EVIDENCE FOR DIRECT LEXICAL ACCESS FROM RESPONSE TIME EXPERIMENTS

Kari Suomi
Deparment of Finnish, Saami and Logopedics, University of Oulu, Finland

ABSTRACT

Monitoring experiments are reported that compared response times (RTs) to three target types in Finnish: whole words, word-final syllables, and word-final phonemes. Care was taken to ensure that the targets could not be responded to on the basis of only partial analysis of the stimuli. Throughout, whole words were detected faster than coterminous syllables and phonemes, suggesting that words are not recognised through intermediate phonemic or syllabic representations.

INTRODUCTION

Both the idea of lexical access through compulsory intermediate phonemic and/or syllabic representations and the idea of direct access postulating some other kind of sound representation to mediate between the sensory input and the lexicon have support in the literature on spoken word recognition (see e.g. [1], [2]). Measurement of RTs to monitoring targets, a paradigm in which shorter RTs are interpreted to indicate earlier on-line processing, is a tool that, potentially at least, could be profitably used to inquire whether lexical access is direct or not. To my knowledge, however, the paradigm has never been used to explicitly address this question. The virtual lack of reported comparisons of RTs to words and their phonological constituents in the last two decades seems to be due, in part, to a preconception among some researchers who use this paradigm to study lexical access that there is no alternative to access involving intermediate phonological units and representations. The conclusions of McNeill & Lindig [3] seem to have been influential in shaping the preconception, these authors claiming that that minimum RT in target monitoring experiments occurs whenever the linguistic level of the target and the search list is the same, and since the level where the target and the list match is entirely determined by the experimental design, it is no possible that monitoring experiments can reveal the perceptual units of speech. However, these widely cited conclusions are not warranted by the authors' experiments that have been shown to suffer from a number of methodological weaknesses that render the results highly unreliable. Thus, due to the way in which the stimuli were constructed, subjects were able to base their responses on only the initial portions of the target-matching stretches of the stimuli (see [4]), they were in fact urged to do so, and the experiments included conditions involving what the authors call *downward search* in which subjects were required e.g. to detect target sentences and target words in search lists consisting of syllables and phonemes (sic!) (see [5], which also contains a more detailed account of the present experiments).

In brief, there is no evidence from target monitoring studies that would force the conclusion that phonemes and/or syllables must be identified before a word is accessed and recognised. Prompted by a desire to test the DAPHO model [6] that postulates one version of direct access, the present experiments were designed to measure RTs to words, syllables and phonemes under as comparable bottom-up conditions as possible. Each target-bearing or target-constituting stimulus word contained all three target types. E.g., RTs were measured to each of the targets "PALKKI", "KI" and "I" in the stimulus word *palkki*. In a given stimulus word, the three target types were all coterminous, and thus the time course of how subjects were exposed to the distinguishing auditory information in the stimulus was exactly the same for each target type. And since RTs were always measured from exactly the same temporal location in a given experimental word for each target type, any systematic differences observed in RTs to these targets must be due to differences in the central processing of simultaneously available peripheral input.

PROCEDURE

Experiment 1 is described in some detail below, but for experiments 2 and 3 only major deviations from the procedure of experiment 1 are indicated.

Experiment 1

In experiment 1, the target-carrying stimuli were a set of 36 disyllabic words, each occurring in a list containing from three to six words. In addition, subjects were presented 10 practice word lists at the beginning of the test the responses to which were ignored, and also, dispersed among the experimental lists, 18 no-response distractor lists and 9 filler lists. All subjects heard exactly the same stimulus material. The target-carrying stimuli were chosen in 12 triplets so that, within each triplet, all three words had a phonemically identical second syllable, and the first syllable of each word had the same general structure in terms of the C and V class affiliation of its segments. A further requirement for a word to be included in a triplet was that at least one further familiar word must exist that diverges from the experimental word with respect the the final phoneme alone, to guarantee that the uniqueness point of the experimental words was not reached until the portion corresponding to the final phoneme.

