final stress (type count). Thus, there is no clear preference for a particular position. If one takes frequency of occurrence into account (token count), only 11.7% of the trisyllabic noun tokens have initial stress, but 53.0% have pre-final stress, and 35.3% have final stress. That is, the majority of the trisyllabic noun tokens has pre-final stress in Dutch and some of them have a relatively high frequency of occurrence compared to the trisyllabic nouns with initial stress. The picture changes quite a bit when all trisyllabic noun items are included in the analysis, i.e., also compounds and derivations. In that case, there are 70.4% words with initial stress, 18.9% with pre-final stress, and 10.7% with final stress (type count).[4]

This lexico-statistical analysis showed that for trisyllabic nouns the situation of the dominant stress pattern is less clear than for the bisyllabic nouns. If all trisyllabic nouns are taken into account, there still is a bias towards initial stress (see previous paragraph and footnote 4). However, if only monomorphemic nouns are considered, pre-final stress is occurring most frequently. Important at this point is, however, that according to the psycholinguistic theory by Levelt et al. (1999), monomorphemic words with both pre-final and final stress are irregular (non-default) in Dutch (unless they start with a schwa syllable, e.g., *jeNEver* /jənevər/ ''[type of] liquor''). Therefore, any argument centred on the distinction between default and non-default stress in Dutch would not apply to differences between these two stress positions.

[4] Again, trisyllabic compounds usually have initial stress in Dutch because stress falls on the first part of the compound (e.g., WOON.ka.mer ''living room'') and suffixes are usually also unstressed (e.g., HE.mel.rijk ''kingdom of heaven'', consisting of the noun morpheme hemel ''heaven'' and the nominal suffix -rijk). Again, taking frequency into account, a token count revealed that 42.4% of the trisyllabic nouns in Dutch have initial stress, while 39.3% have pre-final stress, and 18.3% have final stress, showing that many initial stress compounds have a relatively low frequency of occurrence as compared with words with pre-final or final stress.

Method

Participants. Twenty-eight native Dutch participants from the same pool as for the previous experiments took part in Experiment 2.

Materials. Twenty-eight trisyllabic, monomorphemic picture names were selected for pictures, which were available in the picture database of the Max Planck Institute for Psycholinguistics. Half of the picture names had pre-final stress; the other half had final stress. The complete list of materials can be found in Appendix B. All items were between five and nine segments (phonemes) long and the item categories had a mean frequency of occurrence between 7 and 9 per million as determined by CELEX (see Baayen et al., 1995), i.e., all items were of low to moderate frequency. Picture names with pre-final and final stress were also matched for delayed naming latencies (for details see Table 2). For the object/non-object decision we also selected 28 pictures of pseudo-objects described earlier.

Procedure and design. The procedure and design for Experiment 2 were identical to Experiment 1 with the following exception: Instead of pressing the button for first or second syllable stress in the monitoring part, participants were asked to press the button when a picture name had pre-final stress and withhold the button-press response in cases when the picture name had final stress. In a second monitoring block, the instructions were switched. Half of the participants actively responded to pre-final stress first and final stress afterwards, the other half received the reverse order of blocks.

Results

The picture naming data of one participant was lost due to technical problems. Picture names with pre-final stress (808 ms; $SD = 79$) were named slightly more slowly than picture names with final stress (799 ms;

TABLE 2
Lexico-statistical characteristics of the target words in Experiment 2

Stress location	*Example*	*Mean CELEX frequency (per one million words)*	*Mean delayed naming latencies (in ms)*	*Mean length in segments*
Pre-final	*asperge*	9.2	347	7.4
Final	*artisjok*	7.5	349	7.5

Note: The mean delayed naming latencies are based on a study similar to the delayed naming study described in the Discussion of Experiment 1 ($n = 14$ native Dutch participants).

$SD = 80$). This 9 ms difference was not significant, however, $t_1(26) < 1$; $t_2(26) < 1$. The errors revealed a similar picture. Altogether, there were 3.8% errors. There were slightly more errors in the pre-final stress condition (4.0%) than in the final stress condition (3.7%). Again, this difference was not significant, $t_1(26) < 1$; $t_2(26) < 1$.

