VOWELS AS PHONEME DETECTION TARGETS

Anne Cutler, Dennis Norris and Brit van Ooyen

MRC Applied Psychology Unit,
15 Chaucer Rd., Cambridge CB2 2EF, UK.

ABSTRACT

Phoneme detection experiments, in which listeners' response time to detect a phoneme target is measured, have typically used consonant targets. This paper reports two experiments in which subjects responded to vowels as phoneme detection targets. In the first experiment, targets occurred in real words, in the second in nonsense words. Response times were long by comparison with consonant targets, and error rates were high. Targets in initial syllables were responded to much more slowly than targets in second syllables. Full vowels were responded to faster and more accurately than reduced vowels in real words, but *not* in nonwords. Vowel duration correlated negatively with response time. We conclude that the process *of* phoneme detection in English is more difficult for vowels than for consonants, and vowels in words are relatively likely to be responded to on the basis of a lexical representation. We speculate that vowel detection may be less difficult in languages with sparser vowel distributions than English.

1. INTRODUCTION

Phoneme detection is a psycholinguistic task in which listeners are presented with speech input and are instructed to press a response key as fast as they can when they hear an occurrence of a pre-specified phoneme target. The experimental variable is the speed with which listeners respond, i.e. their reaction time (RT). The task (developed by Foss [1]) has typically been used as a tool for studying components of human speech recognition, such as segmentation of continuous speech, word recognition, syntactic processing, *etc;* it has been of little interest in its own right, and the choice of which phonemes to use as detection targets has often been assumed to be arbitrary.

Typically, detection tasks have used stop consonant targets, because stop bursts are relatively easy to locate in a speech signal, and this eases the chore of aligning a mark to initiate response timing at the onset of the target. Other consonantal targets have also been used, but vowels have rarely served as targets. Partly this may have occurred because most phoneme detection experiments require responses to targets in word-initial position only, and fewer English words begin with vowels than with consonants.

Response times in phoneme detection experiments usually average half a second or less, and for detection of word-initial targets, there appear to be no differences in RTs to the six stops [2]. Longer consonants (such as fricatives) are associated with longer RTs than shorter consonants (such as stops) [3]; this is presumably an artefact of the fact that for stops the riming mark tends to be synchronised with the release burst, while for fricatives the mark is synchronised with the onset of frication. If subjects' instructions are to detect targets occurring anywhere in a word rather than in word-initial position only, RTs to word-initial targets are somewhat slower, but in general there is little difference between RTs to targets in initial versus word-internal position [4]; instructions to seek targets anywhere in the word do, however, produce large associative-context and lexicality effects, suggesting that postlexical responses are more likely in such a case [4] [5].

Cutler and Norris [6] proposed that phoneme detection can be performed on the basis of either a prelexical or a lexical representation, with each individual response being the outcome of a race between lexical processing and computation of an explicit phoneme representation from prelexical information. In some experiments different phoneme targets have produced different patterns of effects. Lexicality effects (faster RTs to targets in words than in nonwords) appear with [b] targets but not with [s] [3]; and they are stronger with [b] targets than with [d] [7] [8].

Cutler, Mehler, Norris and Segui [8] explained these differences in terms of Cutler and Norris' Race Model. Studies of perceptual confusions among consonants [9] [10] show patterns of confusability which relate directly to articulatory similarity (so a fricative is most likely to be misperceived as another fricative, and so on). However, these patterns are also influenced by response biases (a very frequent sound in the language is more likely to be erroneously chosen than an infrequent sound is, for example). Goldstein [11] separated out the relative contributions of intrinsic distinctiveness and response bias to confusion matrix patterns; consonants with higher distinctiveness than response bias rankings he labelled relatively unambiguous, consonants with higher response bias than distinctiveness rankings were labelled relatively ambiguous. On this metric, [b] is more ambiguous than either [d] or [s], and as such is harder *to* perceive; thus the

computation of a phonemic representation is slowed and the lexical route is more likely to win the race to produce a detection response. The lower perceptibility of [b] does *not* result in slower RTs; the Race Model assumes more perceptible phonemes are detected via prelexical representations, less perceptible ones via lexical representations, but neither route is intrinsically faster (or there would be no meaningful race).

