Towards a New Model of Phonological Encoding

https://escholarship.org/uc/item/4cr2r9pz

Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Meyer, Paul J. A.

1994

Towards a New Model of Phonological Encoding

Drs. Paul J. A. Meijer
Max Planck Institute for Psycholinguistics
P.O. Box 310
NL-6500 AH Nijmegen
The Netherlands
`pmeyer@mpi.nl`

Abstract

The sound-form generation of a word in speech production involves the retrieval of segmental and suprasegmental information from the mental lexicon. A translation task experiment showed that the naming latencies of target items can be reduced when prime words are presented that have the same placement of the lexical stress as the target. However, this reduction will only occur when primes and targets have the same word onset. A second experiment showed that primes that have the same number of segments as the targets will cause naming facilitation compared to primes that have different numbers of segments. I have developed a new model of phonological encoding that incorporates ordered selection of the various elements. Lexical stress is chosen first, followed by information about the number of slots, the word onset, the second segment, and the other segments, until all segments have been selected. The model further employs mechanisms that allow for the retrieval of the initial segment to influence the retrieval of lexical stress. Various simulations show that the model can replicate the findings of the two experiments. Other models of phonological encoding largely neglect suprasegmental retrieval and cannot explain these results.

Speech production theories generally discern three main processing stages involved in the generation of fluent speech (Levelt, 1989). The first step in the generation of an utterance consists of Conceptualization. The speaker has to determine what she wants to express. This message then has to be Formulated. A surface structure will be created, consisting of an ordered string of lemmas grouped in syntactic chunks such as phrases. A lemma only contains an item's meaning and syntactic structure, but its sound-form has to be generated also. This phonological encoding will happen next. Finally, the articulatory program is executed during Articulation, leading to overt speech. The present paper is concerned with sound-form generation, in particular the role of suprasegmental information (such as lexical stress) in this process.

In normal conversation we utter about two words per second, but this speed can be increased to about five words (Maclay & Osgood, 1959; Deese, 1984). Phonological encoding must therefore be capable of generating the sound-forms of several words per second, allowing for partial overlaps between successive generations. A model of sound-form generation should take such overlapping generation into account. I will present a new model of phonological encoding that meets this criterion and I will apply the model to two experiments on the retrieval of lexical stress and the CV structure of a word. I will also show how other models fail to explain these effects.

My model is an example of a symbolic spreading activation model (e.g., see Dell, 1986). In these models, there are a number of nodes in a network that is often subdivided into several layers. A node has a certain activation value. A fraction of its activation spreads to other nodes via weighted links. A decay function lessens a node's activation value over time. The first layer in the network of my model contains all the lexemes. A lexeme is a singe node, pointing to an item's phonological information that is stored in the mental lexicon (see figure 1). The lexeme layer has unidirectional connections to the segmental and suprasegmental elements that form its phonological make-up. There is a stress layer, where each lexical stress pattern is stored. There is a slot layer containing the various slot numbers. Finally, there are several phoneme sets. The first set contains all the possible word-initial phonemes in the language, the second set holds all the word-second segments, etcetera. Each lexeme is only connected to its defining nodes. For instance, the lexeme *bin* is connected to the metrical node ˈσ, the slot node *3X* (meaning that 3 segments need to be selected), and the phonemes /b/, /i/, and /n/. Different lexemes are connected to the same nodes. For instance, the word *ball* is connected to the same initial segment and stress node as the item *bin*. Each node in a retrieval set has inhibitory connections to all the other nodes in the same set. An active node will spread some inhibition to the other nodes in the set, thereby trying to suppress their activation.

