Metrical stress in speech production: A time course study

Niels O. Schiller[†,‡]
† University of Maastricht, The Netherlands
‡ Max Planck Institute for Psycholinguistics, The Netherlands
E-mail: n.schiller@psychology.unimaas.nl & niels.schiller@mpi.nl

ABSTRACT

This study investigated the encoding of metrical information during speech production in Dutch. In Experiment 1, 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., *LEpel* 'spoon') than for targets with final stress (e.g., *liBEL* 'dragon fly'; capital letters indicate stressed syllables) and revealed that the 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 bi- and trisyllabic picture names. These results demonstrate that metrical information of words is encoded rightward incrementally during phonological encoding in speech production. The results of these experiments are in line with Levelt's model of phonological encoding.

1. INTRODUCTION

Word form encoding or phonological encoding can be further divided into a number of processes (see recent overview in [1]). Levelt and colleagues presented the most fine-grained model of phonological encoding to date [2]. 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. The phonological encoding system must retrieve the corresponding segments and the metrical frame of a word form. 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, 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. /_ '_ _/; but see [3] for a different perspective). Evidence for the relative independence of segmental and metrical retrieval comes, for instance, from speech errors that involve stress exchange, such as *my PROsodic /prosədɪk/ – proSOdic /prosɒdɪk/ colleagues* (example from [4]). These errors are rare [1], but the fact that they exist supports the idea that there is an encoding stage for stress. Furthermore, Roelofs and Meyer found that speakers had to know the number of syllables and the stress pattern of the *entire* response word to be able to prepare the initial segments of a response word ([5]; see details below). This suggests that metrical information is retrieved independently of segmental information.

Then, during segment-to-frame association previously retrieved segments are combined with their metrical frame. The retrieved ordering of segments prevents them from being scrambled. They are inserted incrementally into slots made available by the metrical frame to build a so-called phonological word. This incremental syllabification process respects universal and language-specific syllabification rules, e.g. *ba.NA.na.* [1] Evidence for the incremental ordering during segmental encoding comes from a number of studies using different experimental paradigms (e.g., [6,7,8,9,10,11]). The reason for "spelling out" lexical words only to rebuild them again into phonological words lies in the necessity to form maximally pronounceable syllables (see [2], pp. 19-22 for details): In connected speech, phonological words often have syllable structures deviating from the canonical syllable structures of the lexical words (see example in endnote 1). The domain of syllabification is the so-called phonological word and may be larger (clitics) or smaller (compounds) than the input lexical words themselves [12]. Segment-to-frame association is the process that lends the necessary flexibility to the system.

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 [13]. Syllables in the syllabary may possibly be represented in terms of gestural scores [14] specifying articulatory motor programs for syllable-sized chunks. Although there is very little on-line evidence for the use of syllables in speech production ([15,16]; but see [17,18,19]; see also [20]), the idea of having precompiled syllabic motor programs is very attractive because it decreases the computational load of the phonological/phonetic encoding component ([13,21]; for lexico-statistical support see [22]). One idea is that syllables in the syllabary are activated through their segments and selected on the basis of Luce's choice rule (for details see [2,23]). In case there is no corresponding syllable in the syllabary, it has to be computed on the fly by concatenating individual segments. Once the syllabic gestural scores are made available, they can be translated into neuro-motor programs, sent to the articulators, and then be executed resulting in overt speech. Levelt's theory does not assume that the exact articulatory movement trajectories are programmed, but rather

ISBN 1-876346-48-5 © 2003 UAB

neuromuscular speech tasks to be achieved [24,25].

As stated above, a number of studies showed that phonological words are encoded incrementally. Meyer ([6], [7]) 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 [6,7]. This was taken as evidence that segmental encoding proceeds in an incremental fashion from word onset to word offset during phonological encoding.