Monitoring latencies shorter than 300 ms and longer than 2000 ms were excluded from the analyses. Also, errors (4.3%) were not included in the RT analysis. Monitoring latencies for pre-final (second syllable) stress (1036 ms; $SD = 172$) were faster than for final (third syllable) stress (1097 ms; $SD = 177$). This 61 ms difference was marginally significant (one-tailed), $t_1(27) = 3.07$, $p < .01$; $t_2(26) = 1.70$, $p = .05$). Error rates pointed in the same direction. There were more errors on picture names with final stress (5.4%) than on picture names with pre-final stress (3.3%). However, this difference in error rates was not significant, $t_1(27) < 1$; $t_2(26) < 1$.

Pictures with pre-final stress names were recognised slightly more slowly (483 ms; $SD = 75$) than pictures with final stress names (473 ms; $SD = 72$). However, this 10 ms difference was not significant, $t_1(27) = 1.93$, n.s.; $t_2(26) < 1$. Error rates showed a similar pattern. Overall, there were 5.2% errors. There were slightly more errors on pre-final stress picture names (6.1%) than on final stress picture names (4.3%). This difference, however, was not significant, $t_1(27) < 1$; $t_2(26) < 1$.

Discussion

Like for the bisyllabic picture names, we see an increase in monitoring latencies from pre-final to final position in trisyllabic picture names. Neither target position to be monitored in this experiment corresponded to the default stress position in Dutch. Nevertheless, pre-final stress was monitored significantly faster than final stress. Following the work by Wheeldon and collaborators, one might take this result to demonstrate that stress is encoded rightward incrementally from the beginning to the end of words during speech production. However, an alternative account according to which monitoring takes place after the entire word has been phonologically encoded cannot be refuted on the basis of the present data. Furthermore, the results of the picture naming part of Experiment 2 showed that there was no difference in naming pictures of either stress category. That is, although pictures with pre-final stress names exhibited a clear advantage over pictures with final stress names in the monitoring task, that advantage disappeared in the naming latencies. This demonstrated once more the relative independence of monitoring in tacit naming from naming aloud (see discussion above). Furthermore, according to psycholinguistic theory, none of the two stress conditions is the default stress position in Dutch. Therefore, any claim about a default/non-default

advantage would not hold in this case. In summary, we showed that the incremental nature of stress monitoring also holds for trisyllabic targets in Dutch. This result might be taken to show that the time course of metrical encoding is independent of the distinction about default/non-default stress position in Dutch. The data of the object recognition experiment again did not show any sign of evidence for the hypothesis that the monitoring effect is due to a visual input effect (i.e., pictures whose names have pre-final stress would be recognised faster than pictures of final stress names). As expected, there was no difference in recognition speed between the two stress categories.

EXPERIMENT 3: MONITORING METRICAL STRESS WITH TRISYLLABIC TARGETS (INCLUDING INITIAL STRESS)

Experiment 2 convincingly showed that even when stress is not in a default position, the monitoring latencies show an incremental pattern. However, it would be elegant to show that initial stress in a trisyllabic target precedes both second and third syllable stress to further strengthen the argument about incremental metrical encoding. Therefore, in Experiment 3, we tested trisyllabic picture names that were stressed either on the first, the second, or the third syllable and asked participants to make a button-press decision about the stress position. Since this task with three alternatives is presumably more difficult than a task with only two alternatives (as was the case in Experiments 1 and 2), we expected on average longer RTs and more errors than in the previous experiments and consequently adjusted our trimming procedure for the monitoring latencies. However, this effect of task difficulty was regarded as constant across all stress conditions in this experiment.

Method

Participants. Thirty-three native Dutch participants from the same pool as for the previous experiments took part in Experiment 3.

Materials. Forty-two trisyllabic, monomorphemic picture names were selected for pictures, which were available in the picture database of the Max Planck Institute for Psycholinguistics. One third of the picture names had initial stress, one third had pre-final stress, and one third had final stress. Pictures of the latter two stress categories were the same as in Experiment 2. The complete list of materials can be found in Appendix C. All items were between five and nine segments (phonemes) long. The item categories had a mean frequency of occurrence between 7 and 9 per

million as determined by CELEX (see Baayen et al., 1995), i.e., all items were of low to moderate frequency. Picture names with initial, pre-final, and final stress were also matched for delayed naming latencies (for details see Table 3). For the object/non-object decision part, we selected 42 pictures of pseudo-objects described earlier.