Each word in each of such highly controlled triplets functioned as carrier of each of the target types Word, Syllable and Phoneme but in three different, rotated target conditions. The target conditions were rotated in such a way that, for a given carrier word, subjects in one condition were given a word target, those in a second condition a syllable target, and those in a third condition a phoneme target. Target assignments were balanced across the conditions so that each triplet yielded three instances of RTs to each target type. Consequently, the RTs to the three target types to be reported were obtained using exactly the same set of words.

In the no-response distractor lists the Word, Syllable and Phoneme targets were similarly rotated, but the Word target specified for a list did not occur in that list. Instead, the list contained a word that deviated from the specified Word target by the last phoneme only. E.g., one such list had the specified targets "HELMA", "MA" and "A", and the list consisted of the words *kuori kuusi potti rove helmi tossu*, in which the penultimate word is the intended distractor. Thus in each no-response list, the distractor conditions were exactly the same for the three target types, and the appearance of finally-diverging distractors in the Word target condition should induce subjects to respond only after a complete analysis of the stimulus words. Subjects should not respond to the distractor lists if they were reacting accurately, and therefore subject reacting to more than a predetermined number of such lists were discarded. 27 of the 30 tested subjects were accepted.

Individual subjects were seated before a computer terminal, and the lists were presented through earphones. Subjects were told that they would hear word lists and that their task was to monitor for whole-word targets, targets consisting of a consonant-vowel sequence, or vowel targets, and they were instructed to press the space key as soon as they were certain that they had heard the target valid for a given word list. Before each new word list, an alert beep was sounded and the (fully phonemic) written target specification appeared on the screen where it stayed 2.5 seconds, after which the list was heard.

For each target-carrying word, the raw RTs were measured from the estimated onset of the final vowel, but the raw values were adjusted to give RTs from the common acoustic end point of the three target types.

Experiment 2

In experiment 1 vowel-final disyllabic real words were used as stimuli, whereas in experiment 2 phonologically well-formed nonsense items were used, to allow for more variable yet native-like structural patterns. Half of the items were disyllabic, half trisyllabic, and within each group, half were vowel-final, half consonant-final. Nonsense items are also insensitive to word frequency effects which were not completely controlled in experiment 1. All subjects again heard exactly the same stimulus material, and the 48 stimuli

carrying the three target types were chosen in 16 triplets following the same criteria as in experiment 1. Subjects were instructed to treat the whole-item nonsense targets as novel words, e.g. as names of new products. Otherwise, the procedure was as in experiment 1, including the use of finally-diverging distractor stimuli and rotation of the target types. There were 24 accepted subjects.

Experiment 3

Experiments 1 and 2 required detection of targets in lists of real and nonsense items that were separated by pauses, and such lists may favour the detection of whole-word targets because the input has already been segmented into stretches that correspond exactly to the target units, whereas the onsets of the phoneme and syllable targets have to be located in coarticulated speech within the stimuli. Therefore, experiment 3 was conducted in which all target types had to be segmented from both preceding and following continuous speech. Experiment 3 contained the same target-carrying words as experiment 1, but this time embedded in short sentences that were semantically fully neutral with respect to the probability of occurrence of either the specified Word target or its implicit, finally-diverging lexical competitor(s). 21 subjects were accepted.

RESULTS

The mean RTs observed in experiments 1-3 are shown in Tables 1-3.

Table 1. Mean RTs (in ms) to detect the target types Word, Syllable and Phoneme in disyllabic vowel-final real words in experiment 1.

Wrd	Syl	Pho	Mean
173	271	314	253

In experiment 1, whole words were detected about 100 ms faster than final syllables, which were in turn detected about 40 ms faster than final phonemes; both of these differences were significant.

Experiment 2 replicated the major results of experiment 1. Thus while responses to consonantal Phoneme targets and consonant-final Syllable and Word targets were faster than responses to vocalic or vowel-final targets, RTs to Word targets were again faster than those to the phonological targets irrespective of the type of final segment, and final syllables were again detected faster than final phonemes (all these differences were significant).

Table 2. Mean RTs to detect the target types Word, Syllable and Phoneme as a function of target-final segment class in nonsense items in experiment 2.

