Word Recognition without Phonemes or Syllables?

Kari Suomi
Department of Logopedics and Phonetics, University of Oulu
P.O.B 111, 90571 Oulu, Finland

ABSTRACT

Two experiments are reported that were intended to distinguish between two conceptions of how spoken words are recognised. According to one view words in the lexicon are accessed through intermediate representations in terms of phonemes and/or syllables, whereas an alternative view assumes that lexical access is direct in the sense that it involves no discrete units between the acoustic input and the lexicon but, instead, matching of the input against lexically stored continuous representations. The experiments compared response times to three acoustically coterminous target types in Finnish, namely whole-word targets, word-final syllable targets, and word-final phoneme targets, the words constituting or bearing the targets being non-unique until the final phoneme. Word targets were detected about 100 ms faster than syllable and phoneme targets in both experiments, strongly suggesting that the representations used in identifying whole words did not contain phonemes or syllables. Instead, the results support the idea of direct lexical access.

INTRODUCTION

The DAPHO model (Suomi, 1993) proposed that phonemes and syllables are not involved in word recognition as intermediate units between the auditory input and the lexicon. Instead, DAPHO claims that lexical access is direct in the sense that it involves matching of the auditory input against word-size holistic auditory prototypes, without intermediate discrete units. According to this proposition, phonemes and so-called lexical syllables are essentially units of the planning of speech production. The model further claims that, in listening to speech, the production units are not used in the normal, communication-oriented mode, and that they are employed only in the special phonetic mode in which the listener's attention is directed to the phonic medium rather than the linguistic message being transmitted (as, typically, in phoneme and syllable monitoring tasks). These claims amount to postulating a lexicon with two qualitatively different sound representations for each lexical item, an articulatory-gestural representation consisting of lexical syllables and phonemes, used mainly in production, and another representation in terms of holistic auditory prototypes, used only in recognition.

But a very strong preconception prevails against the feasibility of direct lexical access. For example, some scholars argue that intermediate processing units are inescapable in lexical access in order to avoid cumbersome exhaustive search of all lexical representations and to enable, instead, a more efficient preliminary classification using a relatively small set of units in terms of which lexical items are arranged (Norris & Cutler, 1988). Others assume that the extensive acoustic variability in speech rules out direct mapping and forces the input to be transformed into a prelexical representation that must be both abstract and discrete (Otake, Hatano, Cutler & Mehler, 1993). And still others suggest that intermediate units "are, in fact, *tacitly* assumed by all contemporary models of word recognition. Without this assumption, it would not be possible to recover the internal structure of words and access their meanings" (Pisoni & Luce, 1987, p. 33; emphasis in the original). But alternative solutions to avoiding exhaustive lexical search and to accounting for constant perception in the face of acoustic variability are conceivable, and what theory-independent grounds are there for making the presupposition that *both* recovery of internal structure *and* access to meaning take place every time a word is recognised?

Measurement of response times (RTs) to phonological targets in monitoring tasks is an extensively used tool in exploring the manner in which words are extracted from the acoustic signal, with shorter RTs interpreted to indicate earlier processing. But these studies too have almost invariably taken it for granted that intermediate phonological units must be detected in the input before contact with the lexicon, and in hardly any experiments have attempts been made to actually test whether or not phonological targets are detected sooner than whole-word targets. Consequently, there is no evidence in the results of target monitoring studies that would force the conclusion that detection of phonological units must precede word recognition, given the alternative of direct lexical access, an alternative that has never been experimentally refuted. But at the same time, and for the same reasons, evidence in support of direct lexical access is similarly lacking in the target monitoring paradigm. A true test of the relative accessibility of words and phonological units must involve direct comparison of response times to these entities under as comparable and controlled conditions as possible, and this is what the present experiments in Finnish attempted to do.

