Phonological encoding in speech production

Niels O. Schiller

Department of Cognitive Neuroscience, Maastricht University, The Netherlands

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Leiden Institute for Brain and Cognition (LIBC), The Netherlands

<https://doi.org/10.36505/ExLing-2006/01/0008/000008>

Abstract

Language production comprises conceptual, grammatical, and word form encoding as well as articulation. This paper focuses on word form or phonological encoding. Phonological encoding in speech production can be subdivided into a number of sub-processes such as segmental, metrical, and syllabic encoding. Each of these processes is briefly described and illustrated with examples from my own research. Special attention is paid to time course issues introducing behavioural and electro-physiological research methods such as LRPs and ERPs. It is concluded that phonological encoding is an incremental planning process taking into account segmental, metrical, and syllabic encoding.

Models of spoken language production

Models of speech production (e.g., Caramazza, 1997; Dell, 1986, 1988; Fromkin, 1971; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, and 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 Figure 1). 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., *table* /tEɪb↔l/) 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., /t/, /Eɪ/, /b/, /↔/, /l/), while during metrical retrieval the metrical frame of a word is re-retrieved, which consists at least of the number of syllables and the location of the lexical stress (e.g., for *Table* – capital letters mark stressed syllables – this would be a frame consisting of two syllables the first of which is stressed).

Figure 1. A model of phonological encoding in speech production (slightly adapted from Levelt and Wheeldon, 1994).

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 ($/t/_1$, $/EI/_2$, $/b/_3$, $/\leftrightarrow/_4$, $/l/_5$). 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. *TA.ble* (dots mark syllable boundaries). 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* /gEI.vIt/. Roelofs (1997) 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 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, 2006; Cholin, Schiller, & Levelt, 2004; 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).

Segmental encoding

Word forms activate their segments and the rank order in which these segments have to be inserted into a phonological frame with slots for each segment (slot-filler-theory; Shattuck-Hufnagel, 1979 for an overview). Evidence for this hypothesis comes, for instance, from speech errors such as “**q**ueer old **d**ean” instead of “**d**ear old **q**ueen”, a spoonerism. These errors show that word forms are not retrieved as a whole, but rather they are computed segment by segment. Retrieving all segments separately and putting them together into word frames afterwards may seem more complicated than retrieving word forms as a whole. However, this mechanism has an important function when it comes to the production of more than one word. Usually, we do not speak in single, isolated words, but produce more than one word in a row. Let us take the above example *gave it* /gEI.vIt/. Whereas *gave* is a monosyllabic CVC word, the phrase *gave it* consists of a CV and a CVC syllable. That is, the syllable boundaries straddle word or lexical boundaries. In other words, the syllabification process does not respect lexical boundaries because the linguistic domain of syllabification is not the lexical word, but the phonological word (Booij, 1995). Depending on the phonological context in the phonological word, the syllabification of words may also change. Therefore, it does not make a lot of sense to store syllable boundaries with the word forms in the mental lexicon since syllable boundaries may

change during the speech production process as a function of the phonological context (Levelt & Schiller, 1998). Syllable boundaries will be generated on-line during the construction of phonological words to yield maximally pronounceable syllables. This architecture lends maximal flexibility to the speech production system in different phonological contexts.

Time course of segmental processing

One important question in word form encoding is the time course of the processes involved. For instance, are the segments of a word encoded one after the other or are they encoded in parallel? It was argued above on the basis of empirical evidence (e.g., sound errors) as well as on theoretical grounds that word forms are planned in terms of abstract units called segments or phonemes. Behavioural evidence for these claims has been provided in priming studies by Meyer (1990, 1991) and in self-monitoring studies by Wheeldon and Levelt (1995), Wheeldon and Morgan (2002), and Schiller (2005). For a summary of these studies see Schiller (2006).

However, there are also electrophysiological studies investigating the time course of segmental encoding. Van Turennout et al. (1997), for instance, investigated the time course of segmental encoding using lateralized readiness potentials (LRPs). The LRP is a derivative of the electroencephalogram (EEG) which can be measured by using scalp electrodes. Participants in Van Turennout et al.'s experiment named pictures on a computer screen, one at a time. Whenever a visual cue was presented, participants were requested to carry out a dual task (retrieve certain properties about the to-be-named word) and afterwards name the picture. For instance, participants were asked to make a decision about the animateness of the target concept and about the identity of the initial and final segment of the word. Interestingly, the onset of the nogo-LRP started to develop about 80 ms earlier when the segment was at the onset of words than when it was at the offset of words. This has been interpreted as reflecting the time course of the availability of phonological segments during phonological encoding in speech production planning.