Procedure and design. The procedure and design for Experiment 3 were similar to Experiment 2 with the exception that in the monitoring part three blocks were presented: one when participants were required to press the button if the target picture name had initial stress, one when they were required to press the button for second syllable stress, and one when they were asked to press the button for final stress. In each block, there were 14 YES-responses and 14 NO-responses. The NO-responses consisted of an equal number of pictures from the other two stress conditions, i.e., each picture was shown twice, once as a YES-response and once as a NO-response, just as in the previous experiments. The order of blocks followed a Latin square design and each order of blocks was assigned an equal number of participants. The order of trials was randomised individually for each block and participant.

Results

Picture naming latencies for picture names with initial stress (804 ms; $SD = 102$) were slightly faster than for both picture names with pre-final stress (838 ms; $SD = 110$) and picture names with final stress (835 ms; $SD = 106$). These differences were significant by participants, but not by items, $F_1(2, 64) = 3.82$, $MSE = 3129.64$, $p < .05$; $F_2(2, 39) < 1$. The errors revealed a similar picture. Altogether, there were 7.3% errors. There were slightly fewer errors in the initial stress condition (6.5%) than in the pre-final stress condition (7.3%) and in the final stress condition (8.2%). However, these differences were not significant, $F_1(2, 64) < 1$; $F_2(2, 39) < 1$.

TABLE 3
Lexico-statistical characteristics of the target words in Experiment 3

Stress location	*Example*	*Mean CELEX frequency (per one million words)*	*Mean delayed naming latencies (in ms)*	*Mean length in segments*
Initial	*ananas*	8.6	347	6.8
Pre-final	*asperge*	9.2	347	7.4
Final	*artisjok*	7.5	349	7.5

Note: The mean delayed naming latencies are based on a study similar to the delayed naming study described in the Discussion of Experiment 1 ($n = 14$ native Dutch participants).

Monitoring latencies shorter than 300 ms were excluded from the analyses (0.1% of the cases). Also, time-outs (27.1%) and errors (12.2%) were not included in the RT analysis. As expected, the error rate increased significantly compared with the previous experiment. Monitoring latencies for first syllable stress (1207 ms; $SD = 190$) were faster than for second syllable stress (1337 ms; $SD = 192$), and the latter in turn were faster than for third syllable stress (1409 ms; SD = 193). The overall effect of stress position was significant, $F_1(2, 64) = 24.78$, $MSE = 13979.42$, $p < .01$; $F_2(2, 39) = 12.67$, $MSE = 8262.61$, $p < .01$). Furthermore, one-tailed *t*-tests showed that the 202 ms difference between initial and final stress was significant, $t_1(32) = 5.90$, $p < .01$; $t_2(26) = 5.93$, $p < .01$) and also the 130 ms difference between initial and pre-final stress, $t_1(32) = 4.37$, $p < .01$; $t_2(26) = 3.00$, $p < .01$. The 72 ms difference between the pre-final and final stress was significant by participants and marginally significant by items, $t_1(32) = 3.29$, $p < .01$; $t_2(26) = 1.64$, $p = .06$.

Error rates point in the same direction. There were more errors on picture names with final stress (16.5%) than on picture names with pre-final stress (10.6%) and picture names with initial stress (9.5%). A similar picture emerges for the time-outs: There were more time-outs in the final stress condition (30.7%) than in the pre-final stress condition (29.9%) or in the initial stress condition (20.8%). For the error analysis, errors and time-outs were collapsed. The main effect of stress position was significant for the error rates, $F_1(2, 64) = 4.72$, $MSE = 9.89$, $p < .05$; $F_2(2, 39) = 5.79$, $MSE = 19.04$, $p < .01$. The individual differences in error rates were significant between the initial and the final stress conditions, $t_1(32) = 2.65$, $p < .05$; $t_2(26) = 4.25$, $p < .01$ and between the initial and the pre-final stress conditions, $t_1(32) = 1.79$, $p < .05$; $t_2(26) = 1.93$, $p < .05$, but not between the pre-final and the final stress conditions, $t_1(32) = 1.54$, n.s.; $t_2(26) = 1.20$, n.s.

