Speeded detection of vowels:
A cross-linguistic study

ANNE CUTLER
MRC Applied Psychology Unit, Cambridge, England
and Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands

BRIT van OOIJEN
MRC Applied Psychology Unit, Cambridge, England
and Laboratoire de Sciences Cognitives et Psycholinguistique, Paris, France

DENNIS NORRIS
MRC Applied Psychology Unit, Cambridge, England

and

ROSA SÁNCHEZ-CASAS
Universidad Rovira I Virgili, Tarragona, Spain

In four experiments, listeners' response times to detect vowel targets in spoken input were measured. The first three experiments were conducted in English. In two, one using real words and the other, nonwords, detection accuracy was low, targets in initial syllables were detected more slowly than targets in final syllables, and both response time and missed-response rate were inversely correlated with vowel duration. In a third experiment, the speech context for some subjects included all English vowels, while for others, only five relatively distinct vowels occurred. This manipulation had essentially no effect, and the same response pattern was again observed. A fourth experiment, conducted in Spanish, replicated the results in the first three experiments, except that miss rate was here unrelated to vowel duration. We propose that listeners' responses to vowel targets in naturally spoken input are effectively cautious, reflecting realistic appreciation of vowel variability in natural context.

The recognition of spoken words is an extremely rapid process, and seems, to the listener, mostly effortless. Words appear to be apprehended as wholes, and we certainly do not have the impression of processing them phoneme by phoneme. Certainly, listeners *can* pay attention to phonemes; for instance, we can easily notice a speech error involving an exchange of phonemes ("woken spurds"), or realize that a speaker has consistent difficulty with a particular sound. But, in general, our attention is not devoted to the level of individual speech sounds.

Current models of spoken-word recognition have achieved considerable sophistication in simulating the time course of human recognition of spoken words; in particular, models that incorporate processes of simultaneous activation of and competition between alternative word candidates have been successful in simulating experimental results. The state of the art is represented by Shortlist (Norris, 1994), a model that can simulate spoken-word recognition in a realistically sized vocabulary (tens of thousands of words). In principle, such computational models lend themselves to integration with a front-end processor which would take real speech input (see McClelland & Elman, 1986a; Norris, 1990); in practice, the input to the models is given, for reasons of current computational expediency, in the form of phonemes or features. TRACE (McClelland & Elman, 1986b), for example, incorporates a featural level in which seven features can activate, to greater or lesser degrees, the set of phonemes on the adjacent phonemic level. Norris's Shortlist at present operates on a phonemic input.

To facilitate the eventual integration of psycholiguistic models of spoken-word recognition with speech-processing front ends, we can attempt to approximate

such simulated input more closely to the characteristics that real speech input presents to listeners. In the current models, it is assumed that there are no differences between individual phonemes in difficulty or type of processing. Yet, speech sounds can differ considerably in their acoustic form, as a result of differences in the type of articulation they involve. Some sounds are uttered with unobstructed vocal tract and are relatively continuous; some involve obstruction of the airflow from the lungs, even to the point of blocking it off entirely for a brief period; and so on. (These differences may be summed up in various phonological descriptions: distinctive features of phonemes, the sonority hierarchy.) Should the input to spoken-word recognition models be adjusted to reflect differences between individual phonemes in the type or intrinsic ease of processing?

It may seem odd to propose that some phonemes might be more difficult to process than others—why should languages have developed phonemes that lend themselves less than optimally to processing for word recognition? In fact, such intrinsic differences may be a necessary consequence of achieving optimal contrast via the human articulatory system. The primary dimension of interphonemic variation is, as mentioned above, sonority, with speech sounds varying from the very continuous (vowels) to the very punctual (stop consonants); it is true of all languages that utterances consist of a roughly alternating sequence of consonants and vowels, this alternation presumably allowing speech sounds to contrast effectively with one another. The classes that contrast most effectively, however, may differ in how they are best processed.

