Short Report
Orthographic Effects on Phoneme Monitoring

TON DIJKSTRA *NICI, University of Nijmegen*
ARDI ROELOFS *Max Planck Institute for Psycholinguistics and Massachusetts Institute of Technology*
STEFFEN FIEUWS *Laboratory of Experimental Psychology, University of Leuven*

Abstract A widely used task in the research on spoken word recognition is phoneme monitoring, in which subjects have to detect phonemes in spoken words. It is generally assumed that this task is performed using phonetic or phonological representations of words only. To test whether an orthographic representation of the words is employed as well, an experiment was conducted in which Dutch subjects monitored for phonemes with either a primary or secondary spelling in phonologically matched spoken words and nonwords. Phoneme monitoring times were slower when the phoneme had a secondary spelling than when it had a primary spelling. The effect was greater after than before the uniqueness point of the word, and monitoring times were faster for words than for nonwords. These findings indicate that an orthographic representation of words is engaged in phoneme monitoring.

Résumé Un test souvent utilisé dans la recherche sur la reconnaissance du langage parlé est celui du contrôle phonémique, c'est-à-dire là où on demande aux sujets de déceler les phonèmes dans les expressions orales. En général, on admet que les sujets ont recours à des représentations de mots qui ne sont que phonétiques ou phonologiques. Afin de vérifier si on avait en outre recours à des représentations orthographiques des mots, on a mené une expérience dans laquelle des sujets néerlandais recherchaient des phonèmes d'épellation primaire ou secondaire dans des mots ou barbarismes agencés selon leur phonologie. Le délai de contrôle des phonèmes était plus long lorsque le phonème était d'épellation primaire plutôt que secondaire. L'effet était plus sensible avant qu'après le point caractéristique du mot, et les délais de contrôle plus rapides pour les mots que pour les barbarismes. Ces constatations révèlent que la représentation orthographique des mots joue un rôle dans le contrôle des phonèmes.

A standard research technique in the domain of auditory word recognition is the phoneme-monitoring task, in which subjects push a response button as soon as they identify a target phoneme in a spoken word or nonword (e.g.,

Canadian Journal of Experimental Psychology, 1995, 49:2, 264-271

Foss & Swinney, 1973; Foss & Gernsbacher, 1983). In one variant of this task, known as the "generalized phoneme-monitoring procedure", the phonemes to be identified may appear at any position in the item. Research by Frauenfelder and Segui (1989; Segui & Frauenfelder, 1986) has shown that this technique can be fruitfully applied in assessing the contribution of various mental information-sources to auditory word recognition. Results indicate that the phoneme detection response may be based not only on information directly derived from the speech signal (i.e., "prelexical" information) but may also be affected lexically, that is, by stored phonological knowledge about words (e.g., Frauenfelder & Segui, 1989). Both autonomous and interactive views have been brought forward to explain how and when lexical information may intervene in the detection response.

According to the autonomous race model (Cutler & Norris, 1979), subjects performing the phoneme monitoring task respond on the basis of either a prelexical or a lexical code dependent on which code becomes available first. The prelexical code becomes available from the speech signal, whereas the lexical code is obtained through accessing a particular lexical entry. Because the lexical code can contribute to the response for words only, phoneme targets are detected faster in words than in nonwords, provided that the lexical code wins the race. In contrast, in an interactive-activation model such as TRACE (McClelland & Elman, 1986), the target phoneme is detected when it attains a criterial level of activation with respect to all other phonemes in the language. In this model, bottom-up excitation of target phonemes occurs through distinctive feature units activated by the sensory signal. Top-down excitation may also contribute to target detection when word units, activated by earlier non-target phonemes, send activation to their upcoming phonemes. Activation of the target phoneme by such feedback is especially likely when the target is located later in the word. Because spoken nonwords activate word units only partially, reaction times (RTs) will be faster for targets in words than in nonwords.

Although these two accounts differ, both assume that the phoneme monitoring task is performed using *phonetic or phonological* representations only. An observation by Frauenfelder, Segui, and Dijkstra (1990), however, suggests that this assumption may be incorrect. In a phoneme monitoring experiment conducted in French, a main effect of phoneme type was found. Slower detection times were obtained for the phoneme /k/ than for /p/ and /t/ in words compared to nonwords. The authors suggested that this effect may be attributed to the spelling of the phonemes in the words. In French, the phoneme /k/ can be realized orthographically by three different letters: "c", "k", or "q(u)". This does not hold for the /p/ and the /t/. If subjects consult an orthographic representation of the word in addition to phonetic or phonological codes, the variety of orthographic representations for /k/ as opposed to /p/ or /t/ may have affected the monitoring response. Similar effects of orthography have been reported for other auditory task situations, such as syllable detection (see Dupoux & Mehler, 1992) and rhyme monitoring (Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981).