Pictures with initial stress names (501 ms; $SD = 88$) were recognised slightly faster than pictures with pre-final stress names (510 ms; $SD = 95$) and pictures with final stress (509 ms; $SD = 81$). However, these differences were not significant, $F_1(2, 64) = 1.02$, $MSE = 829.09$, n.s.; $F_2(2, 39) < 1$. Error rates showed a similar pattern. Overall, there were 3.5% errors (0.9% errors and 2.6% time-outs). There were slightly more errors on initial stress picture names (3.4%) than on pre-final stress picture names (2.3%). Final stress picture names (4.8%) caused most errors. These differences were not significant, however, $F_1(2, 64) < 1$; $F_2(2, 39) < 1$.

Discussion

As in the previous experiments, we observed an increase in monitoring latencies from pre-final to final position, but also from initial to pre-final

position. Following Wheeldon and Levelt (1995; Levelt et al., 1999; Wheeldon & Morgan, 2002), these results might be interpreted as suggesting that stress is encoded rightward incrementally from the beginning to the end of words, although the alternative account according to which lexical stress is monitored after phonological encoding has been completed cannot be completely refuted on the basis of these data. However, the alternative account would predict the monitoring intervals to be approximately equal because the stress-bearing segments are approximately even-spaced in the abstract phonological word representation. In fact, the monitoring interval of the lexical stress between the pre-final and the final syllable (130 ms) is, however, faster than the monitoring interval of lexical stress between the pre-final and the final syllable (72 ms). Although this effect might be problematic for the alternative account, the view stressing the phonological encoding component of the results might be able to account for this difference. Wheeldon and Levelt (1995) found that segmental monitoring increased in speed as target segments occurred towards the end of words. They accounted for this effect by assuming that placing the syllable boundary takes time during which the segmental encoder keeps making available the segments for the second syllable. However, these segments could only be inserted into their slots after the syllable boundary has been computed. Since the segments were already available at that point, they could be inserted and monitored relatively faster than the segments of the first syllable (Wheeldon & Levelt, 1995). Hence, the monitoring system could catch up and yield shorter monitoring latencies than for the segments in the first syllable. A similar account might be offered for the data of Experiment 3. Assuming that the computation of syllable boundaries takes some time and assuming that segments are retrieved at a relatively constant speed, segments of the final syllable could be associated with their metrical frame at a faster speed than segments of the pre-final and initial syllable. This might be reflected in faster stress monitoring latencies and therefore be taken as an argument that the monitoring latencies reflect the genuine phonological encoding process. Unfortunately, the monitoring latencies from the first two experiments cannot directly be compared with each other to test this hypothesis.

The picture naming results of Experiment 3 showed that—at least for participants but not for items—there was a difference between the different stress categories. This difference, however, did not match the observed monitoring pattern. Whereas for monitoring we observed a continuous increase in reaction times from first to third syllable stress, in naming the second syllable stress condition was the slowest. This demonstrated once more the relative independence of monitoring from naming (see above). Furthermore, two of the three stress conditions are

non-default stress positions in Dutch. Therefore, any claim about a default/non-default advantage would not hold in this case. This latter result replicates, in fact, the outcome of Experiment 2 with different participants. In summary, we showed that the incremental nature of stress monitoring holds for each single stress position of trisyllabic targets in Dutch.

As in the previous two experiments, the data of the object/non-object recognition experiment again did not show any sign of evidence for the hypothesis that the monitoring effect is due to a visual input effect (i.e., pictures whose names have initial stress would be recognised faster than pictures with pre-final stress names, which in turn would be recognised faster than pictures with final stress names). As expected, there was no difference in recognition speed between the three stress categories.