In both experiments, RTs were measured to three types of coterminous targets, namely (real or meaningless) whole-word targets, word final syllable targets, and word final phoneme targets, each target-bearing stimulus word containing each of the three target types in three experimental conditions. E.g., RTs were measured to each of the targets "PALKKI", "KI" and "I" in the stimulus word *palkki.* Since RTs were always measured relative to the common end point of each target type, any systematic differences observed in RTs to these targets must be due to differences in central processing. When given the written monitoring target "KI", subjects activate the syllabic representation corresponding to this specification, and when given the target "I", they activate the representation of the phoneme /i/. But when given a whole-word monitoring target like "PALKKI", subjects locate that word in their lexicon and activate its recognition-oriented lexical sound representation, whether or not this is the only lexical sound representation belonging to that word. Now if the recognition-oriented lexical sound representation is a phonemic code and if, accordingly, constituent phonemes have to be identified before lexical access and before a spoken word can be unambiguously recognised against a finally-diverging competitor, then detection of a stimulus word should take longer than detection of its final phoneme. And if, instead, lexical access and storage are in terms of syllables and if, therefore, constituent syllables have to be recognised before the whole word can be identified, then it should, by the same reasoning, take longer to detect a word than its final syllable. But if, finally, the recognition-oriented lexical sound storage involves a representation that does not contain phonemes or syllables, a possibility not envisaged in the target monitoring literature, then DAPHO predicts that a stimulus word, which is represented by an auditory prototype in the lexicon, is detected faster than its final syllable or final phoneme, and that there is no difference in the detection times of these latter units (for the motivation of these predictions see Suomi, 1993).

EXPERIMENT 1

The stimuli carrying or constituting the targets were a set of 36 disyllabic Finnish words, each occurring in a list containing from three to six words. In addition, subjects were presented 10 practice word lists and 18 no-response 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. For example, one such triplet consisted of the words *pol-la, hel-la, pul-la.* 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 final phoneme (thus Finnish also has the words *pol-le, hel-lä, pul-lo).* Because of these stringent requirements all target-bearing words were disyllabic and ended in a vowel, since Finnish phonotactics makes it difficult to find suitable minimal pairs that differ with respect to the final consonant, or with respect to the final vowel in words with more than two syllables. 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. For example, the list containing the experimental word *hella* had the rotating targets "HELLA", "LA" and "A". In the remaining two words of the triplet, the target assignments were different across the conditions so that each triplet yielded three instances of each target type to be responded to. In the 18 no-response foil lists the Word, Syllable and Phoneme targets were rotated as in the response lists, but the Word target specified for a list did not occur in that list. Instead, the list contained, in similarly variable serial positions, 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 spoken words *kuori kuusi potti rove helmi tossu,* in which the penultimate word is the intended foil item. Thus whatever expectations subjects might entertain about the identity of partially analysed words, the expectations should affect all three targets similarly. In particular, the finally-diverging foils in the Word target condition should teach subjects to refrain from responding on the basis of initially matching information and to induce them, instead, to respond only after a complete analysis of the stimulus words. The 27 subjects were assigned to three rotated target conditions of the same stimulus material, with 9 subjects in each condition. The experiment was administered using EASYST, a DOS-compatible RT measurement system constructed by Einar Meister (see Meister & Suomi, 1993). Subjects were given the target valid for a list in written form on the computer screen, after which they heard the spoken list through headphones, and their task was to press a button as soon as they were sure that they had identified the target. The results are shown in Table 1.

Table 1. *Mean response times (msec) to detect the coterminous target types Word, Syllable and Phoneme in disyllabic vowel-final real words, as measured from the common end point.*

Word	Syllable	Phoneme	Mean RT
173	271	314	253

Analyses of variance showed that the mean of the Word target was significantly different from the mean of the Phoneme target and from the mean of the Syllable target, and that the difference between the means of the Phoneme target and the Syllable target was not significant. The results are thus grossly at variance with predictions based on the assumption that lexical access and word recognition involve prior identification of phonemes and/or syllables. In contrast, the results agree with the general idea of direct lexical access without intermediate phonological units, and they are fully in agreement with the predictions of DAPHO to the effect that a word is detected faster than its final syllable or final phoneme, and that there is no difference in the detection times of the coterminous phonological units. Whole-word targets were detected no less than 100 ms faster than the phonological targets; no previous results from corresponding experiments are available with which this figure could be compared.