The sounds at the most sonorous end of the sonority continuum, vowels, have the attribute that they are continuously variable. A human speaker can, in fact, produce a sound that is intermediate between two vowel categories, but mostly cannot produce a sound intermediate between two consonant categories. Synthetic ambiguous sounds of all phonemic types of course can be, and repeatedly have been, constructed for use in speech-perception experiments, but the fact remains that at the vowel end of the continuum we can easily carry out this exercise ourselves and at the consonant end of the continuum we cannot. This simple fact prompted a great deal of research in phoneme perception, which in turn motivated for some time the claim that vowels and (at least) stop consonants were perceived in different ways (see, e.g., Liberman, Mattingly, & Turvey, 1972; Pisoni, 1973). Ades (1977), however, used the differences in continuous variability to argue *against* an intrinsic difference in perception. The effective range (defined in numbers of just noticeable differences) of the average continuum from one clear vowel exemplar to another, he argued, is larger than the range from any one consonant to another. In the psychophysical model of intensity resolution proposed by Durlach and Braida (1969), size of continuum range is inversely correlated with accuracy of identification performance. Ades pointed out that the larger range of vowel continua would on this model produce more variable identifications for vowels than for consonants, and this variability would result in vowels having less clearcut identification and discrimination functions than consonants.

Identification and discrimination tasks are the classic methodology of speech-perception studies. In the case of vowels, listeners may be asked to *classify* a sound as a member of a particular category (e.g., Andruski & Nearey, 1992; Strange, 1989b); or it may be determined whether they can *discriminate* between two sounds belonging to the same or different categories (e.g., Gottfried, Jenkins, & Strange, 1985; Schouten & van Hessen, 1992); or detection thresholds for different sound categories may be measured as a function of context (e.g., Rakerd, Verbrugge, & Shankweiler, 1984) or at varying intensity levels (e.g., Kewley-Port, 1991). These tasks have the valuable property of forcing a response from the subject, even if it is only a best guess. They are, however, poor measures of ease of processing, since this must be indirectly inferred from relative response variability. Thus, for the psycholinguist, who is interested in the processing of speech sounds in natural situations (such as how they function in word recognition), such tasks cannot supply the full picture; what is further required is some method of tapping the processing of phonemes on-line and assessing ease versus difficulty of processing more directly.

Among the methods available to psycholinguistics is the phoneme detection task (Foss, 1969), in which listeners are presented with speech input and have to press a response key as fast as they can when they hear an occurrence of a prespecified phoneme target. The experimental variable is the speed with which the listener presses the response key to signal that detection has occurred. Phoneme detection is not exactly a natural task, of course, but it has several advantages for the present objectives. First, in requiring a speeded response, it is "on-line"—the listener has no time to engage in conscious deliberation. Second, it gives a direct measure of relative ease/difficulty via relative response time. Third, since it does not force a response, the miss rates can be further informative and, in fact, offer a separate measure of processing ease versus difficulty. And finally, by requiring subjects to detect a prespecified target, phoneme detection encourages them to set a criterion for response; just as response criteria in identification and discrimination tasks can be manipulated by altering factors in the experimental design, so too can such manipulation be effective in phoneme detection, as many studies have shown (see, e.g., Newman & Dell, 1978; Norris & Cutler, 1988).

In practice, phoneme detection has been used more as a tool to investigate lexical and sentential processing than as a window on phonetic processing (see Cutler & Norris, 1979, for a review), and as a result the choice of which phonemes to use as targets for detection has generally been made on the basis of convenience. Most de-

tection experiments have used stop-consonant targets, and prior to our work there appeared to be virtually no phoneme-detection results available for vowels. A study by Dijkstra, Schreuder, and Frauenfelder (1989), which measured choice response time (yes–no decisions) to signal which of two vowels had occurred, used extremely simple materials (CV syllables), and the principal independent variable was manipulation of a visually presented letter representing a consonant; this experimental design is unrepresentative of phoneme-detection tasks. Two older findings, however, seemed to suggest that detection of vowels is difficult. First, in a control experiment reported by Mehler, Dommergues, Frauenfelder, and Segui (1981), response times (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. Second, post hoc analysis of the results of a study by Hakes (1971), in which vowels as well as consonants were used as (word-initial) targets, showed likewise that RTs to vowel targets were significantly longer than RTs to stop [b,d,g,p,k], nasal [m,n], or glide targets [l,w]. More recently, both Cutler and Otake (1994) and van Ooijen (1994) have reported slower phoneme detection RTs from English-speaking subjects for vowel targets than for certain consonant targets.