GENERAL DISCUSSION

In this paper, we modified a methodology introduced by Wheeldon and Levelt (1995) to investigate monitoring of metrical stress during language production. The results of Wheeldon and Levelt's study demonstrated that the representation on which the monitoring response for individual segments is based is phonological and syllabified in nature. Participants are monitoring an internal abstract code, i.e., the output of the process that assigns segments (phonemes) to a syllabified prosodic frame. Our present results support the view that the nature of the representation underlying the monitoring of internal speech is prosodified.

Here, we were especially interested in metrical stress. The results of Experiments 1, 2, and 3 showed that metrical information becomes available incrementally to the monitoring system. Participants were significantly faster in deciding about the stress location when a bisyllabic picture name had initial stress than when it had final stress. The metrical frame of bisyllabic words is presumably monitored from beginning to end. If a word has initial stress (e.g., *KAno*), this information is available earlier in monitoring than when a word has final stress (e.g., *kaNON*). This demonstrates that not only segmental but also metrical monitoring proceeds incrementally. Similar results were found for trisyllabic targets, i.e., the earlier the stress was located in the word, the faster participants were to respond. The serial order information about metrical monitoring is important because it shows that not only relatively concrete elements like segments are monitored sequentially; suprasegmental units such as metrical stress are also monitored in that way (see Figure 1). Meyer (1990, 1991), Levelt and Wheeldon (1994), and Van Turennout et al. (1997) showed that the segmental encoding of speech is essentially an incremental process. Of course, overt speech is a sequential process and

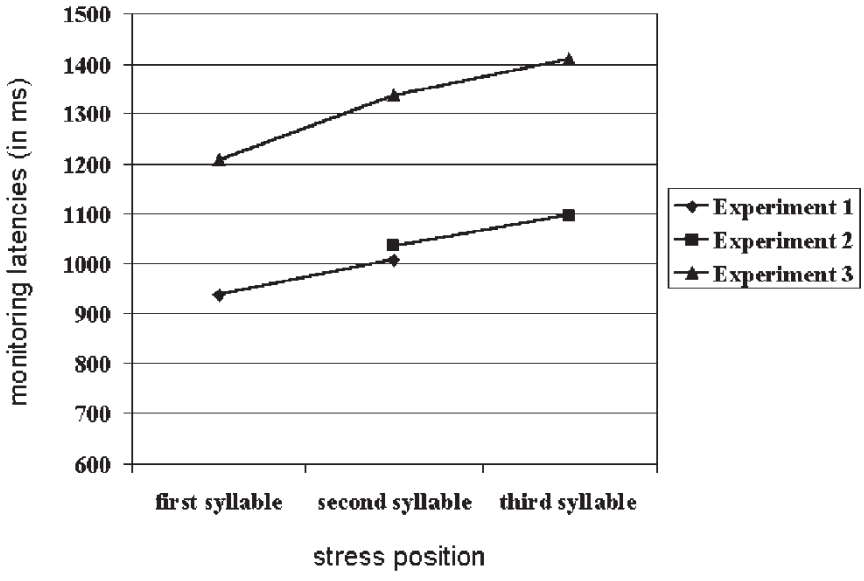


Figure 1. Monitoring latencies in Experiments 1, 2, and 3 as a function of the stress position of the target picture names.

necessarily has to proceed from beginning to end. But the studies mentioned above investigated the phonological *planning stage* of word generation and found strict serial ordering effects. Our present data might be interpreted in a similar way. If the present data reflect effects of temporal ordering during phonological encoding in speech production, then one can conclude that the time course of metrical stress encoding is also rightward incremental.

However, one may argue that monitoring for metrical stress may not be independent of segmental monitoring. For instance, when the phonological word is incrementally constructed, particular segments (e.g., vowels) might be marked for stress (e.g., [+ stress] or [– stress]). If this were the case, then the time course of metrical encoding might just be a by-product of the incremental segmental encoding process. This would imply that metrical stress forms part of the segmental representation. However, Roelofs and Meyer (1998) showed that when speakers could prepare the initial segments (e.g., *ma*) of to-be-produced words (e.g., *ma.RI.ne*, *ma.nus.CRIPT*, *ma.TE.rie*, *ma.de.LIEF*), but not the (non-default) metrical structure (e.g., /_'_ _/ vs. /_ _'_/), there was no preparation effect. Note that the first syllable *ma* was unstressed in all target words in the example above. If stress was encoded directly on the segments instead of in a separate frame, a preparation effect should have occurred because the first two segments of all target words were identical

to the encoder (i.e., [– stress]) in the example above. The fact that such an effect did not occur, however, suggests that segmental and metrical spell-out are—to some extent—independent. Roelofs and Meyer (1998) only obtained a preparation effect when target words shared initial segments, had the same stress pattern, and the same number of syllables (i.e., metrically identical).