In the current paper, we report a direct test of the hypothesis that orthographic codes may affect the RTs in phoneme monitoring by examining the monitoring times for target phonemes that have more than one spelling in Dutch words. For example, in Dutch (as in English), the primary spelling of the phoneme /k/ is "k" and a secondary spelling is "c". Similarly, the primary spellings of /s/ and /t/ are, respectively, "s" and "t", and secondary spellings are, respectively, "c" and "d". If orthographic codes become available during speech processing and are consulted in phoneme monitoring, secondary spellings may lead to interference effects because they are not congruent with the canonical spelling of the phonemes. In the Dutch materials used, whether a phoneme has a primary or secondary spelling in a word could only be determined on the basis of the identity of that word, thus requiring lexical information.

METHOD

Subjects

Thirty University undergraduates, all native speakers of Dutch, were individually tested.

Materials and design

The target phonemes were the voiceless stops /k/ and /t/ and the voiceless fricative /s/. The spoken carrier items consisted of 60 Dutch words with a primary spelling of the target phoneme (Primary Spelling), 60 words with a secondary spelling of the target phoneme (Secondary Spelling), and 60 nonwords (Nonword).[1] The words were nouns consisting of two or more syllables. First, the secondary spelled words were selected from the CELEX data base (Baayen, Piepenbrock, & van Rijn, 1993) by combining orthographic and phonological search criteria (e.g., the orthographic form of the item should contain a "c", while this letter was pronounced as /k/). Next, primary spelled words were chosen to match the secondary spelled words. On average, the words were of the same frequency and of the same length in terms of number of phonemes. The local phonological context of the target phoneme in the words was kept as similar as possible. In addition, an attempt was made to match the items with respect to global word structure (i.e., CV structure and stress pattern). Finally, nonwords were derived from the primary spelled words by replacing two to four phonemes by phonemes differing in one or two distinctive features.

The target phonemes were located either before or after the uniqueness

1 These materials are available from the author.

TABLE 1
Examples of stimulus items and (in parentheses) average distance in number of phonemes separating the target phonemes from Uniqueness Point/Nonword Point

POSITION	SPELLING		
	Primary	Secondary	Nonword
Before UP/NWP	kabouter [-2.7]	cabaret [-2.7]	kadoupel [-2.4]
After UP/NWP	paprika [+0.9]	replica [+1.1]	taplika [+1.9]

point (UP) of the target-bearing words. Following Frauenfelder et al. (1990), the UP was defined as the point in a word at which its initial part is shared by no other morphologically unrelated word in a phonetic dictionary. Targets in the nonwords were located at the same serial position as in the matched primary spelled words, but were before or after the nonword point (NWP). The nonword point was defined as that point moving from onset to offset at which the item becomes a nonword. Examples of the word and nonword items can be found in Table 1.

For the Before-UP condition not enough words could be found with a secondary spelling of the phoneme /t/. Therefore, in this condition a number of compounds were included (e.g., *badpak* (*swimming suit*)), in which the "d" is pronounced as /t/). For these items, the target phoneme was located just before or right at the point in the compound where the first part (e.g., *bad*) becomes uniquely identifiable. If the first lexical part of the compound is retrieved before the whole, a lexical effect on phoneme monitoring might already be expected in the Before-UP condition, not only in the After-UP condition. Thus, for these items, the Before-UP and After-UP conditions become similar, reducing a possible interaction between spelling (primary or secondary) and target position in the word.

In addition to the 60 test items for each target phoneme (20 words with Primary Spelling, 20 words with Secondary Spelling, and 20 nonwords), another 15 items (10 words and 5 nonwords) were included as filler items. These were bisyllabic items in which the target phoneme was located somewhere in the middle. An additional 75 filler items (50 words and 25 nonwords) did not contain the phoneme target.

By combining test and filler items, a list of 150 items was made for each target phoneme. In addition, a practice list was created with the target phoneme /p/. This list contained 20 items, and was similar in structure to the three experimental lists.

All lists were digitally recorded (with a sample frequency of 22 kHz) in a soundproof room by a female native speaker of Dutch, using the Farallon MacRecorder Sound System on a Macintosh SE/30 computer in combination with a Sennheiser microphone.