EXPERIMENT 2

Experiment 2 involved 48 target-bearing phonologically and phonotactically well-formed nonsense items which permitted more variable structural patterns than the real words in experiment 1; a test with non-sense materials was also desirable to exclude the possibility that the results of experiment 1, in which word frequency was not controlled, were due to lexical effects (word frequency was not controlled in experiment 1). Thus half of the items were disyllabic, half trisyllabic, and within each group of items, half were vowel-final, half consonant-final. The target types Word, Syllable and Phoneme were rotated in three conditions as in the first experiment, and subjects were instructed to treat the whole-item nonsense targets as novel words, e.g. as names of new products. All items in all lists were pseudowords structurally similar to the target-bearing items. In all other relevant aspects, the two experiments were essentially similar. The results are shown in Table 2 as a function of the two significant main effects, target type and of final segment type, which did not interact.

Table 2. *Mean response times (msec) to detect the coterminous target types Word, Syllable and Phoneme as a function of target-final segment class in nonsense items.*

	Word	Syllable	Phoneme	Mean RT
V-final	175	301	344	273
C-final	156	260	285	234
Mean RT	165	280	314	253

As in the first experiment, the mean of the Word target was significantly different from the means of the phonological targets, but the difference of the means of the phonological targets was not significant. The results of experiment 1 were thus replicated in experiment 2: responses to consonantal phoneme targets and consonant-final syllable and whole-word targets were considerably faster than responses to vocalic or vowel-final targets, but RTs to Word targets were again faster than those to the phonological targets irrespective of the type of final segment.

DISCUSSION

The results indicate undisputably that, at lest in Finnish, whole words can be recognised well before their constituent final syllables and final phonemes even when the words are lexically non-unique up to the last but one phoneme, and when the possibilities of guessing on the basis of only partial acoustic information have been further eliminated by stringent foil conditions; this experimental design has not been used in other languages. The results support the claim that lexical access and word recognition are direct in the sense of involving no intermediate levels of representation in terms of phonemes or syllables, and they are at variance with all models of lexical access that maintain that phonemes and/or syllables have to be identified before a spoken word can be recognised. In contrast, the results fully agree with the prediction of DAPHO that words should be detected earlier than their coterminous phonemes and syllables, and that the latter units should be detected simultaneously. These findings and conclusions will be reported in more detail elsewhere.

REFERENCES

Meister, E. & K. Suomi. 1993. 'EASYST - A computer system for reaction time measurements'. In Iivonen, A. & R. Aulanko (eds.), *Fonetiikan päivät – Helsinki 1992. Papers from the 17th Meeting of Finnish Phoneticians* (A. Iivonen & R. Aulanko, editors), 131-143. Publications of the Department of Phonetics, University of Helsinki, 36.

Norris, D. & A. Cutler. 1988. 'The relative accessibility of phonemes and syllables'. *Perception and Psychophysics* 43(6), 541-550.

Otake, T., G. Hatano, A. Cutler & J. Mehler. 1993. 'Mora or syllable? Speech segmentation in Japanese'. *Journal of Memory and Language* 32, 258-278.

Pisoni, D. & P. Luce. 1987. 'Acoustic-phonetic representations in word recognition'. In Frauenfelder, U. & L. Tyler (eds.), *Spoken Word Recognition,* 21-52. Cambridge, MA: The MIT Press.

Suomi, K. 1993. 'An outline of a developmental model of adult phonological organization and behaviour'. *Journal of Phonetics* 21, 29-60.