There are at least two possibilities as to how internal monitoring of stress (or segments) might work. We will refer to them as *production monitoring* and *perception monitoring*. Let us first describe production monitoring (see also Laver, 1980). A production monitor may be a device that can monitor for certain entities (e.g., segments, stress, etc.) in the course of building a phonological word during phonological encoding (Levelt et al., 1999; see introduction above). For instance, every time a new segment or a new syllable has been encoded, the monitor moves a segment or a syllable closer towards the end of the phonological word. If a target segment or stress value is detected, the monitor sends a response to the central executive system such that the button-press can be executed. Since the production monitor is moving in parallel with phonological word encoding, targets occurring at the beginning of a phonological word would be detected earlier than targets occurring towards the end. This could, for instance, account for the sequential effect in segment monitoring found by Wheeldon and Levelt (1995) and in our metrical stress monitoring data.

Alternatively, a perception monitor may handle the process of monitoring during speech production. When (part of) a phonological word has been encoded, it may be transferred into a buffer where the speech plan can be stored temporarily (Hartsuiker & Kolk, 2001; Levelt, 1989). A perception monitor may be a device that can monitor this buffered representation for certain entities (e.g., segments, stress, etc.). This would be a perceptual process since the monitor scans a previously created phonological representation. The perception monitor starts at the beginning of a phonological word and moves towards the end. If a target segment or stress value is detected, the monitor sends a response signal to the central executive system such that the button-press response can be executed. Since the perception monitor moves from beginning to end of a stored phonological word, targets occurring at the beginning of a phonological word would be detected earlier than targets occurring towards the end. Therefore, perception monitoring could also account for the sequential pattern Wheeldon and Levelt (1995) found in their segment monitoring experiments since perceptual monitoring reflects the time course of speech production processes such as phonological encoding. The same holds for our data, but they speak to the temporal ordering of metrical stress, and therefore fit into the general picture of incremental phonological encoding.

Our data do not allow us to decide unambiguously whether the monitoring effect is due to the incremental creation of phonological words (production monitoring) or rather to the left-to-right nature of a monitoring process that scans a previously created phonological representation (perception monitoring). However, as argued in the Discussion of Experiment 3, there are certain aspects of the present data that could potentially be accounted for by production monitoring but not so easily by perception monitoring. The presumably simplest way to account for the monitoring effects we reported in this study is to assume that monitoring is done via the general comprehension system (Levelt, 1989). However, this perceptual monitoring reflects the time course of phonological encoding, i.e., a genuine speech production process.

CONCLUSION

Planning stages in speech production can be taken as a particular instance for the study of serial order in behaviour (Lashley, 1951). The results of the present study are likely to reflect effects of serial order in speech production and compare nicely to results of other studies on phonological encoding planning (especially Wheeldon & Levelt, 1995). We now have on-line evidence from different paradigms for the incremental encoding of phonological information in spoken language production although the possibility that monitoring latencies do not reflect genuine production processes cannot be refuted completely. Segments are first assigned to the slots made available by the metrical frame of the first syllable. Once the first syllable has been encoded and the syllable boundary has been placed, the following syllabic frame is filled from beginning to end until the prosodic frame of the whole phonological word has been filled with segments. Then this phonological word can be phonetically encoded and the phonological encoder eventually moves on to the next phonological word. Our data not only fit into the general incremental ordering during phonological encoding, but they specifically speak to the temporal processing of stress.

REFERENCES

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

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

Cappa, S., Nespor, M., Ielasi, W., & Miozzo, A. (1997). The representation of stress: Evidence from an aphasic patient. *Cognition*, *65*, 113.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

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

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.

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*, 635–646.