Phonemes • come in two varieties: vowels and consonants. There seems to be no qualitative difference in how vowels versus consonants are identified [12] [13]; but they may well differ in relative perceptibility. Studies of spontaneous slips of the ear [14] suggest that consonants are misperceived more often than vowels; in particular, vowels in stressed syllables tend to be accurately perceived. Thus a first prediction about vowels as phoneme detection targets is that they may be easier to detect than consonants. Since most vowels occur in word-medial position, one might also make a second prediction, that vowels are quite likely to be responded to post-lexically.

In fact, the few phoneme detection results available for vowels suggest that vowel detection RTs may actually be *longer* than consonant RTs. RTs to detect [a] in the first syllable of (French) words like *balance* and *balcon* were about twice as long as RTs to detect the first syllable *{ba* or *bal)* of the same words [15]. Even in word-initial position vowel detection appears to be comparatively difficult. In a study by Hakes [16], RTs to stop [b,d,g,p,k], nasal [m,n] and glide targets [r,l,w] were in the expected 400-500 msec range; fricatives [s,f] produced slightly longer responses; but RTs to vowel targets were considerably longer.

The present experiments were designed to assess the characteristics of vowels as phoneme detection targets.

2. EXPERIMENT 1

Method

Materials. The five target vowels used were the full vowels /a/, /E/, /I/ and /^/, and the reduced vowel schwa. 120 disyllabic nouns, verbs and adjectives were chosen, 24 for each target vowel. For the full vowels, the words formed sets of four, with the target vowel occurring once in the first and once in the second syllable of words with initial stress and final stress respectively (examples for /a/: *CARton, disCARD, carTOON, PLAcard).* Schwa does not occur in stressed syllables, so for schwa there were only couplets of initial and final stress, with the target always in the unstressed syllable (e.g. *conFUSE, FALcon).* Within each set, the words were matched for frequency and where possible for phonemic environment 50 further mono- and disyllabic words, 10 for each vowel set, were dummy target items, and 1000 words of one, two or three syllables were filler items. Except for a few words containing schwa, no filler items contained a target vowel.

Experimental design. The materials were arranged in five blocks, one for each target vowel. Each block consisted of 44 lists of two to six words in length; of these, 24 lists contained an experimental word in the penultimate (third, fourth or fifth) position, ten lists contained a dummy target in first or second position, and ten lists contained no occurrence of the target. The blocks, plus a short practice set and a small set of example words were recorded by a male native speaker of British English. Five different orders of presentation of the experimental tapes were used. (Because of the acoustic similarity of /^/ and schwa, these two blocks were never adjacent.)

Subjects. 37 students of St. John's College, Cambridge served as paid volunteers for the experiment. All were native speakers of British English with normal hearing. The data for 12 subjects were lost by equipment failure. Five of the remaining subjects heard each order of presentation of the experimental tapes.

Procedure. Subjects were tested individually; they listened to the tapes over headphones and were instructed to press the response key as soon as they heard the specified vowel. Before each block they heard examples of words containing the appropriate target. Response timing was initiated by marks aligned with the onset of experimental words, inaudible to subjects. The data were collected by a microcomputer. The 120 experimental words were digitized and word length, target vowel duration, and the time from target vowel onset to timing mark were measured.

