Speech Communication 11 (1992) 101 108
North-Holland

Detection of vowels and consonants with minimal acoustic variation

Brit van Ooyen, Anne Cutler and Dennis Norris
MRC Applied Psychology Unit, 15 Chaucer Rd. Cambridge CB2 2EF, UK

Received 26 September 1991
Revised 7 February 1992

Abstract. Previous research has shown that, in a phoneme detection task, vowels produce longer reaction times than consonants, suggesting that they are harder to perceive. One possible explanation for this difference is based upon their respective acoustic/articulatory characteristics. Another way of accounting for the findings would be to relate them to the differential functioning of vowels and consonants in the syllabic structure of words. In this experiment, we examined the second possibility. Targets were two pairs of phonemes, each containing a vowel and a consonant with similar phonetic characteristics. Subjects heard lists of English words and had to press a response key upon detecting the occurrence of a pre-specified target. This time, the phonemes which functioned as vowels in syllabic structure yielded *shorter* reaction times than those which functioned as consonants. This rules out an explanation for response time difference between vowels and consonants in terms of function in syllable structure. Instead, we propose that consonantal and vocalic segments differ with respect to variability of tokens, both in the acoustic realisation of targets and in the representation of targets by listeners.

Zusammenfassung. Fruhere Forschung hat gezeigt. dass Vokale langere Latenzzeiten ergeben als Konsonanten, wenn die Methode der Phonemdetektion gebraucht wird. Dies impliziert, dass Vokale schwieriger zu erkennen sind. Eine mogliche Erkliirung nimmt Bezug auf die respektiven akustischen und artikulatorischen Eigenschaften beider Phonemkategorien, eine zweite auf die verschiedene Funktion der Vokale und Konsonanten in der Silbenstruktur des Wortes. In diesem Experiment untersuchten wir diese zweite Erklarungsmoglichkeit. Zwei Phonempaare, jedes bestehend aus einem Vokal und einem Konsonant mit ahnlichen phonetischen Eigenschaften, wurden als Zielphoneme gebraucht. Die Versuchspersonen horten Listen englischer Worter und driickten auf einen Antwortknopf sobald sie ein vorher spezifiziertes Phonem erkannten. In diesem Experiment ergaben die Phoneme *kurzere* Latenzzeiten, die in der Silbenstruktur wie Vokale funktionieren, als diejenigen die wie Konsonanten funktionieren. Dieses schliesst eine Erkliirung auf Grund verschiedener Funktionen in der Silbenstruktur aus. Stattdessen schlagen wir vor. dass Vokale und Konsonanten sich sowohl in der akustichen Variabilitat individueller Sprachlautbildung wie beim Horer in ihrer mentalen Representation von einander unterscheiden.

Resume. Dans les etudes precedentes, on a constate que dans une tache de detection, les voyelles donnent lieu a des temps de reaction plus longs que les consonnes. ce qui suggere que les voyelles sont plus difficiles a percevoir. II y a deux explications possibles pour ceci. La premiere est qu'il s'agit de differences acoustiques et articulators. Une seconde explication met en cause les roles differents des voyelles et des consonnes dans la structure syllabique des mots. Dans la presente experience, on a examine la seconde possibility. On a utilise comme cibles deux paires de phonemes, chaque paire etant constitute d'une voyelle et d'une consonne qui se ressemblent en termes des caracteristiques phonetiques. Les sujets devaient appuyer sur un bouton aussitot qu'ils avaient repere dans une liste de mots anglais un phoneme cible prespecifie. Cette fois, les phonemes jouant le role de voyelle dans une structure syllabique ont donne des temps de reaction *inferieurs* a ceux des phonemes jouant le role de consonne. Ceci elimine une explication de la difference des temps de reaction qui serait basee sur le role du phoneme dans la structure syllabique. Nous proposons, en revanche, que cette difference provienne de la variation acoustique telle qu'elle se manifeste dans la realisation des phonemes cibles et dans leur representation mentale chez les auditeurs.

Keywords. Speech perception; phoneme detection; vowels; consonants; semivowels.

1. Introduction

The sounds of speech come in two varieties: vowels and consonants. All languages have both kinds of phonemes, and language users usually have some awareness of the distinction. However, it is generally agreed that a precise dividing line between the two categories is hard to draw, and where it is drawn can depend on whether articulatory. acoustic or phonological criteria are invoked.