The targets in the Van Turennout et al. (1997) study were 1.5 syllables long on average. Dividing 80 ms by 1.5 corresponds well to the 55 ms difference reported by Wheeldon and Levelt (1995) for the monitoring of syllable onset vs. offset phonemes. One may assume that phonological encoding of a whole syllable takes approximately 50 to 60 ms.

Metrical encoding

The above-mentioned studies investigating the time course of segmental encoding all have in common that they assume the measured effects to take

place at the level of the phonological word. This holds both for the priming studies by Meyer (1990, 1991) and for the monitoring studies by Wheeldon and Levelt (1995) as well as Van Turennout et al. (1997). However, it is unclear how the metrical stress of words is retrieved and encoded. Roelofs and Meyer (1998) found evidence that metrical stress of words is retrieved from the lexicon when it is in non-default position. However, Schiller, Fikkert, and Levelt (2004) could not find any evidence for stress priming in a series of picture-word interference experiments. Schiller et al. (2004) suggested that lexical stress may be computed according to language-specific linguistic rules (see also Fikkert, Levelt, & Schiller, 2005). Furthermore, lexical stress may be encoded incrementally – just like segments – or it may become available in parallel.

Time course of metrical processing

To investigate the time course of metrical processing, Schiller and colleagues employed a tacit naming task and asked participants to decide whether the bisyllabic name of a visually presented picture had initial or final stress. Their hypothesis was that if metrical encoding is a parallel process, then there should not be any differences between the decision latencies for initial and final stress. If, however, metrical encoding is also a rightward incremental process – just like segmental encoding –, then decisions to picture names with initial stress should be faster than decision latencies to picture names with final stress. The latter turned out to be the case (Schiller, Jansma, Peters, & Levelt, 2006). However, Dutch – like other Germanic languages – has a strong preference for initial stress. More than 90% of the words occurring in Dutch have stress on the first syllable. Therefore, this effect might have been due to a default strategy. However, when pictures with trisyllabic names were tested, participants were still faster to decide that a picture name had penultimate stress (e.g., *asPERge* 'asparagus') than that it had ultimate stress (e.g., *artiSJOK* 'artichoke'). This result suggests that metrical encoding proceeds from the beginning to the end of words, just like segmental encoding.

Recently, Schiller (in press) extended this research into the area of electrophysiology. Event-related brain potentials have the advantage of being able to determine processes more precisely in time, whereas behavioural studies such as reaction time studies can only measure the end of processes. In his study, Schiller (in press) used N200 effects to measure the availability of lexical stress in the time course of speech planning. He replicated the behavioural effect demonstrated by Schiller et al. (2006) and showed that the N200 peak latencies were significantly earlier when stress was on the first as compared to the second syllable. Furthermore, the N200 effects occurred in a

time window (400-500 ms) previously identified by Indefrey and Levelt (2004) for phonological encoding.

Syllabic encoding

We have already mentioned above that syllables are presumably created on the fly during speech production. There is quite some linguistic and psycholinguistic evidence (see Cholin et al., 2004 for a recent review and some new data) for the existence of syllables. However, in Levelt's model syllables form the link between the phonological planning process and the articulatory-motor execution of speech in a so-called *mental syllabary* (Levelt, 1989; Levelt et al., 1999). Such a mental syllabary is part of long-term memory comprising a store of syllable-sized motor programs. Ferrand and colleagues (1996, 1997) reported on-line data confirming the hypothesis about a mental syllabary, but Schiller (1998, 2000; see also Schiller, Costa, & Colomé, 2002 and Schiller & Costa, in press) disconfirmed this finding. Rather the results of these latter studies support the idea that syllables are not retrieved, but created on-line during phonological encoding.

The existence of the mental syllabary hinges on the existence of syllable frequency effects. Levelt and Wheeldon (1994) were the first to report effects of syllable frequency effects. However, segment frequency was not controlled well enough and therefore these results are not conclusive. Recently, Cholin et al. (2006) were able to demonstrate syllable frequency effects in very controlled set of materials. Following Schiller (1997), they used quadruples of CVC syllables controlling the segments in onset and offset position (e.g., HF *kem* – LF *kes* and HF *wes* – LF *wem*; HF = high frequency, LF = low frequency). In two experiments, Cholin et al. (2006) showed that HF syllables were named significantly faster than LF syllables. So far, this study includes the best controlled materials demonstrating a syllable frequency effect and hence evidence in favour of a mental syllabary, which may be accessed during phonological encoding.