Results

RTs were adjusted for measured timing mark displacement to give RTs from target vowel onset. Two analyses of variance were conducted, with subjects and with words as random factors; we report only effects significant in both. The mean RT was 759 msec (much slower than typical RTs for stop consonants with the same subject population). Fig. 1 shows mean RTs for full vowels vs. schwa in first- vs. second-syllable position. The mean RT to schwa was slower than the the mean RT to full vowels ($F1\ [1,24] = 6.59$, $p < .02$; $F2\ [1,44] = 3.75$, $p < .06$). Vowels in first syllables were detected significandy more slowly than vowels in second syllables ($Fl\ [1,20] = 52.31$, $p < .001$; $F2\ [1,100] = 84.13$, $p < .001$), and the difference between first and second syllables was much greater for schwa than for full vowels ($Fl\ [1,24] = 12.47$, $p < .01$; $F2\ [1,44] = 16.81$, $p < .001$). RT to full vowels was not affected by whether the vowel bore primary (e.g. *CARton, disCARD)* vs. secondary stress (e.g. *carTOON, PLAcard).*

The error rate for the experiment was high, with 23% of targets missed; but the error rate for full vowels was 20%, for schwa significantly higher at 35%. Thus there was no speed-accuracy tradeoff - the vowels most often missed were also responded to slowest. First- and second-syllable targets did not differ in error rate.

A correlation analysis showed that the longer the duration of the vowel, the faster it was detected ($r\ [119] = -.30$, $p < .001$); this was not simply a reflection of the long RTs to the (short) vowel schwa, because the correlation also held for the full vowels alone ($r\ [95] = -.34$, $p < .001$).

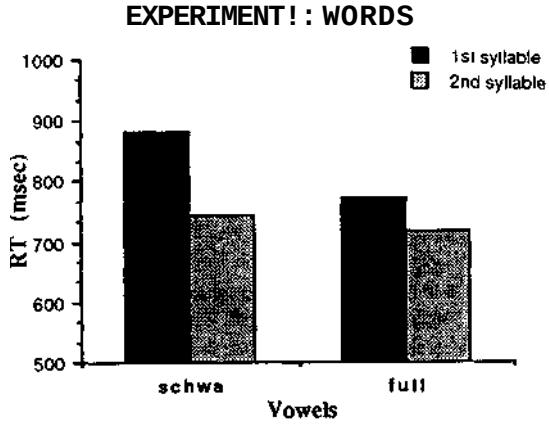

Fig. 1. *Mean reaction time (msec) as a function of vowel quality (schwa or full) and syllable position (first or second), Experiment 1.*

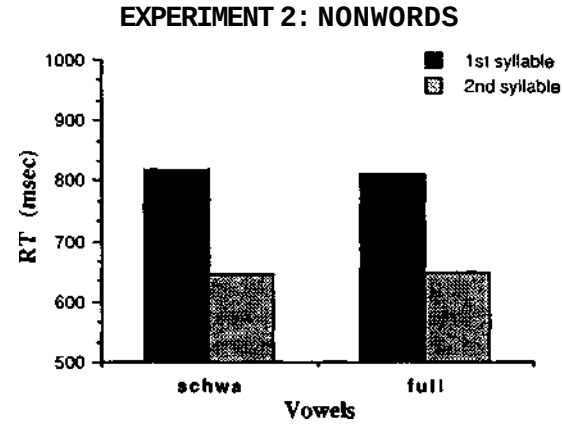

Fig. 2. Mean reaction time (msec) as a function of vowel quality (schwa or full) and syllable position (first or second), Experiment 2.

There was *no* correlation between RT and the duration of the words in which the target vowels occurred; this is evidence that subjects were not waiting till the end of the word before responding.

The RT advantage for targets in second syllables strongly suggests that a significant proportion of responses may have been post-lexical. In similar tasks requiring post-lexical responses (e.g. detection of a mispronounced phoneme), RT decreases steadily across the word [17]. The added difficulty of schwa compared with full vowels also offers indirect evidence for lexical involvement, since inspection of the individual item means showed that an orthographic effect was operative: responses to schwa were faster when the orthographic representation was "e", suggesting that *"e"* may act as a canonical orthographic representation for schwa. In the experimental words the vowels /E/, /I/ or /^/ all had constant representations, and in all but three words /a/ was represented by *"ar"* (the mean RT for the remaining three words was long by comparison with the /a/ mean). Schwa, however, was orthographically represented in our word set in four different ways, with "e" being the most common representation (9 of 24 items).