Vowels can be characterised on articulatory/ acoustic dimensions as relatively steady-state, periodic sounds, produced with vibration of the vocal cords and without obstruction of the airflow from the lungs. In phonological terms, they form syllabic nuclei. Consonants, in contrast, are relatively transient, often aperiodic, produced with full or partial obstruction of the airflow from the lungs, and with or without vocal cord vibration. Phonologically, they occur in the margins of syllables in onsets and codas.

The voiceless plosives /p,t,k/ are perhaps the '"most consonantal" phonemes. Other consonants share certain characteristics with vowels; the nasals /m,n/. for instance, are produced with vocal cord vibration throughout - indeed, nasalised vocalic segments in some languages resemble consonantal segments in other languages. The most vowel-like of the consonant phonemes of English are the so-called semivowels /w/ and / j / . Acoustically. they are relatively steady-state and periodic, and hence could quite plausibly be classified as vowels. Phonologically. however, they are consonants, since they cannot function as the nucleus of a syllable.

Consonants and vowels provide each other with facilitating context in speech perception. Vowels are easier to identify if they are bounded by consonants (Strange et al., 1976; Strange et al., 1979). They are also easier to detect in consonantal context (Rakerd et al., 1984). Consonants are likewise easier to identify if they are bounded by vowels (Liberman et al., 1954).

Many experimental studies have addressed the question of whether the acoustic/articulatory differences between vowels and consonants are reflected in human speech processing. Identification and discrimination of both vowels and consonants have been extensively investigated, and it was claimed that identification of vowels and of consonants was fundamentally different, with perception of consonants being categorical, perception of vowels continuous (Liberman et al., 1967). This view was called into question by, among others, Ades (1977) who pointed out that the effective perceptual range of any vowel category, defined in numbers of just noticeable differences (JNDs), is larger than the effective range of consonant categories - that is, within any vowel category there are more members between which listeners can just perceive a difference than within any consonant category. This difference would produce better discrimination performance for vowel tokens across a continuum than for consonant tokens, which in turn would make vowel identification appear less categorical than consonant identification. Thus the evidence from identification and discrimination tasks cannot be taken as indicating fundamentally different perceptual functions for the two phoneme categories.

However, there is evidence that vowels and consonants produce unexpectedly differing performance patterns in another experimental task: phoneme detection or phoneme monitoring. This task (devised originally by Foss (1969)) requires listeners to respond as soon as they hear a prespecified target phoneme; the dependent variable is response time (RT). The detection task has been a valuable psycholinguistic tool for the investigation of processes such as lexical access; it has not really been studied as a topic in its own right. Because of this, the choice of which phonemes to use as targets has generally been motivated by which sounds are comparatively easy to locate in a speech signal, rather than by questions pertaining to the sounds themselves. In practice, most detection experiments have used stop consonants, and vowels have rarely served as targets.

In most phoneme-monitoring experiments targets occur in word-initial position only. In such experiments RTs typically average half a second or less. In so-called generalised phoneme monitoring (GPM; Frauenfelder and Segui, 1989), targets may occur anywhere in a word. Here, RTs to word-initial targets are somewhat longer, but in general there is little difference between RTs to targets in initial versus word-internal position. (GPM does however produce large associative-context and lexicality effects, suggesting that post-lexical responses are more likely in such a case.)

Results from other perceptual studies suggest that, if anything, vowels should have proved easier to detect than consonants. Studies of spontaneous slips of the ear show that consonants are misperceived more often than vowels (Bond and Garnes, 1980). In particular, vowels in stressed syllables seem to be accurately perceived. Assuming that this accuracy can be associated with their comparative resilience to perceptual distortion as a result of their relative prominence in the acoustic signal, it would seem plausible to expect that vowels would stand out equally in a detection task. Yet the very few phoneme detection results previously available for vowels suggested this is not the case. Mehler et al. (1981) found longer RTs for detection of vowels than for detection of the syllables in which they occurred, and analysis of data from Hakes (1971) revealed longer RTs for detection of vowels than of consonants.

The same pattern of results was observed by the present authors in a series of experiments designed to assess the characteristics of British English vowels as phoneme detection targets. In two experiments, using real words and nonsense words, Cutler et al. (1990) found that vowel RTs were very long in comparison with RTs reported in previous work on consonants. Moreover, error rates were high. It was concluded that detection of vowels is difficult.