In the present study, we used the phoneme-detection task to investigate in greater detail the on-line processing of vowels. The characteristics of the task allowed us to ask whether natural tokens differed in difficulty of processing, for instance as a function of whether they occurred earlier versus later in a word, or in a stressed versus unstressed syllable. That is, we assumed that categorization tasks would encounter no problem in eliciting the same response to, say, the vowel in the first syllable of *tenant* and *technique* and in the second syllable of *nutmeg* and *condemn*; discrimination tasks would encounter no problem in saying that those vowels were all totally different from the vowel common to *carton*, *cartoon*, *placard*, *discard*. But, in order to improve our modeling of how vowels were processed in natural word tokens, it would be useful to know whether, for instance, there was a difference between the ease with which those two vowels were detected or a difference in the ease with which the same vowel was processed near the beginning of a word (*tenant*, *technique*, *carton*, *cartoon*) versus near the end (*nutmeg*, *condemn*, *placard*, *discard*) or in a syllable that bore primary stress (*tenant*, *condemn*, *carton*, *discard*) versus one that did not (*technique*, *nutmeg*, *cartoon*, *placard*). If we were to find differences of this kind, we would further like to know whether they might be explained by characteristics of vocabulary structure (such as differences in predictability of syllable types; such effects should then also be expected to appear in appropriate large-vocabulary simulations) or whether they were due to extralexical factors (in which case we might improve our simulations by modeling them in the input). Phoneme detection would offer us the ability to answer all these questions.

Our first choice concerned which vowels to use as targets, given the possibility that vowels might differ among themselves in ease of processing. To be sure, we found no indication in the previous phoneme-detection literature that phoneme targets within a manner of articulation category differed in how difficult they were to detect. For detection of stops in initial position, there are claimed to be no differences in absolute RTs to the six stops [p,t,k,b,d,g] (Martin, 1977). Several researchers have reported that fricatives produce slower RTs than do stops (Foss & Swinney, 1973; Morton & Long, 1976; Rubin, Turvey, & van Gelder, 1976; Savin & Bever, 1970), but a recent series of experiments, using a mixture of stop, fricative, nasal, and glide targets (Pitt & Samuel, 1990), reported no significant interphoneme RT differences. We found no interphoneme comparisons of miss rates (which are typically low—5% or less—in phoneme-detection experiments with consonant targets). Nevertheless, we preferred to control the nature of the vowel target explicitly, and we chose to investigate five targets. This allowed us to keep the experiment to a manageable length (since subjects might become fatigued if testing continued for too long in any speeded response task) while still yielding a substantial number of responses per vowel target.

Because a series of studies from our laboratory (e.g., Cutler & Butterfield, 1990, 1992; Cutler & Norris, 1988) had shown that the strong–weak vowel distinction in English was exploited by listeners in speech segmentation, a comparison between strong (full) and weak (reduced, central) vowels was of interest to us. Thus, one of our chosen targets in Experiment 1 was the weak vowel schwa. The four remaining full vowels in our experimental target set were chosen according to several criteria. First, we excluded diphthongs; in a phoneme-detection task, it could be argued that a diphthong constituted a multiple target. Second, since we wanted to test vowels in first and second syllables and with differing levels of stress, we chose vowels with a high frequency of occurrence in the English vocabulary in order to ensure that we had sufficient test words. Third, we attempted to provide a reasonable range of intrinsic durations. Schwa is much shorter than full vowels, so a difference between schwa and full vowels could be interpreted as an effect of either centrality or duration; to decide this issue, it would be necessary to have an independent yardstick of duration effects. Measurements of vowel durations were available for the materials used by Cutler, Mehler, Norris, and Segui (1987), and a comparable set of measurements were available to us for American English from a study by Diehl, Kluender, Foss, Parker, and Gernsbacher (1987). The four full vowels we chose (/a/, /ɛ/, /ʌ/, and /ɪ/) ranked, respectively, 1st (longest), 5th, 8th, and 10th in our set of measurements of 10 British English vowels, and 3rd, 5th, 8th, and 10th in the American English set of 10.