When phonological encoding begins, the target lexeme is activated for a certain amount of time, increasing its activation value. Its activation spreads to the connected elements. These elements are activated and will at a certain point become eligible for selection. Retrieval of the segmental and structural elements is ordered in time. First the set of lexical stress patterns delivers an element, followed by the set of slots. Then the segmental sets select elements. The set of word onsets delivers a segment first, then the set of second segments, the set of third segments, etcetera, until all segments are selected (Meyer, 1990; Meyer & Schriefers, 1991). An element can be selected when its activation has exceeded a threshold value and when its set is allowed to deliver an element. The probability of selection is a simple function of the total amount of activation in the set and the activation of the target element such that the probability increases as the target's activation value increases and the values of the other elements in the set decrease.

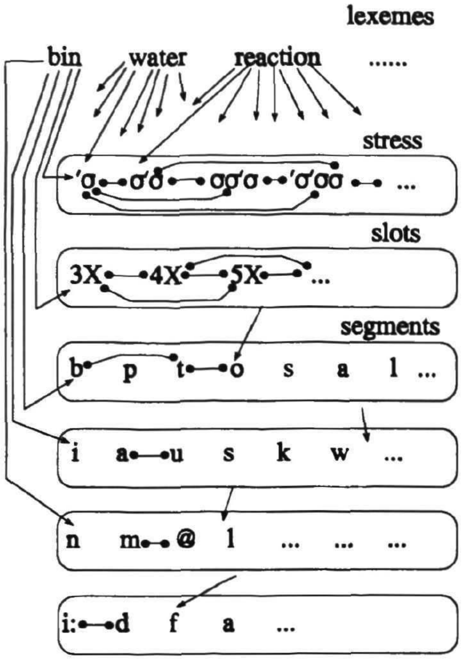

Figure 1: A description of the spreading activation network. Boxes indicate retrieval sets, arrows indicate excitatory activation flow, and lines with circular ends indicate inhibitory activation flow.

I assume that a selected element is sent to the Associator that takes each new component and associates it to the phonological representation (Levelt, 1992). The lexical stress pattern is delivered first because it forms the framework to which the segments are connected. The slot information is used by the retrieval process of the segments to determine how many segment sets must be allowed to select a segment. The Associator can use this information to know how many segments are to be expected. The segments fill the slots and a syllabified representation is formed

When a retrieval set has selected an element, the next set is allowed to deliver an element. The sets that have already given an element have finished the encoding cycle for the current lexeme and are ready to encode a new lexeme, thus allowing for overlapping word-form generations. A problem with such overlap is that both the now current lexeme and the previous lexeme whose activation is decaying are activating their connected elements. These multiple streams of activation activate several elements within a set and this might lead to competition and even selection errors. The multiple activation streams will only lead to problems if the target elements get activated to a lesser extent than non-targets. This is the case for the generation procedure of the new lexeme, because the old lexeme will initially spread more activation than the new lexeme. This disadvantage should be overcome, especially by the first retrieval set. This set forms the bottleneck of selection because it has to deliver an element earlier and in more difficult circumstances - that is, while having received less input from the lexeme - than the other sets.

The multiple activation problem is solved by a word onset inhibition mechanism and a curbing mechanism for the set of lexical stress patterns. The onset inhibition mechanism inspects the set of word onsets right before the word-form generation of a new lexeme, and suppresses all active word-initial segments. The old word onset's activation value is pushed down while the new initial segment gets activated by the new lexeme. This mechanism prevents interference by the old lexeme and maximizes the average speed of word onset retrieval. The word onsets are not as susceptible to the multiple activation problem as the set of lexical stress patterns, but the onset inhibition mechanism indirectly guides retrieval of the lexical stress through the curbing mechanism.

A curbing mechanism on stress nodes is a less powerful form of interference control that in principle applies to all stress nodes. This mechanism puts an upper limit on the amount of activation of each stress node. It is assumed that the onset inhibition mechanism cancels all activation from other than the intended source. The activation value of the most active word onset is therefore a result of the activation that was received from the current lexeme. The curbing mechanism takes this value and makes it the maximum activation value of the stress patterns. All activation that exceeds that maximum must originate from other than the intended source and should be removed.