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

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

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

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). San Diego, CA: Academic Press.

Garrett, M. F. (1980). 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.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

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.

Guenther, F. H. (2003). Neural control of speech movements. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 209–239). Berlin: Mouton de Gruyter.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113–157.

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

Jansma, B. M., & Schiller, N. O. (2004). Monitoring syllable boundaries during speech production. *Brain and Language*, *90*, 311–317.

Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

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

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 862–885.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, *23*, 39–66.

Laganaro, M., Vacheresse, F., & Frauenfelder, U. H. (2002). Selective impairment of lexical stress assignment in an Italian-speaking aphasic patient. *Brain and Language*, *81*, 601–609.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium* (pp. 112–146). New York, NY: John Wiley & Sons.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 287–305). New York, NY: Academic Press.

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

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

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 122.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*, 13464–13471.

Levelt, W. J. M., Roelofs, A, & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–37.

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

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. (2000). Form representation in word production. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 49–70). Philadelphia, PA: Taylor & Francis.

Miceli, G., & Caramazza, A. (1993). The assignment of word stress in oral reading: Evidence from a case of acquired dyslexia. *Cognitive Neuropsychology*, *10*, 273–296.

Morgan, J. L., & Wheeldon, L. R. (2003). Syllable monitoring in internally and externally generated English words. *Journal of Psycholinguistic Research*, *32*, 269–296.

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

Nickels, L., & Howard, D. (1999). Effects of lexical stress on aphasic word production. *Clinical Linguistics and Phonetics*, *13*, 269–294.

Pechmann, T., Reetz H., & Zerbst, D. (1989). Kritik einer Meßmethode: Zur Ungenauigkeit von Voice-key Messungen [Critique of a method of measurement: On the unreliability of voice-key measurements]. *Sprache & Kognition*, *8*, 65–71.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*, 97–131.

Price, C. J., Wise, R. J., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S. J., & Friston, K. J. (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain*, *119*, 919–931.

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

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

Roelofs, A. (2000). WEAVER++ and other computational models of lemma retrieval and word-form encoding. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 71–114). Philadelphia, PA: Taylor & Francis.

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.

Schiller, N. O. (in press). Verbal self-monitoring. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schiller, N. O., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language*, *90*, 231–240.

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*, 828.

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., van Lieshout, P. H. H. M., Meyer, A. S., & Levelt, W. J. M. (1999). Does the syllable affiliation of intervocalic consonants have an articulatory basis? Evidence from electromagnetic midsagittal articulography. In B. Maassen & P. Groenen (Eds.), *Pathologies of speech and language. Advances in clinical phonetics and linguistics* (pp. 342–350). London: Whurr Publishers.

Trommelen, M., & Zonneveld, W. (1989). *Klemtoon en metrische fonologie* [Stress and metrical phonology]. Muiderberg: Coutinho.

Trommelen, M., & Zonneveld, W. (1990). Stress in English and Dutch: A comparison. *Dutch Working Papers in English Language and Linguistics* (Vol. 17). Leiden, The Netherlands: Rijksuniversiteit Leiden.

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

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, *280*, 572–574.

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

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

Zonneveld, W., Trommelen, M., Jessen, M., Bruce, G., & Árnason, K. (1999). Word-stress in West-Germanic and North-Germanic languages. In H. van der Hulst (Ed.), *Word prosodic systems* (pp. 477–603). Berlin: Mouton de Gruyter.

Appendix A

Materials used in Experiment 1 (with English translation between brackets)