We reduced the length of the experiment by presenting the materials in blocks such that subjects listened for one target across a block of 44 trials, and then changed to another target for the next block; this saved the time taken by target specification when targets vary from trial

to trial. Finally, all the vowel targets occurred in real words, in medial position. Frauenfelder and Segui (1989; Frauenfelder, Segui, & Dijkstra, 1990) showed that phoneme targets in medial position afford the best conditions for the display of lexical effects if these exist; in the present case, lexical effects are most likely to be observed for targets occurring late in the word (*nutmeg*, *condemn*, *placard*, *discard*).

In Experiment 1, therefore, we asked whether response times and miss rates to vowels would show evidence of processing difficulty (as suggested by Ades's, 1977, interpretation of the identification and discrimination results, and by the skimpy evidence from previous phoneme-detection studies); whether responses were affected by position of the target within the word and by syllable stress level; whether there were differences in ease of processing between schwa and full vowels; and whether vowel duration played a role in vowel detection.

EXPERIMENT 1

Method

Materials. Five vowels served as targets: the full vowels /a/, /ɛ/, /ɪ/, and /ʌ/, and the reduced vowel /ə/ (schwa). One hundred twenty disyllabic words (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*, *PLAcard*, *carTOON*, *disCARD*; upper case denotes syllabic stress. Note that Southern British English is a nonrhotic dialect; thus these four example words have the structure CVCVC, CCVCVC, CVCVC, and CVCCVC, respectively). 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., *FALcon*, *conFUSE*). The 120 experimental items are listed in the Appendix. Within each set, the words were matched for frequency (Johanson & Hofland, 1989) and, where possible, for phonemic environment. Fifty further mono- and disyllabic words, 10 for each vowel set, were dummy target items, and 1,000 words of one, two, or three syllables were filler items. Except for a few words containing schwa, no filler items contained any target vowel. A practice set was also constructed; the target vowel for this set was /əʊ/ as in *oatmeal*, *poker*, *corrode*.

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, 10 lists contained a dummy target in first or second position, and 10 lists contained no occurrence of the target. Order of occurrence and position in list of items in each stress and target-syllable condition were matched across the blocks. The number of syllables and stress pattern of the word immediately preceding each target item was also matched across blocks. In each block, 5 warm-up lists preceded the first list containing an experimental target. The practice set contained 16 lists, 4 of which had no occurrence of the target. The five experimental blocks plus the practice set and a small set of example words were recorded by a male native speaker of British English, with no obvious regional dialect. The lists were recorded at a rate of one word per 1.5 sec, with 3 sec between lists.

The experimental tapes were presented in five different orders. (The blocks for /ʌ/ and schwa, which are acoustically similar, were never adjacent.) Each block lasted approximately 6 min.

Subjects. Twenty-five students at Pembroke College, Cambridge, served as paid volunteers for the experiment. All were native speakers of British English with normal hearing. Five subjects heard each order of presentation of the experimental tapes.

Procedure. The subjects were tested individually in a quiet room; they listened to the tapes over headphones, and were given written instructions in which they were asked to press the response key as soon as they heard an occurrence of the specified vowel. Before each block, the subjects heard examples of words containing the appropriate target. Response timing was initiated by marks aligned with the onset of experimental words, inaudible to the 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 using the waveform editor CAMSED. Vowel onset and offset were determined by a combination of visual and auditory methods; the cursor was placed at the point where, in the judgment of the measurer (the second author, for all four experiments, but jointly with the fourth author in the case of Experiment 4), the transition occurred, in either direction, between perceptible vowel information and no perceptible vowel information.