TABLE 2
Mean RTs (ms) and (in parentheses) error percentages for POSITION by SPELLING collapsed across target phonemes

POSITION	SPELLING			
	Primary	Secondary	Nonword	Mean
Before UP/NWP	620 (2.0)	641 (3.4)	675 (4.3)	645 (3.3)
After UP/NWP	359 (2.4)	415 (5.8)	468 (6.5)	414 (4.9)
Mean	489 (2.2)	528 (4.6)	571 (5.4)	529 (4.1)

Procedure

Subjects were asked to make a speeded detection response to the target phonemes occurring anywhere in spoken words and nonwords by pressing a button with their preferred hand. Targets were specified auditorily to the subjects before each list by phrases like "listen now for the sound /s/ as in Simon".[2] Furthermore, subjects were instructed to listen to the items as they were spoken, disregarding the speaker's dialect, their own pronunciation, or spelling. This remark was intended to direct the subject's attention to the auditory signal, at the same time making clear that spelling differences were of no consequence for performing the task.

The experiment was run on an Apple Macintosh SE/30 computer, using the Experimental Control System (Sikuta, MacWhinney, & Clynes, 1990). During testing, each experimental item was preceded by a 200 ms (1 kHz) warning signal and a silence period of 1 s.

RTs were measured from burst onset for the stop targets and from frication onset for the fricatives. The onsets were determined for each item before the experiment both visually and auditorily by two independent judges. Each presented item was followed by an inter-trial interval of 2 s. An experimental session lasted about 40 minutes.

RESULTS

Mean RTs were computed for each subject and each experimental item. For word-initial targets, all responses with a latency less than 150 ms were considered to be errors and excluded from further analysis. For all items, responses with a latency greater than 1500 ms were treated as errors as well.

First, to determine the effect of spelling, the RTs were submitted to by-subject and by-item ANOVAS with SPELLING (Primary or Secondary), POSITION (Before-UP or After-UP) and TARGET (/k/, /s/, or /t/) as main factors. The ANOVA yielded main effects for SPELLING [$F_1(1,29) = 36.86$, $MS_e = 3715$, $p < .001$; $F_2(1,9) = 10.25$, $MS_e = 3892$, $p < .01$], POSITION [$F_1(1,29) = 628.53$,

2 Subjects may have set up their internal target representation on the basis of this example of a primary spelled word. The point of this study is, however, to assess whether orthography has some influence.

$MS_e = 8491$, $p < .001$; $F_2(1,9) = 763.52$, $MS_e = 2319$, $p < .001$] and TARGET [$F_1(2,58) = 17.31$, $MS_e = 15742$, $p < .001$; $F_2(2,18) = 21.72$, $MS_e = 4575$, $p < .001$]. More importantly, the effect of spelling depended on the position of the target, as a significant interaction between SPELLING and POSITION showed [$F_1(1,29) = 8.56$, $MS_e = 3059$, $p < .01$; $F_2(1,9) = 6.82$, $MS_e = 1240$, $p < .05$]. As expected from the deviant make-up of the /t/ items (see Materials), there was an interaction between TARGET and POSITION [$F_1(2,58) = 8.79$, $MS_e = 5583$, $p < .001$; $F_2(2,18) = 3.65$, $MS_e = 3842$, $p < .05$]. However, these items did not affect the basic effects, because there was neither an interaction between TARGET and SPELLING [$F_1(2,58) = 2.94$, $MS_e = 3892$, $p > .05$; $F_2(2,18) < 1$], nor between TARGET, SPELLING and POSITION [$F_1(2,58) = 1.20$, $MS_e = 3867$, $p > .3$; $F_2(2,18) < 1$].

Second, to obtain independent evidence for lexical mediation of the spelling effect, an ANOVA was run with LEXICAL STATUS (Primary words or Nonwords) and POSITION (Before-UP/NWP or After-UP/NWP) as main factors.[3] The ANOVA yielded main effects for both LEXICAL STATUS [$F_1(1,29) = 186.48$, $MS_e = 3236$, $p < .001$; $F_2(1,9) = 101.83$, $MS_e = 1973$, $p < .001$] and POSITION [$F_1(1,29) = 481.23$, $MS_e = 10217$, $p < .001$; $F_2(1,9) = 767.03$, $MS_e = 2137$, $p < .001$]. Furthermore, a significant interaction was found between LEXICAL STATUS and POSITION [$F_1(1,29) = 31.79$, $MS_e = 2038$, $p < .001$; $F_2(1,9) = 25.40$, $MS_e = 797$, $p < .001$]. RT-differences between Primary Spelled words and Nonwords increased when the target phoneme was positioned later in the item.