In a follow-up experiment two highly distinct vowel targets were compared with two relatively confusible stop consonants (van Ooyen et al., 1991). RTs to the vowels were significantly longer than to the plosives. Only in word-initial position did detection times for vowels approach those for stop consonants.

In the present study, we address the question *of why* vowels are difficult to detect in monitoring tasks. As discussed above, there are two principal ways in which vowels and consonants differ: in acoustic/articulatory characteristics, and in their function in syllable structure. Each of these differences could potentially offer an explanation of the RT effects. However, there is one sense in which syllabic function may seem to be the more likely candidate. This is because of the discrepancy between the detection RT task (in which there is evidence that vowels produce greater difficulty than consonants) and identification and discrimination tasks (in which there is no evidence that vowels produce greater difficulty than consonants). The latter tasks have typically used only the simplest phonological sequences, which may perhaps mean that syllabic function had little opportunity to affect performance. It may likewise be reasonable to suppose that the syllabic function should play a larger role in phoneme detection tasks, since such tasks usually use real words; therefore it may be only in the detection task that the syllabic function has had an opportunity to exercise an effect.

We therefore concentrated in this experiment on the role of the syllabic function in the detection RT difference between consonants and vowels. To enable as pure a test as possible of syllabic function alone, we compared vowels and consonants which differed minimally in articulatory/acoustic characteristics; that is, we compared detection times for vowels and for the consonants which most closely resemble vowels.

2. Experiment

2.1. Method

2.1.1. Materials

To test syllabic function while controlling acoustic/articulatory differences as closely as possible, we chose as target phonemes those English consonants which are, acoustically, minimally different from vowels, but which function syllabically as consonants, namely the semivowels /j/ and /w/. These are characterised by Ladefoged (1982) as "non-syllabic versions of the English high vowels /i/ and /u/, respectively" (p. 209). Therefore we compared detection performance of */j/* with that of */i/* and of /w/ with that of /u/. If the RT difference between vowels and consonants is due to the function played by each type of phoneme in syllable structure, we would expect that semivowels, which function as consonants, would produce faster RTs than vowels in the same way that stop consonants did. On the other hand, if the differences are due to acoustic structure of vowels versus consonants, then we would expect that semivowels, which acoustically resemble vowels more closely than stop consonants, should produce a response pattern more similar to that of vowels.

The four target phonemes in the experiment were, therefore, high front /i/ and high back /u/ plus their corresponding semivowels /j/ and /w/, respectively. Due to restrictions of the English vocabulary, it was not possible to compare the phonemes in all word positions. There are no English words ending with /j/ or /w/, and very few beginning with /u/. Thus comparisons of /j/ and /i/ were limited to initial and medial position, and comparisons of /w/ and /u/ were limited to medial position only.

144 monosyllabic and disyllabic words were chosen, 36 for each target phoneme. For /j/ and /i/, 14 of the 36 words had the target phoneme in word-initial position, and seven of these were monosyllabic, seven disyllabic. The remaining 22 words (11 of which were monosyllabic, 1 disyllabic) had the target phoneme in word-medial position. For /w/ and /u/, all of the 36 words had the target phoneme in word-medial position. Of these 36 words, 20 were monosyllabic and 14 were disyllabic. Target phonemes always occurred in a stressed syllable. The words were matched for mean frequency across target phonemes within each position sub-group. A further 40 words, ten per target phoneme, were dummy target items. About 100 further words served as fillers.

The words were arranged in four blocks, one for each target phoneme. Each block consisted of 55 lists, of two to six words in length. Of these, 36 lists had an experimental item in third, fourth or fifth position, ten lists had a dummy target phoneme in first or second position and ten lists contained no occurrence of the target phoneme.

The materials were recorded in a sound-dampened chamber by a male native speaker of British English. The recording was made on DAT tape using an AKG 1000 S microphone. Timing marks, inaudible to the subjects, were placed on the second channel of the tape, aligned approximately with the onset of each target-bearing word.

2.1.2. Procedure

The subjects were presented with taped instructions that requested them to press the response key as soon as they detected an occurrence, anywhere in a word, of the specified target phoneme. The instructions emphasised that subjects should concentrate on how the word was spoken, rather than how it was spelt.

The blocks were presented in four different orders. Before each block the target for that block was specified with examples.

Response timing was initiated by the timing mark aligned with each experimental word, and terminated by the subject's keypress. Timing and data collection and storage were controlled by a Zenith microcomputer.