It is possible under the curbing mechanism that the activation value of more than one stress element is on that maximum. The activation from other sources is not completely suppressed but their influence is controlled because the use of an upper limit does not allow any element to be more active than can be reasonably the case in uncontaminated retrieval. Selective inhibition (such as in the onset inhibition mechanism), affecting specific elements while leaving other segments unaffected, is less suited for the set of lexical stress patterns because this set only has a few members and an even smaller number of those - corresponding to monosyllabic and disyllabic words - are selected very frequently. Consequently, there is a high probability that the pattern that gets inhibited is exactly the one to be chosen.

Below I will present the results of two experiments, using the translation task paradigm. In these experiments, subjects read an English word and gave the Dutch translation aloud. Meanwhile a prime was presented auditorily at the same time as the stimulus (for SOA 0 ms), or after the onset of the stimulus (for positive SOAs). The subject was supposed to ignore the prime but primes often influenced the word-form generation process.

For the model I have assumed that auditory word-recognition influences the production system via the lexeme level. The acoustic signal corresponding to the prime activates several lexemes. These lexemes send activation down their encoding paths, activating their (supra)segmental elements. The model uses the mechanisms of cohort theory to simulate the recognition of the prime (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). That is, as the acoustic signal is identified, a cohort is set up containing the words matching the incoming speech. The cohort is reduced as more acoustic material gets recognized. The cohort contains all and only the words that exactly match the incoming phonemes.

In the next sections I will also discuss simulations of the two experiments. In a simulation, all trials of an experiment were run with the model (only the average results are given below). The duration of phonological encoding of an item was determined by computing at each time step the probability that the word-final segment was selected, implying that all other elements were selected at earlier time steps and that retrieval had finished. The model's estimates of the duration of phonological encoding are similar to another estimate derived by Levelt et al.'s (1991) mathematical model of picture naming. In the figures below, a constant was added to the model's reaction times, symbolizing the contribution of the processes preceding and following word-form generation to the subjects' naming latencies.

Experiment 1: Lexical Stress

Twenty-four disyllabic targets were selected for the first experiment. Twelve of these items were stressed on the first syllable. For each of these items an English translation was chosen. The English word and its translation never shared the initial segments, and had an overall minimal segmental overlap. In addition, the English and the Dutch item had a different number of syllables in most cases, and had a different placement of lexical stress in the remaining cases. Twenty-four disyllabic primes were combined with the targets in four prime conditions (the target in the examples below is ˈbajes [jail]):

1. Double Similarity. An acoustic prime (such as ˈbuidel [pouch]) and the target translation shared lexical stress and word onset.

2. Onset Similarity. The target and the prime (such as ben´zeen [benzine]) shared the word onset but had lexical stress on different syllables.

3. Metrical Similarity. The prime (such as ˈleugen [lie]) had the same lexical stress as the target but a different word onset.

4. Dissimilarity. The target and the prime (ko´meet [comet]) differed in word onset and position of lexical stress.

In each condition, the majority of targets and primes had the same CV structure. Twelve additional disyllabic targets and primes were included in the experiment and formed forty-eight filler trials. In those trials there was no overlap in stress placement and word onset. The forty-eight filler and ninety-six critical trials were presented to subjects in a pseudo-randomized order. The subject's naming response triggered a voicekey. The experiment was run at SOAs 0, 150, 300, and 450 ms. For each SOA, sixteen subjects, all native speakers of Dutch, were recruited from the Max Planck Institute for Psycholinguistics subject pool. They were paid Hfl. 8.50 for their participation.

The results are given in figure 2. The SOA manipulation did not cause differential results and the reaction times in the figure are therefore collapsed over SOA. The data show that priming of the lexical stress does not reduce naming latencies when the prime does not share the word onset with the target. If the word onset is shared, priming of the lexical stress reduces the naming latencies. In addition, naming latencies are larger when the initial segment is shared rather than different.