Results and Discussion

RTs longer than 1,500 msec or shorter than 100 msec were discarded (this resulted in the loss of 2.03% of all responses). The RTs were adjusted for measured timing mark displacement to give RTs from target-vowel onset. Two analyses of variance were conducted, with subjects and words as random factors; we report only effects significant in both. Table 1 shows mean response times and mean percentage responses missed for each vowel in first- versus second-syllable position.

The overall mean RT was 606 msec. The main effect for vowel was significant [$F1(4,80) = 23.99$, $p < .001$; $F2(4,100) = 21.12$, $p < .001$]; Newman-Keuls post hoc comparisons showed that response times were significantly slower to schwa than to the other four vowels, responses were significantly slower to /ʌ/ than to the other three full vowels, while response times to /a/, /ɛ/, and /ɪ/ did not differ significantly.

There was also a main effect of the order in which the subjects had heard the five vowel blocks [$F1(4,20) = 3.68$, $p < .05$; $F2(4,400) = 61.37$, $p < .001$], but this between-subjects effect did not interact with any within-subjects variable (in particular, the vowel factor) and presumably represents a simple difference in mean RT across the subject groups.

Table 1
Mean Response Time (in Milliseconds) and Proportion of Missing Responses for the Five Vowel Targets of Experiment 1 as a Function of Position in First Versus Second Syllable

Vowel	First Syllable		Second Syllable	
	Response Time	Responses Missed	Response Time	Responses Missed
/a/	565	3%	547	10.3%
/ɛ/	581	14%	490	13%
/ɪ/	594	10%	507	8.7%
/ʌ/	675	12.3%	599	15%
/ə/	776	63.7%	729	53%

Vowels in first syllables were detected significantly more slowly than vowels in second syllables [$F1(1,20) = 48.19, p < .001; F2(1,100) = 30.35, p < .001$]. An analysis excluding schwa was conducted to investigate the effect of the stress-pattern factor. RT to full vowels was faster when the vowel bore primary (e.g., *CARton*, *disCARD*; 550 msec) as opposed to secondary stress (e.g., *carTOON*, *PLAcard*; 589 msec; $F1(1,20) = 14.72, p < .001; F2(1,80) = 5.76, p < .05$]; however, a significant interaction with the syllable-position factor [$F1(1,20) = 42.99, p < .001; F2(1,80) = 10.37, p < .01$] showed that this difference was significant in first syllables (83-msec difference) but not in second syllables (6-msec difference).

The number of targets missed was high, at 20.3% of all targets in the experiment; but again there was a significant difference across the five vowels [$F1(4,80) = 108.16, p < .001; F2(4,100) = 90.16, p < .001$]. Newman-Keuls post hoc comparisons revealed that schwa was missed significantly more often (58%) than any of the four full vowels, which did not differ among themselves (range: 6.7%–13.7%).

The order and syllable-position effects were not significant in the miss-rate analysis, and there were no interactions between the variables. An analysis excluding schwa showed that full vowels were missed significantly more often in secondary-stressed syllables (14.3%) than in primary-stressed syllables [7.2%; $F1(1,20) = 15.16, p < .001; F2(1,80) = 13.7, p < .001$].

A correlation analysis showed that the longer the duration of the vowel, the faster it was detected [$r(119) = -.47, p < .001$]; 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 [$r(95) = -.34, p < .001$]. It also held separately for targets in first [$r(59) = -.47, p < .001$] and second syllables [$r(59) = -.4, p < .01$]. An analysis of variance on the vowel duration measurements showed a highly significant difference between vowels in first syllables (mean duration 98 msec; 105 msec without schwa) and vowels in second syllables [mean duration 140 msec; 158 msec without schwa; $F(1,100) = 147.38, p < .001$]. There was no correlation between RT and the overall duration of the words in which the target vowels occurred. As the similarity between the pattern of results in the RT and missed-response analyses would suggest, these two response measures were correlated: items that were responded to slowly were also missed more often [$r(119) = .60, p < .001$]. Thus, also, the correlations of vowel duration with RT were mimicked by correlations of vowel duration with miss rate [$r(119) = -.46, p < .001$, overall; $r(95) = -.22, p < .05$, without schwa; $r(59) = -.46, p < .001$, for first syllables; $r(59) = -.57, p < .001$, for second syllables].