Discussion

In the experiment, spelling differences between primary and secondary spelled Dutch words resulted in RT-differences for target phonemes dependent on their location in the item. This result was expected because the spelling difference was lexically determined, and lexical effects are expected to arise predominantly after the UP (cf. Frauenfelder et al., 1990). The sensitivity of the experiment for lexical effects was independently attested by an increasing difference between the RTs for primary spelled words and matched nonwords going from the before-UP/NWP to the after-UP/NWP target position. Following Frauenfelder et al. (1990), this RT-difference can be considered as a measure of the lexical contribution to the phoneme-detection process.

In short, the experiment showed an effect of lexically mediated orthography on phoneme monitoring. This task is therefore not performed using phonetic or phonological representations of words only. An orthographic representation of the words appears to be employed as well. To explain this influence of orthography, the existing acounts of the representations and processes

3 The Secondary Spelled words were not included in the comparison because their recognition could be affected by inhibition effects due to their deviating spelling.

involved in the phoneme monitoring task have to be modified.

A first possible explanation is that activated lexical orthographic codes facilitate the monitoring response by feeding back activation to their associated phonemes. Thus, the RT-pattern observed in our experiment could be based on activation feedback from the lexical orthographic to the sublexical phoneme level, either through a direct link between these two levels, or through a mediating sublexical orthographic level (e.g., that of graphemes; Dijkstra, Frauenfelder, & Schreuder, 1993).

Second, as Dupoux and Mehler (1992) suggested, simultaneous use of different types of codes (including orthographic codes) could also be a consequence of the architecture of the human language processing system in relation to task requirements. A non-modular decision mechanism might perform a matching of stimulus and target representation across all active codes (irrespective of their origin), followed by a weighing of the evidence to arrive at a single decision.

Third, subjects performing the phoneme monitoring task could keep in mind an orthographic representation of the target phoneme in addition to a phonological one, even when this would be detrimental to task performance. One reason for doing so could be that the simultaneous use of different codes provides them with a more stable representation of the target in Working Memory over time (Conrad, 1972).

Whichever of these (or still other) possibilities proves to be correct, both models introduced earlier need to be adapted. The TRACE model (McClelland & Elman, 1986) must assume that there is a fast spreading of activation from a lexical representation to a graphemic representation that can affect the phoneme detection response. The race model (Cutler & Norris, 1979) must adopt the assumption that auditory lexical access also involves the retrieval of orthographic representations, which can contribute to the detection decision in addition to phonological representations.

To conclude, researchers should seriously consider the possibility that both phonological and orthographic codes become quickly available during the perception of a spoken word, and that phoneme monitoring may be affected by each of these sources of information. Given the size of the effects (on the order of 35 ms), they should consider controlling the orthographic characteristics of their auditory stimuli to reduce noise in their data and especially to avoid artefactual results.

We are indebted to Colin MacLeod, Len Katz, and two anonymous reviewers for useful comments on the text and to the students of the EPO-course for their help in running the experiment.

Address correspondence to Ton Dijkstra, *NICI*, University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands (e-mail: dijkstra@nici.kun.nl).

References

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database.* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Conrad, R. (1972). Speech and reading. In J.F. Kavanagh and I. Mattingly (Eds.), *Language by ear and by eye* (pp. 205-240). Cambridge, MA: MIT Press.

Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In W.E. Cooper and E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113-134). Hillsdale, NJ: Erlbaum.

Dijkstra, A., Frauenfelder, U.H., & Schreuder, R. (1993). Bidirectional grapheme-phoneme activation in a bimodal detection task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 931-950.

Donnenwerth-Nolan, S., Tanenhaus, M.K., & Seidenberg, M.S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 170-180.

Dupoux, E. & Mehler, J. (1992). Unifying awareness and on-line studies of speech: A tentative framework. In J. Alegria, D. Holender, J. Junca de Morais, and M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 59-75). Amsterdam: Elsevier Science Publishers.

Foss, D.J. & Gernsbacher, M.A. (1983). Cracking the Dual Code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior, 22*, 609-632.

Foss, D.J., & Swinney, D.A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *Journal of Verbal Learning and Verbal Behavior, 12*, 246-257.

Frauenfelder, U.H. & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition, 17*, 134-140.

Frauenfelder, U.H., Segui, J., & Dijkstra, A. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 1, 77-91.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Segui, J. & Frauenfelder, U.H. (1986). The effect of lexical constraints upon speech perception. In F. Klix and H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 795-808). Amsterdam: Elsevier Science Publishers.

Sikuta, G., MacWhinney, B., & Clynes, D. (1990). *Documentation Experimental Control System.* Pittsburgh, PA: Carnegie Mellon University, Psychology Department.

Date of acceptance: March 3, 1994