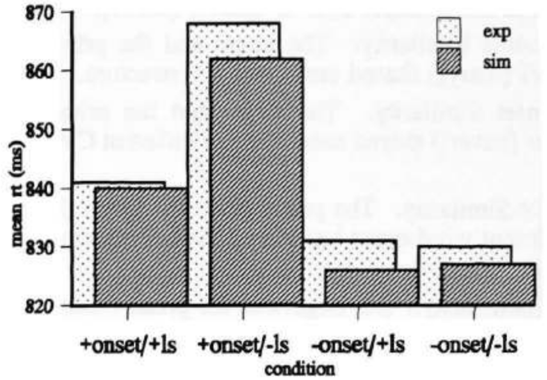

Figure 2: Experiment 2, real data and simulation data collapsed over SOA (in ms). +/- onset = same/different word onset; +/- ls = same/different lexical stress.

Figure 2 also show the results of a simulation that was run with the model (again, the data are collapsed over SOA). The simulation data closely follow the experimental findings. Metrical Similarity and Dissimilarity priming (conditions 3 and 4) produce similar naming latencies. In those conditions, the prime activates its initial segment but the activation value of this word-initial segment is suppressed when word-form generation of the target starts . The target lexeme activates another word onset that starts growing in activation. Its activation value forms the maximum value that any lexical stress node is allowed to have. In the Metrical Similarity condition (condition 3), the target lexical stress is preactivated by the prime. However, this priming is canceled by the curbing mechanism. Consequently, selection of the initial elements occurs at similar moments and a similar naming latency is obtained for both conditions.

The retrieval process is greatly affected by inhibition in the Double Similarity and Onset Similarity conditions (conditions 1 and 2). The prime's initial segment is inhibited at the start of word-form generation. This segment is shared with the target and is the intended element of selection. The facilitating effect of priming is less influential than the inhibition, leading to a slower initial rise in activation value of the target word onset than in the other two conditions. Furthermore, the curbing mechanism prevents an early rise in activation of the stress node. Once the word initial segment has overcome its inhibitory effect and starts growing in activation rapidly due to the extra priming, the stress node can start growing in activation also. Because there is no metrical priming in condition 2 (the Onset Similarity condition), the stress node grows slower in activation than in the Double Similarity condition. Consequently, the average naming latency is larger in the Onset Similarity condition than in the Double Similarity condition.

Experiment 2: Slot Retrieval

In the second experiment, twenty-six disyllabic targets were combined with primes to create the following conditions (the target in the example here is ´z̲enuw [nerve]):

1] Double Similarity. The target and the prime (such as ´z̲uivel [diary]) shared onset and CV structure.

2] Onset Similarity. The target and the prime (such as ´z̲ilver [silver]) shared onset but had different CV structures.

3] Slot Similarity. The prime (such as ´b̲izon [bison]) had a different word onset but shared the CV structure with the target.

4] Dissimilarity. The target and the prime (such as ´b̲orstel [brush]) had different word onsets and different CV structures.

Targets and primes had the same stress placement in the 104 critical trials and the word onset was shared in half of the trials. In addition, forty trials were formed in which the target and the prime always differed in word onset and stress placement. Sixteen subjects were run on SOA 150 ms and sixteen subjects were run on SOA 300 ms. The SOA manipulation again failed to cause differential effects. Therefore I present the data collapsed over SOA.

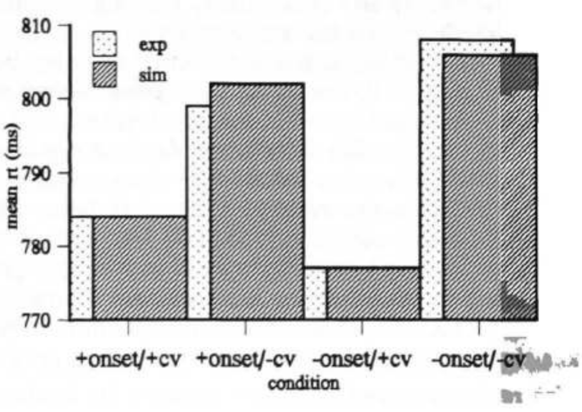

Figure 3: Experiment 2, real data and simulation data collapsed over SOA (in ms). +/- onset = same/different word onset; +/- cv = same/different CV structure.