The 144 experimental words were digitized and measurements were made of word length, target phoneme duration, and the time between target phoneme onset and timing mark. RTs were adjusted for these latter measurements to give RTs exactly from target phoneme onset.

2.1.3. Subjects

Twenty-four members of the Applied Psychology Unit subject panel, aged between 19 and 46, participated in the experiment for a small payment. All were native speakers of British English and all reported normal hearing. Six heard each order of presentation of the blocks.

2.2. Results

Response times below 100 msec or greater than 1500 msec were discarded. (This resulted in the loss of 1.6% of the data.) Two analyses of variance, with subjects and with words as random factors, were carried out. We report only effects significant in both.

Figure 1 shows mean RTs in ms for the vowels (521 ms) and the semivowels (628 ms). The two vowels were responded to significantly faster than the two semivowels {F\ [1,20] = 88.3, p<0.001; $F2$ [1, 140] = 80.41,p< 0.001). The difference was in the same direction and significant for all subcomparisons: medial /u/ (517 ms) versus medial /w/ (601ms; $t1$ [23] =4.27, p< 0.001, tl [35] = 6.03, p<0.001); initial /i/ (481 ms) versus initial /j/ (605 ms; t\ [23] = 7.44,p<0.001, $t2$ [13] = 7.48, p< 0.001); medial /i/ (557 ms) versus medial /j/ (691ms; $t1$ [23] = 5.87, p< 0.001, $t2$ [21] =9.37, p< 0.001). Figure 2 shows mean RTs in ms for the vowels and the semivowels separately for each phoneme pair and each word position.

An error analysis revealed that 14.8% of the semivowels were missed; this was a significantly higher error rate than the one for the vowels at