Final segment class	Target type			
	Wrd	Syl	Pho	Mean
Vowel	175	301	344	273
Consonant	156	260	285	234
Mean	165	280	314	253

The results of experiment 3 replicate the major finding: Whole words were detected faster than final syllables and final phonemes, even when also whole words had to be segmented from continuous speech. In experiment 3 all mean RTs are on average about 70 ms longer than in experiment 1; this increase may be due to a greater difficulty of performing target monitoring in material that is semantically coherent.

Table 3. Mean RTs to detect the target types Word, Syllable and Phoneme in disyllabic vowel-final real words in experiment 3.

Wrd	Syl	Pho	Mean
263	340	370	324

DISCUSSION

Against the background that both lexical access through necessary intermediate phonemic and/or syllabic representations and direct lexical access using some alternative sound representations find ample support among researchers of spoken word recognition, the present

series of experiments set out to investigate response times to three acoustically coterminous target types, namely whole words, word-final syllables, and word-final phonemes, in an attempt to distinguish between the rival broad views of lexical access. The results indicate that whole words were detected before their final syllables and final phonemes even when the words were lexically non-unique prior to the last phoneme, and when the possibility of responding on the basis of guessing was eliminated by stringent distractor conditions. The whole-word advantage was observed in experiments whose materials jointly contained variable and phonotactically representative targets of each type, it was observed with real words as well as pseudowords, with words separated from others by pauses and semantic incohesiveness, and with words in connected speech in meaningful sentences. Differences in the manner in which response times and other temporal data on on-line speech behaviour have been measured and reported make it impossible to compare the present results with previous ones, but the temporal distances here observed between the end of a word and the detection of that word are not inconsistent with the intuitive immediateness with which words seem to be recognised outside the laboratory.

On the assumption that shorter RTs reflect earlier processing, I interpret the results as support for the idea that lexical access and word recognition are direct in the sense that they do not involve compulsory intermediate levels of representation in terms of phonemes or syllables. There is no direct evidence from monitoring studies against direct access in any language, e. g. findings to the effect that RTs to whole words are longer than those to their constituent phonemes or syllables, and consequently there is no principled reason for dismissing the present results as specific to Finnish alone (which nevertheless remains a testable possibility).

A counterargument against the above interpretation that I have come across is that the results are most probably irrelevant to the question of the nature of the sound representation involved in lexical access because word detection involves identifying a familiar unit whereas syllable or phoneme detection does not, and therefore, even if phonological units were used implicitly to identify words, it does not necessarily follow that they would be detected faster than words, in an explicit detection task. If this argument is taken as a sufficient explanation of the observed detection advantage of whole-word targets, then obviously the conclusion follows that it is *a priori* impossible to distinguish between lexical access through intermediate units and direct access using the target monitoring paradigm, because the familiar-unfamiliar distinction can always be invoked to annihilate any data that seemingly support direct access. But if phonemes and/or syllables are regularly and compulsorily identified prior to word recognition, can they really be characterised as unfamiliar units, especially in comparison to pseudowords as used in experiment 2? And if a familiar-unfamiliar effect is operative, does its magnitude fully account for the observed word-advantage in response times?

REFERENCES

[1] Pisoni, D. & Luce, P. (1987), "Acoustic-phonetic representations in word recognition". In *Spoken word recognition* (U. Frauenfelder & L. Tyler, editors), pp. 21-52. Cambridge, MA: The MIT Press.

[2] Klatt, D. (1979), "Speech perception: A model of acoustic-phonetic analysis and lexical access", *Journal of Phonetics*, vol. 7, pp. 279-312.

[3] McNeill, D. & Lindig, K. (1973), "The perceptual reality of phonemes, syllables, words and sentences", *Journal of Verbal Learning and Verbal Behavior*, vol. 12, pp. 419-430.

[4] Norris, D. & Cutler, A. (1988), "The relative accessibility of phonemes and syllables", *Perception & Psychophysics*, vol. 43, pp. 541-550.

[5] Suomi, K. (submitted), "Evidence of word recognition using direct lexical access without phonemes or syllables", a paper submitted for publication.

[6] Suomi, K. (1993), "An outline of a developmental model of adult phonological organization and behaviour", *Journal of Phonetics*, vol. 21, pp. 29-60.