The results of the experiment are shown in figure 3. Priming the CV structure when target and prime share lexical stress leads to fast reactions irrespective of the onset similarity between the onset and the target. Unlike the previous experiment, naming latencies are not larger when the initial segment is shared rather than different.

Figure 3 also shows the results of a simulation that was run with the model (again, the data are collapsed over SOA). The simulation data closely follow the experimental findings. Lexical stress retrieval does not contribute to differences in reaction time because retrieval is fast due to priming in all conditions. Retrieval of the initial segment also does not contribute to differences in reaction time. In conditions 1 and 2, the prime facilitates the target's word onset, but this extra activation is canceled by the inhibitory influence at the start of word-form generation (in this simulation a lower level of inhibition was assumed than in the previous simulation). In conditions 3 and 4, there is neither facilitation nor inhibition of the target's word onset. The main difference is formed by retrieval of the target CV structure. In the Onset Similarity and Dissimilarity conditions (conditions 2 and 4), a different CV structure is primed and this leads to competition in the slot layer. Priming the target CV structure in the other two conditions speeds up retrieval, leading to reduced naming latencies.

General Discussion

The research presented in this paper begins to address the role of suprasegmental information during sound-form generation. Two experiments showed that suprasegmental information forms part of the information that is stored in the mental lexicon and accessed during phonological encoding. I presented a new model about this retrieval process. In the model, selection of various elements is linearly ordered in time. First the lexical stress of a word is determined, followed by information about the total number of segments. Then the segments are retrieved from left-to-right. In addition, the retrieval of the initial segments is guided by an onset inhibition mechanism that can indirectly slow the retrieval of lexical stress. Various simulations showed that this model can replicate the reaction time results of the two experiments.

The scan and copy model by Shattuck-Hufnagel (1987, 1992) and the activation spreading model of Dell (1986, 1988) form alternative descriptions of word-form generation. Both models are heavily influenced by speech error research. Shattuck-Hufnagel's model describes how certain regularities in segment movements, leading to overt speech errors, come about. It assumes that segments fill slots in a prosodic frame, that is specified for lexical stress and (at least partially) the alternation of vowels and consonants, but it is not clear how this frame is determined. Furthermore, the time course of retrieval is unspecified in this model.

Dell's (1988) spreading activation model assumes that a prosodic frame is selected first, followed by the various segments of a word. This frame consists of a sequence of vowel, prevocalic consonant and postvocalic consonant slots. It is not clear whether the model assumes retrieval of lexical stress and whether it forms part of this frame or is retrieved separately. The model does not assume that the retrieval of frame and the segments can influence each other. As with the model of Shattuck-Hufnagel (1992), it is not the aim of Dell's model to describe the generation of several words in time, that can occur quickly and generally without errors.

Thus, previous models of phonological encoding neglect the suprasegmental aspects of retrieval. The model presented in this paper can explain the pattern of results that were found in two experiments on lexical stress and slot number. Further research will have to determine whether the effects on slot number and lexical stress hold for a larger range of items (including shorter and longer words) and whether other suprasegmental information is retrieved as well.

Acknowledgements

The research reported in this paper is part of a PhD project carried out at the Max-Planck-Institut für Psycholinguistik and at the XEROX Palo Alto Research Center. The research was carried out while I was receiving grants from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, and the Netherlands Organization for Scientific Research (NWO).

References

Deese, J. (1984). *Thought into speech: The psychology of a language*. Englewood Cliffs, NJ: Prentice-Hall.

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.

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

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1-22.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of naming. *Psychological Review, 98*, 122-142.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19-44.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D., & Welsh, A (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

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

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

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