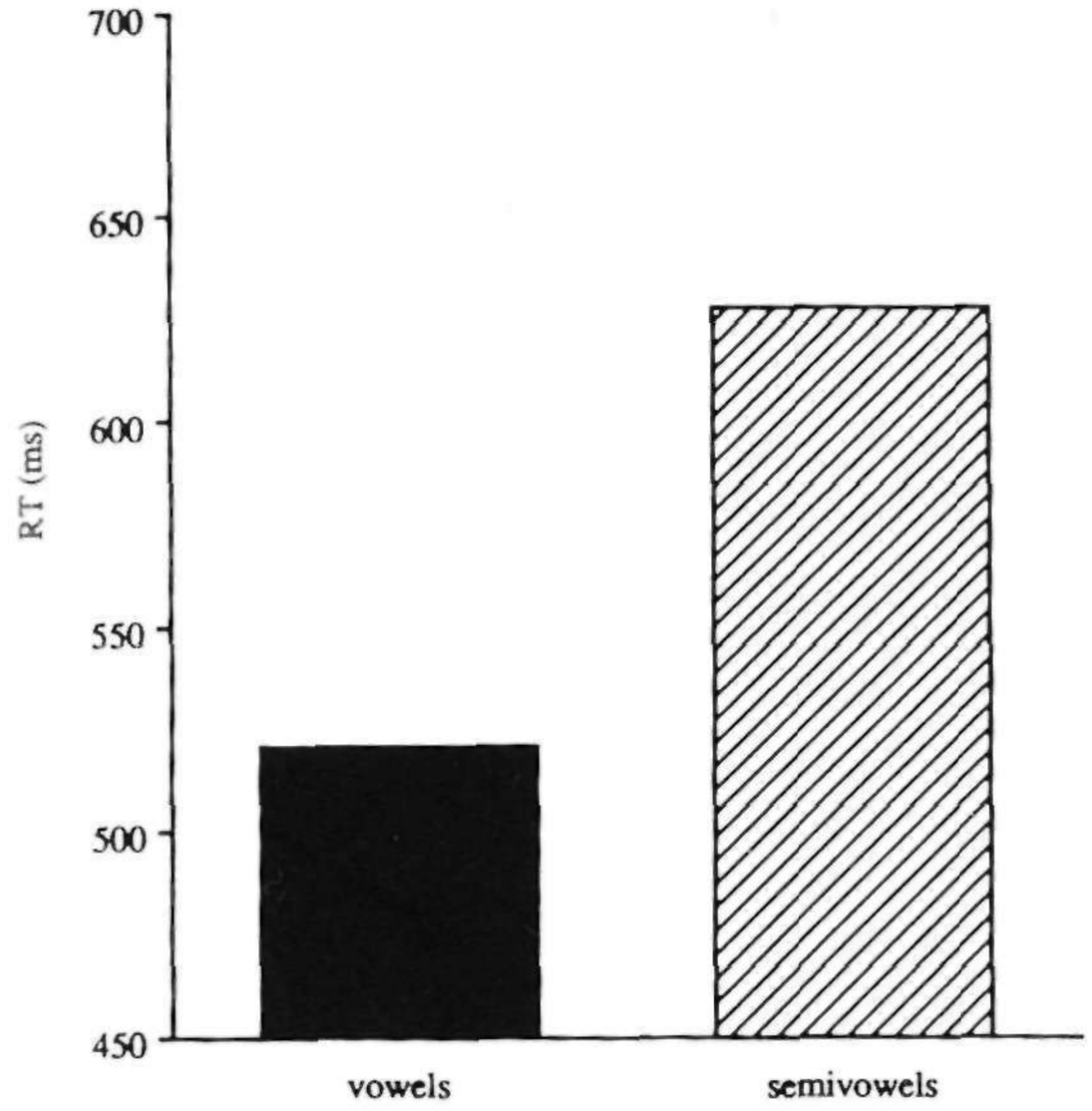

Fig. I. Mean reaction time (ms) as a function of phonemic category (vowels i,u versus semivowels /j,w/).