If these effects indeed represent lexical involvement, they should disappear if lexically mediated responding is ruled out, for instance if the targets are presented in non-words, which have no lexical representations. Accordingly we conducted a second experiment in which the target vowels occurred in nonwords.

3. EXPERIMENT 2

Method

Materials and Design. Using the same target vowels, the same number of items was constructed as in Experiment 1, except that all items were nonwords. Because of the relative freedom of choice in making up nonsense words, all target sets could be controlled for phonemic environment. Examples for the target /a/ are: *LARTome, poLART, larTOACE, DROlart;* for schwa: *penZINE, CLYpen.* The experimental design was as for Experiment 1.

Subjects. Fifty students of Downing College, Cambridge were paid for participating. All were native speakers of British English with normal hearing. Ten heard each order of presentation of experimental tapes.

Procedure. The procedure was as for Experiment 1, except that while half the subjects read the instructions as before, the other half listened to taped instructions, recorded by the same speaker as in the experimental tapes.

Results

RTs were adjusted and analysed as in Experiment 1. The mean RT was again very long (729 msec), indicating that difficulty of vowel detection is not dependent on lexically mediated responding. In this experiment there was no RT difference between full vowels and schwa. However, as with the real words, targets in first syllables were detected significantly less rapidly than targets in second syllables (Fl [1,40] = 566.04, p <.001; F2 [1,100] = 116.28, p < .001). Again, there was no difference between full vowel targets in syllables with primary vs. secondary stress. There was no effect of how the instructions were presented. Fig. 2 shows mean RTs as for Experiment 1.

The error rate was again high (28%), but there was this time no difference between the vowels - all produced a mean error rate in the range 26%-29%. Again, first- and second-syllable targets did not differ in number of errors.

Just as in the previous experiment, measured vowel duration correlated negatively with RT (r [119] = -.28, p < .01), and the correlation held also for the four full vowels alone (r [95] = -.32, p <. 001).

Thus the results of Experiment 2 closely replicate those of Experiment 1 except that schwa was in this case detected neither less rapidly nor less accurately than full vowels.

4. CONCLUSION

These two experiments have shown that (English) vowels are difficult to detect as targets in a speeded response task. This contradicts the prediction that they should have proved easy to detect because their intrinsic perceptibility is relatively high. The second prediction, that post-lexical responding should be likely, received indirect confirmation via the apparent orthographic involvement in the difficulty of schwa detection in real words; since this effect disappeared with nonword materials, it is not a reflection of the acoustic-phonetic structure.

The strongest effect of all was that vowels in the first syllable of disyllables took longer to detect than vowels in second syllables (but this RT effect was not mirrored in the error data). This is *not*, contrary to our earlier suggestion, a lexical effect, since it also appeared with nonwords. We suggest that this finding is an artefact of the tendency in English for word-final syllables to be lengthened, combined with the strong negative correlation which we found between measured vowel duration and RT.

The general difficulty of vowel detection (as reflected both by long RTs and high error rates) is not a function of post-lexical responding, for two reasons: the difficulty is also present with nonword materials, and previous studies have shown that post-lexical responses are no longer than pre-lexical [8]. Likewise it does not arise because vowels (usually) occur word-medially; again, previous studies have shown that word-medial targets are responsed to no more slowly than word-initial [4], and long RTs to vowels also occur in word-initial position [16]. It does not reflect the fact that vowels are relatively long phonemes (by analogy with the finding reported above that longer RTs to [s] than to [b] reflect the greater length of [s] [3]), because measured vowel length correlated *negatively* with RT - longer vowels produced faster responses. We suggest that the problem with vowels as phoneme detection targets lies in the key concept of intrinsic ambiguity as proposed by Goldstein [11]. Goldstein analysed only consonants, and no ambiguity ratings for vowels are available in the literature. However, note that the vowel space of English is relatively densely populated; for distributional reasons alone distinctiveness of vowel types is likely to be low. If this is why vowels proved so difficult in our study, it might be possible to improve vowel detection performance by using only a few, highly distinct vowel targets. An alternative approach would be to compare vowel detection in English with vowel detection in another language in which the vowel space is more sparsely populated; in Japanese, for instance, there are only five vowels, which occupy highly distinct positions in the vowel space. If our interpretation of the relative difficulty of vowel detection in comparison with consonant detection in English is correct, it may be the case that vowel detection would *not* prove harder than consonant detection in Japanese.