Targets with initial stress		Targets with final stress	
CV	CVC	CV	CVC
boter (''butter'')	banjo (''banjo'')	banaan (''banana'')	balkon (''balcony'')
jager (''hunter'')	borstel (''brush'')	beha (''bra'')	dolfijn (''dolphin'')
kabel (''cable'')	bunker (''bunker'')	bureau (''desk'')	garnaal (''shrimp'')
kano (''canoe'')	dokter (''doctor'')	citroen (''lemon'')	gordijn (''curtain'')
kegel (''bowling pin'')	gondel (''gondola'')	fabriek (''factory'')	kalkoen (''turkey'')
ketel (''kettle'')	herder (''shepherd'')	gebit (''dentures'')	karkas (''skeleton'')
koning (''king'')	kansel (''pulpit'')	giraf (''giraffe'')	kasteel (''castle'')
motor (''motor bike'')	lifter (''hitch hiker'')	gitaar (''guitar'')	lantaarn (''lantern'')
nagel (''finger nail'')	panter (''panther'')	kameel (''camel'')	magneet (''magnet'')
ratel (''rattle'')	parfum (''perfume'')	kanon (''canon'')	penseel (''brush'')
robot (''robot'')	pleister (''band aid'')	konijn (''rabbit'')	pincet (''tweezers'')
spijker (''nail'')	scalpel (''scalpel'')	libel (''dragonfly'')	pistool (''gun'')
tijger (''tiger'')	tempel (''temple'')	matras (''mattress'')	portret (''portrait'')
toren (''tower'')	tractor (''tractor'')	raket (''rocket'')	sandaal (''sandal'')
vogel (''bird'')	wortel (''carrot'')	sigaar (''cigar'')	soldaat (''soldier'')
zebra (''zebra'')	zuster (''nurse'')	tomaat (''tomato'')	trompet (''trumpet'')

Appendix B

Materials used in Experiment 2

Targets with pre-final stress	Targets with final stress
capsule (‘‘capsule’’)	artisjok (‘‘artichoke’’)
asperge (‘‘asparagus’’)	baviaan (‘‘baboon’’)
flamingo (‘‘flamingo’’)	batterij (‘‘battery’’)
triangel (‘‘triangle’’)	liniaal (‘‘ruler’’)
komkommer (‘‘cucumber’’)	diamant (‘‘diamond’’)
kabouter (‘‘gnome’’)	dirigent (‘‘conductor’’)
punaise (‘‘thumbtack’’)	envelop (‘‘envelope’’)
horloge (‘‘watch’’)	papegaai (‘‘parrot’’)
diskette (‘‘floppy disk’’)	microscoop (‘‘microscope’’)
gorilla (‘‘gorilla’’)	schilderij (‘‘painting’’)
piano (‘‘piano’’)	hagedis (‘‘lizard’’)
judoka (‘‘judoka’’)	klarinet (‘‘clarinet’’)
trombone (‘‘trombone’’)	astronaut (‘‘astronaut’’)
computer (‘‘computer’’)	krokodil (‘‘crocodile’’)

Appendix C

Materials used in Experiment 3

Targets with initial stress	Targets with pre-final stress	Targets with final stress
ananas (‘‘pineapple’’)	capsule (‘‘capsule’’)	artisjok (‘‘artichoke’’)
adelaar (‘‘eagle’’)	asperge (‘‘asparagus’’)	baviaan (‘‘baboon’’)
boemerang (‘‘boomerang’’)	flamingo (‘‘flamingo’’)	batterij (‘‘battery’’)
camera (‘‘camera’’)	triangel (‘‘triangle’’)	liniaal (‘‘ruler’’)
kakkerlak (‘‘cockroach’’)	komkommer (‘‘cucumber’’)	diamant (‘‘diamond’’)
kandelaar (‘‘candlestick’’)	kabouter (‘‘gnome’’)	dirigent (‘‘conductor’’)
kangoeroe (‘‘kangaroo’’)	punaise (‘‘thumbtack’’)	envelop (‘‘envelope’’)
lucifer (‘‘match’’)	horloge (‘‘watch’’)	papegaai (‘‘parrot’’)
octopus (‘‘octopus’’)	diskette (‘‘floppy disk’’)	microscoop (‘‘microscope’’)
olifant (‘‘elephant’’)	gorilla (‘‘gorilla’’)	schilderij (‘‘painting’’)
ooievaar (‘‘stork’’)	piano (‘‘piano’’)	hagedis (‘‘lizard’’)
caravan (‘‘trailer’’)	judoka (‘‘judoka’’)	klarinet (‘‘clarinet’’)
paprika (‘‘pepper’’)	trombone (‘‘trombone’’)	astronaut (‘‘astronaut’’)
radio (‘‘radio’’)	computer (‘‘computer’’)	krokodil (‘‘crocodile’’)