2. METRICAL ENCODING

Although we know some details about the segmental encoding of phonological words, relatively little is known about their metrical encoding. Roelofs and Meyer [5] found that participants can only prepare the beginning of a word when they know the initial segments AND the metrical frame (e.g., number of syllables and stress location) of the target. Knowing only the segments OR only the metrical frame of the target did not facilitate naming. For instance, when participants knew that a word started with *ma* but they did not know whether the target was bisyllabic (e.g., *maJOOR* 'major'), trisyllabic (e.g., *maTErie* 'matter'), or quadrisyllabic (e.g., *maLAria* 'malaria'), there was no preparation effect. Roelofs and Meyer [5] took their data as evidence that speakers need to have knowledge not only about the segmental structure of words but also about the metrical structure of a to-be-named target to be able to prepare overt speech. Furthermore, their data suggest that metrical frames for entire words are retrieved at once because otherwise preparation effects should have been found for words sharing their initial segments but not their number of syllables. However, so far, this is only one study and we would like to investigate metrical encoding more deeply. One question in metrical encoding is how the metrical frame of a word form is computed.[2] Is metrical information – just like segmental information – also encoded incrementally? This study focuses on the time course of metrical encoding using a speech production task (implicit picture naming).

3. THE EXPERIMENTS

One way to investigate questions about the speech production process is to use picture naming. In picture naming, presumably all stages of the speech production process have to be completed, i.e. conceptualization, lexical access, word form encoding, and articulation (see [26] 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 key when the word fulfilled a certain phonological criterion and withhold the key 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.

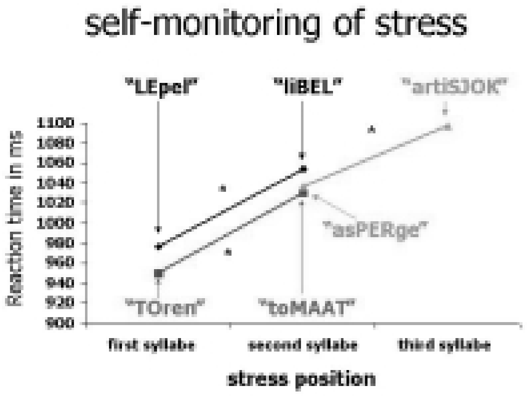
We conducted three experiments, all using the same procedure. Participants all had normal or corrected-to-normal vision and were paid for their participation. All participants were native speakers of Dutch. The materials consisted of bisyllabic (Experiments 1 and 2) and trisyllabic (Experiment 3) monomorphemic Dutch nouns. Line drawings of the corresponding objects were either taken from the set of pictures of the Max Planck Institute for Psycholinguistics or drawn by a professional artist. Items could be divided into groups of equal size depending on the location of their lexical stress. For lexico-statistical details on the materials and complete lists of all items see [27].

Participants were tested individually. They were seated behind a computer screen and asked to place their right index finger on the button of a button box that was placed in front of them. For each experimental trial they were asked to press the button when the picture name had first-syllable stress (e.g., *LEpel* 'spoon'). In case the picture name had second-syllable stress, they were required to withhold the button press (e.g., *liBEL* 'dragon fly'). In a second block, instructions were switched and the same pictures were shown again in order to get a response for every item. In Experiment 3, there were two blocks in which participants were required to decide whether picture names had second or third-syllable stress. An experimental trial consisted of the following events: first, a fixation point appeared for 500 ms in the center of the screen, which participants were asked to fixate. Then, after 300 ms, a picture appeared around the same location on the screen. Pictures were of approximately equal size. They all fitted into a 7 x 7 cm square. As soon as possible after the picture appeared participants were asked to respond. 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.

Before the experimental trials started, participants were familiarized with the pictures. Each picture was shown individually with the picture name underneath until the participant pressed the space bar and the next picture appeared. After picture familiarization, 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. In the experimental trials participants were asked to suppress overt naming of the pictures and – if necessary – press the button as fast and as accurately as possible after a picture appeared on the screen.

The experiments started with a familiarization and a practice block. Then two test blocks followed with reversed instructions. After each block there was a short break. The order of trials was randomized for each block and each participant individually. 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 had to withhold response for names with second syllable stress. They than got a second block with the same material in which the response contingencies were reversed. The other half of the subjects got the reversed block order. In Experiment 3, there were also two groups, one that started with responses to second syllable stress, and one with responses to third syllable stress.

4. RESULTS AND DISCUSSION



The results of the experiments are clear-cut (see Figure above). Native speakers of Dutch are faster and more accurate in deciding that a bisyllabic word has first-syllable stress than in deciding that it has second-syllable stress. Furthermore, they are faster to decide that a trisyllabic picture name has second-syllable stress than that it has third-syllable stress. Control experiments showed that the monitoring latencies were different from the picture naming latencies and the picture recognition latencies. We interpret this result 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 how metrical information is encoded. The above experiments showed that the direction of metrical planning – or the monitoring of metrical information – is also rightward incremental.