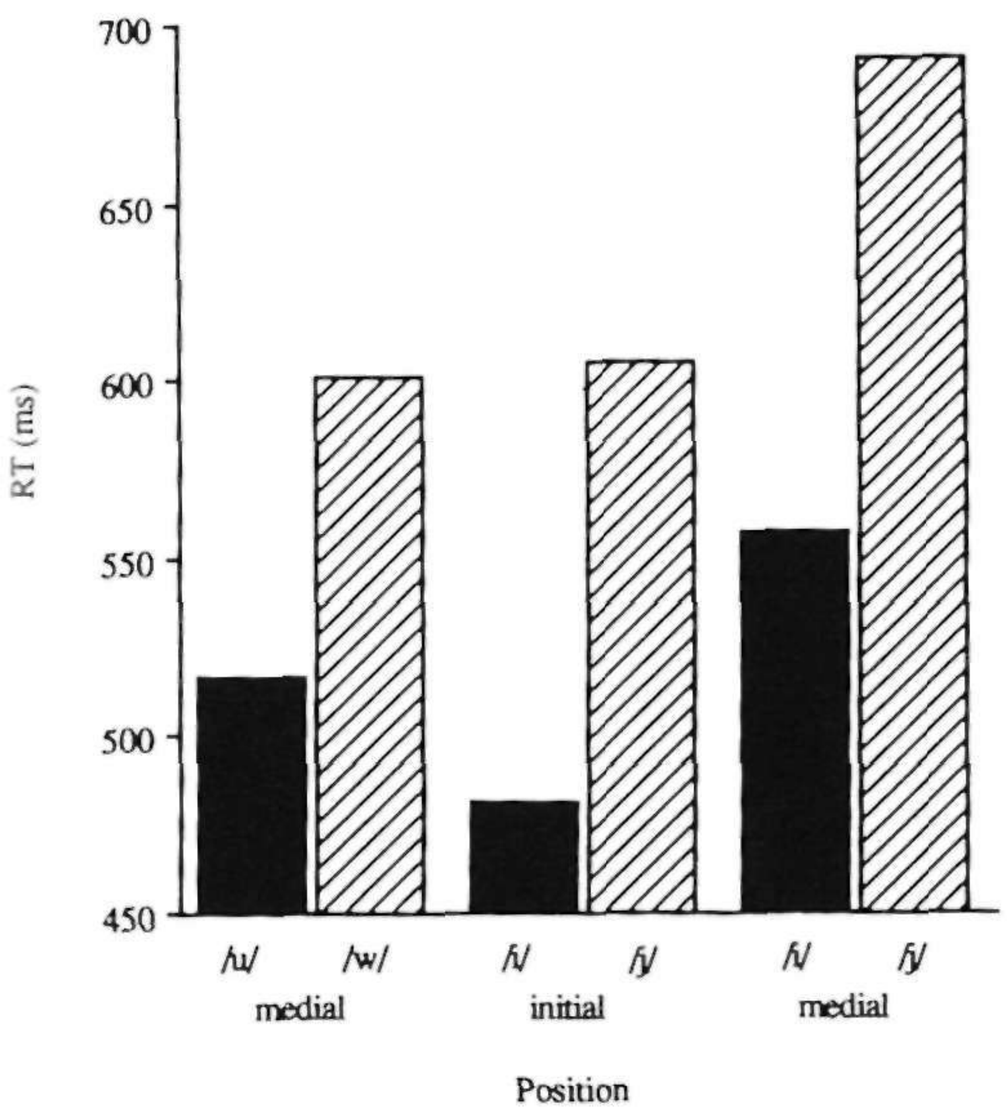

fig. 2. Mean reaction time (ms) as a function of phonemic category (vowels /i,u/ versus semivowels /j,w/) and position of the word.

.8% *(F* [1,20]= 15.38, *p*< 0.001; *F2* [1, 140] = 18, *p*<0.001). The difference was again in the ime direction and significant for all sub-comparisons: medial /u/ (3% error) versus medial /w/ .4%; *t1* [23] = 2.8, p<0.01, *t2* [35] = 3.62, p<0.001); initial /i/ (1.2%) versus initial /j/ (8.3%; *t1* [23] =1.77, /7<0.09, *t2* [13] = 5.06, p<0.001); medial /i/ (4.2%) versus medial /j/ (28.4%; *t1* [23] = 3.64, *p*<0.001, /2 [21]= 13.24, p < 0.001).

There was a negative correlation between RT and duration for the vowels (the longer the vowel, the faster the RT) *r* [71] = 0.27,*p*<0.025), but no such effect for the consonants. Analyses for each phoneme separately showed that this correlation was significant for /u/ alone *(r* [35] = 0.33, p<0.05).

On both RT and error rate measures, therefore, the semivowels in this experiment produced worse performance than the vowels, just as in our previous experiments vowels had produced worse performance than stop consonants on both measures.

3. Discussion

The results of this experiment have shown a clear RT disadvantage for semivowels in comparison to vowels in a phoneme detection task. Therefore it is not the case that consonants of any type will necessarily produce better performance on this task than vowels. This in turn conclusively rules out an explanation of the previous findings in terms of syllabic function. Semivowels function syllabically as consonants; yet they were not privileged in comparison to vowels.

One factor which clearly plays some role in these findings is orthographic interference. Cutler et al. (1990) found evidence for an orthographic effect in detection of the vowel schwa: responses to this vowel were faster when the orthographic representation was "e", suggesting that "e" may act as the canonical orthographic representation for schwa. Similarly, we suspect that orthography played a role in the large number of missed responses for word-medial /j/ in the present experiment. In experimental words such as *dune, cubic, fuse,* there is no corresponding grapheme for the phoneme /j/. (For instance, there is no difference in spelling to indicate the presence of the phoneme /j/ in British English *duty* as opposed to the absence of this sound in some varieties of American English.)

If subjects indeed have a canonical orthographic representation of sounds, and this facilitates

responses to target phonemes which are orthographically represented in the canonical form, then it is likely that responses will be even slower when there is no corresponding orthographic representation whatsoever. The results of the error rate analysis support this suggestion: /j/ was significantly more often missed in word-medial position, where no orthographic symbol was available, than in word-initial position, where the orthographic representation was always "y". It would seem that subjects find it very difficult to make judgements on the basis of phonetic information alone. Since. in a language like English, such judgements can usually be supported by orthographic evidence, making these judgements is particularly difficult where no orthographic evidence is available.

However, it is equally clear that orthography cannot provide the entire explanation. Our results showed that although the RT disadvantage for /j/ in comparison to /i/ was significant in medial position (in which no orthographic representation for /j/ was present), it was just as significant in initial position (in which the orthographic representation of /j/ was consistently "y"). Note also that the orthographic representation of /i/ was highly variable in both positions: *eel, eke, evil, seal, seize, siege,* etc. Similarly, in our previous experiment which compared the vowels /a/ and /i/ with the stop consonants /p/ and t/ (van Ooyen et al., 1991). /a/ was consistently represented by orthographic "ar" (in *art, cigar, sparse,* etc), yet /a/ was detected more slowly than /i/, which had considerable orthographic variation (in *equal, seek. tea, priest, key,* etc). Moreover, if it can be argued that orthographic "a" is an ambiguous symbol because it can also represent other vowel sounds such as in *back,* the same argument should apply to the consonant targets used in that experiment: orthographic "p" occurs in *photo* as well as in *pole,* and "t" occurs in *thin* as well as in *tin.* Yet the stop consonant RTs were significantly shorter than the RTs to each vowel. Finally, the strongest evidence that there is more in these findings than can be explained by orthography comes from our previous work: Cutler et al. (1990) found equivalent RTs and error rates for vowels in real words and in nonsense words. Subjects can have no prior orthographic representation for nonsense words; if they construct an orthographic representation in order to perform the phoneme detection task, then surely they are free to construct it solely in terms of putative canonical representations. For all of these reasons, we hold that orthographic effects, while arguably present, cannot account fully for the results of the present experiment.