5. REFERENCES

[I] D.J. Foss, "Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times," /. *Verb. Learn. Verb. Behav.,* Vol. 8, pp. 457-462, 1969.

[2] M. Martin, "Reading while listening: A linear model of selective attention," /. *Verb. Learn. Verb. Behav.,* Vol. 16, pp. 453-463, 1977.

[3] P. Rubin, M.T. Turvey & P. Van Gelder, "Initial phonemes are detected faster in spoken words than in non-words," *Perc. & Psychophys.,* Vol. 19, pp. 394-398, 1976.

[4] U.H. Frauenfelder & I. Segui, "Phoneme monitoring and lexical processing: Evidence for associative context effects," *Mem. & Cog.,* Vol. 17, pp. 134-140, 1989.

[5] U.H. Frauenfelder, J. Segui & T. Dijkstra, "Lexical effects in phonemic processing: Facilitatory or inhibitory?" *J. Exp. Psy.: Hum. Perc. Perf.,* Vol. 16, pp. 77-91, 1990.

[6] A. Cutler & D. Norris, "Monitoring sentence comprehension," in W.E. Cooper and E.C.T. Walker (Eds.) *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett,* Hillsdale, N.J.: Erlnaum. 1979.

[7] D.J. Foss & M.A. Gemsbacher, "Cracking the dual code: Toward a unitary model of phonetic identification," *J. Verb. Learn. Verb. Behav.,* Vol. 22, pp. 609-632, 1983.

[8] A. Cutler, J. Mehler, D. Norris & J. Segui, "Phoneme identification and the lexicon," *Cog. Psych.,* Vol.19, pp. 141-177, 1987.

[9] M.D. Wang & R.C. Bilger, "Consonant confusions in noise: A study of perceptual features," *J. Acoust. Soc. Amer.,* Vol. 54, pp. 1248-1266, 1973.

[10] G.A. Miller & P. Nicely, "Analysis of perceptual confusions among English consonants," *J. Acoust. Soc. Amer.,* Vol. 27, pp. 338-352, 1955.

[11] L. Goldstein, "Bias and asymmetry in speech perception," in V.A. Fromkin (Ed.) *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand,* New York: Academic Press, 1980.

[12] C.J. Darwin & A.D. Baddeley, "Acoustic memory and the perception of speech," *Cog. Psych.,* Vol.6, pp. 41-60, 1974.

[13] D.H. Whalen, "Vowel and consonant judgements are not independent when cued by the same information," *Perc. & Psychophys.,* Vol. 46, pp. 284-292, 1989.

[14] Z.S. Bond & S. Games, "Misperceptions of fluent speech," in R. Cole (Ed.) *Perception and Production of Fluent Speech,* Hillsdale, NJ: Erlbaum, 1980.

[15] J. Mehler, J.-Y. Dommergues, U.H. Frauenfelder & J. Segui, "The syllable's role in speech segmentation," *J. Verb. Learn. Verb. Behav.,* Vol. 20, pp. 298-305, 1981.

[16] D.T. Hakes, "Does verb structure affect sentence comprehension?" *Perc. & Psychophys.,* Vol. 10. pp. 229-232, 1971.

[17] W.D. Marslen-Wilson & A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech," *Cog. Psych.,* Vol. 10, pp. 29-63, 1978.