References

- Booij, G. 1995. The phonology of Dutch. Oxford, Clarendon Press.
- Caramazza, A. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14, 177-208.
- Cholin, J., Levelt, W. J. M., and Schiller, N. O. 2006. Effects of syllable frequency in speech production. *Cognition* 99, 205-235.
- Cholin, J., Schiller, N. O., and Levelt, W. J. M. 2004. The preparation of syllables in speech production. *Journal of Memory and Language* 50, 47-61.
- 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.
- Ferrand, L., Segui, J., and Grainger, J. 1996. Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language* 35, 708-723.
- Ferrand, L., Segui, J., and Humphreys, G. W. 1997. The syllable's role in word naming. *Memory & Cognition* 35, 458-470.
- Fikkert, P., Levelt, C. C., and Schiller, N. O. 2005. "Can we be faithful to stress?" Poster presented at the 2nd Old World Conference in Phonology (OCP2), 20-22 January 2005 in Trømsø (Norway).
- Fromkin, V. A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47, 27-52.
- Garrett, M. F. 1975. The analysis of sentence production. In G. H. Bower (ed.) 1975, *The psychology of learning and motivation*, Vol. 9., 133-177. San Diego, CA, Academic Press.
- Goldstein, L., and Fowler, C. A. 2003. Articulatory Phonology: A phonology for public language use. In N. O. Schiller and A. S. Meyer (eds.) 2003, *Phonology and phonetics in language comprehension and production: Differences and similarities*, 159-207. Berlin: Mouton de Gruyter.
- Guenther, F. 2003. Neural control of speech movements. In N. O. Schiller and A. S. Meyer (eds.) 2003, *Phonology and phonetics in language comprehension and production: Differences and similarities*, 209-239. Berlin: Mouton de Gruyter.
- Indefrey, P., and Levelt, W. J. M. 2004. The spatial and temporal signatures of word production components. *Cognition* 92, 101-144.
- Levelt, W. J. M. 1989. *Speaking. From intention to articulation*. Cambridge, MA, MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1-75.
- Levelt, W. J. M., and Schiller, N. O. 1998. Is the syllable frame stored? [commentary] *Behavioral and Brain Sciences* 21, 520.
- Levelt, W. J. M. and 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.
- Roelofs, A. 1997. The WEAVER model of word-form encoding in speech production. *Cognition* 64, 249-284.
- Roelofs, A., and 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., Meyer, A. S., Baayen, R. H., and Levelt, W. J. M. 1996. A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3, 8-28.

- Schiller, N. O. 1997. Does syllable frequency affect production time in a delayed naming task? In G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), *Proceedings of Eurospeech '97. ESCA 5th European Conference on Speech Communication and Technology*, 2119-2122. University of Patras, Greece, WCL.
- Schiller, N. O., Meyer, A. S., and Levelt, W. J. M. 1997. The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech* 40, 103-140.
- 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.
- 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.
- Schiller, N. O., Costa, A., and Colomé, A. 2002. Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven and N. Warner (eds.) 2002, *Laboratory phonology* 7, 35-59. Berlin: Mouton de Gruyter.
- Schiller, N. O., Fikkert, P., and Levelt, C. C. 2004. Stress priming in picture naming: An SOA study. *Brain and Language* 90, 231-240.
- Schiller, N. O. 2005. Verbal self-monitoring. In A. Cutler (ed.) 2005, *Twenty-first century psycholinguistics: Four cornerstones*, 245-261. Mahwah, NJ, Lawrence Erlbaum Associates.
- Schiller, N. O. 2006. Phonology in the production of words. In K. Brown (ed.) 2006, *Encyclopedia of language and linguistics*, 545-553. Amsterdam et al., Elsevier.
- Schiller, N. O., Jansma, B. M., Peters, J., and Levelt, W. J. M. 2006. Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes* 21, 112-140.
- Schiller, N. O. in press. Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*.
- Schiller, N. O., and Costa, A. in press. The role of the syllable in phonological encoding: Evidence from masked priming? *The Mental Lexicon*.
- Shattuck-Hufnagel, S. 1979. Speech errors as evidence for a serial ordering mechanism in sentence production. In W. E. Cooper and E. C. T. Walker (eds.) 1979, *Sentence processing*, 295-342. New York, Halsted Press.
- Van Turenhout, M., Hagoort, P., and 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.
- Wheeldon, L., and Levelt, W. J. M. 1995. Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34, 311-334.
- Wheeldon, L., and Morgan, J. L. 2002. Phoneme monitoring in internal and external speech. *Language and Cognitive Processes* 17, 503-535.