How then can we unify the present results with previous findings? Detection RTs for vowels are longer than for stop consonants, but RTs for semivowels are longer still. The present finding rules out syllabic function as an explanation of the difference between vowels and stops. We suggested that the most likely alternative explanation should be sought in the acoustic articulatory characteristics of different phoneme categories. However, the fact that a significant RT difference was found between two phoneme categories with minimal acoustic variance - semivowels and their corresponding vowels indicates that such an acoustic articulatory explanation will also not be a simple matter.

We suggest that the key factor determining phoneme detection difficulty is the perceived variability across individual realisations of members of a phoneme category in the speech signal. As Ades (1977) has pointed out, the effective perceptual range of vowel categories is larger than that of consonant categories. The experiments which Ades reviewed largely compared vowels with stop consonants: thus his argument is most applicable to our results for vowels and stops. With respect to these results, the difference in effective range could affect how subjects perceive individual tokens which represent occurrences of the specified target in a detection task in comparison to the token with which the target was originally specified. Specifically, where a difference between the two tokens exists, the difference is more likely to *be perceptible* (and hence possibly slow down RTs) in the case of vowels than in the case of at least, stop consonants. Such a difference could have two origins: it could arise because the spoken realisation of the phoneme target varied, e.g. as a function of surrounding phonetic context; or it could arise because of change in the subject's mental representation of the specified target.

Both of the latter possibilities seem to us relatively likely. Consistent with the first suggestion is the finding by van Ooyen et al. (1991 ; Experiment I) that RTs to vowels and to consonants were *not*

significantly different in word-initial position, in which no prior context distorts phonemic realisation. Consistent with the second suggestion is the proposal by Cowan and Morse (1986) that mental representations of vowels gradually become more diffuse over time.

Thus an explanation in terms of relative variability seems viable for the contrast between vowels and stop consonants. To incorporate the findings on semivowels into this picture, then, we would assume that, like vowels, semivowels have a comparatively large perceptual range. This seems intuitively plausible: semivowels are acoustically closer to vowels than to consonants. However, in the present experiment semivowels exhibited an added disadvantage in relation to the vowels; the difference appeared even in word-initial position. To extend the variability argument to semivowels it may be necessary to assume further that they are yet more subject to variability as a function of surrounding context than vowels, and/or that their memory representations are yet more variable than the representations for vowels. Unfortunately the available literature does not contain data which would resolve this issue. But we note that one marked difference between our two target sets was in their measured duration: the semivowel durations ranged from 27 ms to 111 ms, the vowel durations from 108 ms to 329 ms, i.e., there was virtually no overlap between the two sets of measurements. The phoneme duration measurements published by Crystal and House (1988) show a similar pattern: the glides in their data set were systematically shorter than the long vowels (such as the /u/ and i used as targets in the present study). However. it is noticeable that Crystal and House's glide measurements are very similar to the values which they report for *short* vowels. It may be, then, that response patterns for short vowels in a phoneme detection task would more closely resemble those for semivowels than those for long vowels. Indeed, in the experiments of Cutler et al. (1990), in which the majority of vowel targets were short, RTs and error frequencies were high, and more similar to the present results for semivowels than to those for ong vowels. Furthermore, we know that duration affects RT to vowels, since here and in our previous experiments we have observed a negative correlation between measured vowel duration and RT. Thus our present results may reflect simply an effect of duration, and no difference at all as a function of vocalic versus consonantal category.