The effects obtained in the experiments seem relatively large compared to other form encoding effects in speech production such as, e.g., form priming (10 to 15 ms per segment, see [17,28]). However, to decide whether a word has initial or final stress in a bisyllabic word, speakers have to generate and monitor two different syllables (first vs. second syllable). Wheeldon and Levelt argued from their second experiment that participants are monitoring a syllabified representation because they found a syllabic monitoring effect [10]. In their third experiment they found a 111 ms monitoring difference between first syllable onsets and second syllable onsets. This number is comparable to our 60-70 ms differences between stress positions. Thus, we have converging evidence that the planning, i.e. the phonological encoding of a syllable plus its metrical structure takes about 100 ms. The overt articulation of syllables, however, is much slower and takes about twice as long [10,29], demonstrating that articulation is an extra step independent of metrical information access.

5. CONCLUSIONS

It has been shown many times that the segmental encoding of speech is essentially an incremental process [6,7,9,10]. Of course, overt speech is a sequential process and 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. We investigated – for the first time – the planning stage of metrical stress assignment, and observed that it takes about 60-70 ms to detect first vs. second syllable vs. third syllable stress, indicating the time course of metrical stress encoding. In general, our results fit nicely into the picture of phonological encoding drawn by Levelt and Wheeldon [13]. These authors proposed that the prosodic frame of a word is made available syllable-by-syllable and that the corresponding segments are subsequently assigned to their slots in the syllable frame.

ACKNOWLEDGEMENTS

Niels O. Schiller was supported by the Royal Dutch Academy of Arts and Sciences (KNAW) and by a Vernieuwingsimpuls grant from the Dutch Organization for Scientific Research (NWO).

REFERENCES

[1] Meyer, A. S. (2000). Form representations in word production. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 49-70). Hove: Psychology Press.

[2] 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.

[3] Schiller, N. O., Fikkert, P., & Levelt, C. C. (submitted). Stress priming in picture naming: An SOA study.

[4] Cutler, A. (1983). Speakers' conceptions of the function of prosody. In A. Cutler & D. R. Lad (Eds.), *Prosody: Models and measurements* (pp. 79-92). Heidelberg: Springer.

[5] 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.

ISBN 1-876346-48-5 © 2003 UAB

[6] 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.

[7] 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.

[8] Schiller, N. O. (in preparation). The incremental nature of phonological in speech production.

[9] 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.

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

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

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

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

[14] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, ***49***, 155-180.

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

[16] Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, ***35***, 458-470.

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

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

[19] 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.), *Laboratory Phonology 7* (pp. 35-59). Berlin: Mouton de Gruyter.

[20] Cholin, J., Schiller, N. O., & Levelt, W. J. M. (submitted). The preparation of syllables in speech production.

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

[22] 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.

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

[24] Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production: Speech and talk*, vol. 1 (pp. 373-420). New York, NY: Academic Press.

[25] Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, ***14***, 29-59.

[26] Glaser, W. R. (1992). Picture naming. *Cognition*, ***42***, 61-105.

[27] Schiller, N. O., Schmitt, B. M., Peters, J., & Levelt, W. J. M. (submitted). Phonological encoding of stress: The time course of metrical encoding in polysyllabic words.

[28] Schiller, N. O. (submitted). The onset effect in word and picture naming.

[29] Waals, J. (1999). *An experimental view of the Dutch syllable*. Ph.D. dissertation (Netherlands Graduate School of Linguistics; 18). The Hague: Holland Academic Graphics.

---

[1] A phonological word is not necessarily identical to the syntactic 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/.

[2] We will not be concerned with the question whether the metrical frame is stored or derived by rule. Levelt et al.'s [2] position is that the metrical structure of regular words is derived by a simple default rule ("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). Only for irregular words (less than 10% of the word tokens) the metrical frame must be stored. However, the evidence for this position is inconclusive at the moment (see also [3]). Nevertheless, whether the metrical frame is generated on-line or retrieved from the mental lexicon, at some point the frame has to be made available to the phonological encoder.