Appendix A. Experimental materials

/j/	/i/	/w/	/u/
word-initial position		word-medial position	
yawl	eel	dwarf	noose
yes	east	swift	shoot
yolk	eke	sweep	soup
your	each	swing	noon
youth	eat	twitch	toot
yard	ease	swerve	soothe
yarn	eve	twin	tomb
yell	epoch	tweed	boom
yonder	ether	swamp	scoop
yellow	evil	dwell	broom
yogurt	equal	quest	food
yodel	easel	quaint	loose
yeoman	eagle	squeeze	roof
yoga	ethos	queen	cool
		quote	loop
word-medial position		twist	crude
		twig	flute
fuel	seal	quick	goose
feud	fiend	swim •	droop
dune	deem	twelve	boot
tune	tease	twinkle	poodle
due	deal	twilight	toothache
fuse	seize	sweater	super
muse	mead	dwindle	noodle
mule	keel	swivel	souffle
pure	beast	framework	foolish
mute	sleep	backward	cooper
cure	siege	sequence	doomsday
futile	feline	seaweed	hoover
nuisance	diesel	upward	rooster
bugle	beadle	meanwhile	movement
tuba	keeper	aware	cocoon
music	reason	reward	gruesome
feudal	feeble	farewell	booster
stupid	speaker	request	ruler
pupil	ceiling	unwise	prudence
fuming	deacon		
tuesday	feature		
cubic	seashore		

Acknowledgment

This research was supported by ESPRIT Basic Research Actions [project P3207 "ACTS"]. A preliminary report of the experiment was presented to *Eurospeech 91,* Genova, September 1991. (N.B. As a result of an error in earlier analyses the figures in that report differed slightly from the corrected figures in the present paper.)

References

A.E. Ades (1977), "Vowels, consonants, speech and non-speech". *Psychological Rev.,* Vol. 84. pp. 524 530.

Z.S. Bond and S. Games (1980). "Misperceptions *o\'* fluent speech", in *Perception and Production oj Fluent Speech,* ed. by R. Cole (Erlbaum. Hillsdale. NJ).

N. Cowan and P. A. Morse (1986), "The use of auditory and phonetic memory in vowel discrimination", *J. Acoust. Soc. Amer..* Vol. 79. pp. 500 507.

T.H. Crystal and A.S. House (1988). "Segmental durations in connected-speech signals: Current results"../. *Acoust. Soc. Amer.,* Vol. S3, pp. 1553 1573.

A. Cutler. D. Norris and B. van Ooyen (1990). "Vowels as phoneme detection targets". *Proc. Internal. Con/. Spoken Language Processing,* Kobe. Japan. Vol. I. pp. 581 584.

D.J. Foss (1969). "Detection processes during sentence comprehension : Effects of lexical item difficulty and position upon decision times". *J. Verb. Learn. Verb. Behav.,* Vol. 8, pp. 457 462.

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

D.T. Hakes (1971), "Does verb structure affect sentence comprehension?", *Perception & Psychophysics,* Vol. 10. pp. 229 232.

P. Ladefoged (1982). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).

A.M. Liberman, F.S. Cooper, D.P. Shankweiler and M. Studdert-Kennedy (1967). "Perception of the speech code". *Psychological Rev..* Vol. 74, pp. 431 461.

A.M. Liberman. P.C. Delattre, F.S. Cooper and L.J. Gerstman (1954), "The role of consonant-vowel transitions in the perception of the stop and nasal consonants". *Psychological Monographs,* Vol. 68 (8, Whole No. 379).

.1. Mehler, J.-Y. Dommergues. U.H. Frauenfelder and J. Segui (1981), "The syllable's role in speech segmentation". ./. *Verb. Learn. Verb. Behav.,* Vol. 20. pp. 298 305.

B. van Ooyen, A. Cutler and D. Norris (1991), "Detection times for vowels versus consonants", *Proc. EUROSPLLCIl '91,* Genova. Italy. Vol. 3, pp. 1451 1454.

B. Rakerd. R.R. Verbrugge and D.P. Shankweiler (1984). "Monitoring for vowels in isolation and in a consonantal context". *J. Acoust Soc. Amer.,* Vol. 76, pp. 27 31.

W. Strange. T.R. Edman and J.J. Jenkins (1979), "Acoustic and phonological factors in vowel identification". *J. Experimental Psychology: Human Perception & Performance,* Vol. 5, pp. 643 656.

W. Strange, R.R. Verbrugge, D.P. Shankweiler and T.R. Edman (1976). "Consonantal environment specifies vowel identity". ./. *Acoust. Soc. Amer.,* Vol. 60, pp. 213 224.