The results of this experiment are somewhat surprising in two respects. First, the miss rates found here were higher than those normally found in phoneme-detection experiments—even /a/, the longest vowel and the most accurately detected, was missed about 7% of the time. There was no speed–accuracy tradeoff—the vowels most often missed were also responded to most slowly. This finding certainly supports our assumption that vowels are not very easy to detect in a speeded response task.

Second, the inverse relationship between target-vowel duration and RT clearly warrants further investigation. It, too, suggests processing differences between vowels and consonants. Diehl et al. (1987) report that consonant detection time was *positively* correlated with the duration of the following vowel: the longer the vowel, the slower the preceding consonant was detected.

To interpret these findings, however, it is important to know at what level subjects in the present study were responding, given that responses in phoneme-detection tasks may be made pre- or postlexically, depending on characteristics of the experimental situation (Cutler et al., 1987). The RT advantage for targets in second syllables may reflect lexical predictability, suggesting that a significant proportion of responses may have been postlexical; in similar tasks requiring postlexical responses (e.g., detection of a mispronounced phoneme), RT decreases steadily across the word (Marslen-Wilson & Welsh, 1978). The added difficulty of schwa compared with full vowels could also reflect lexical involvement, since the orthographic mapping of schwa was less consistent than the mapping of the other four vowels in the experimental words (the vowels /ɛ/, /ɪ/, or /ʌ/ all had constant representations, and in all but three words /a/ was represented by "ar"; schwa, however, was orthographically represented in our word set in four different ways, with "e" as the most common representation, in 9 of 24 items).[1]

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 nonwords, which have no lexical representations. We therefore conducted a follow-up experiment, matched to Experiment 1 except in that the target vowels occurred in nonwords.

EXPERIMENT 2

Method

Materials and Design. Using the same target vowels, the number of items constructed was the same 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*, *DROlart*, *larTOACE*, *poLART*; for schwa: *CLYpen*, *penZINE*. The 120 experimental items are listed in the Appendix. Construction of the lists and practice set was as in Experiment 1, and the materials were recorded by the same speaker.

Subjects. Twenty-five students from Downing and Selwyn Colleges, Cambridge, were paid for participating. All were native speakers of British English with normal hearing. Five heard each order of presentation of the experimental tapes.

Procedure. The procedure was as for Experiment 1.

Results and Discussion

RTs were adjusted and analyzed as in Experiment 1. Responses shorter than 100 msec or longer than 1,500 msec

Table 2
Mean Response Time (in Milliseconds) and Proportion
of Missing Responses for the Five Vowel Targets of
Experiment 2 as a Function of Position in
First Versus Second Syllable

	First Syllable		Second Syllable	
Vowel	Response Time	Responses Missed	Response Time	Responses Missed
/a/	601	7.3%	533	13.3%
/ɛ/	596	12.3%	543	22.7%
/ɪ/	545	11.7%	531	14.3%
/ʌ/	607	26.7%	602	29%
/ə/	749	59.3%	696	58.3%

were again discarded (these amounted to 1.83% of all responses in this experiment).

Table 2 shows the mean response times and percentage of missed responses as does Table 1 for Experiment 1. The mean RT across all conditions was 601 msec. The main effect of vowel was again significant [$F1(4,80) = 13.27, p < .001; F2(4,100) = 5.93, p < .001$]. Newman-Keuls post hoc comparisons showed that schwa was responded to significantly more slowly than were the four full vowels, which did not differ significantly among themselves. The order effect did not reach significance. There was again a syllable-position effect: targets in first syllables were detected significantly less rapidly than were targets in second syllables [$F1(1,20) = 16.47, p < .001; F2(1,100) = 11.58, p < .001$]. An analysis excluding schwa revealed a marginally significant advantage for full-vowel targets in syllables with primary stress (557 msec) over syllables with secondary stress [583 msec; $F1(1,20) = 12.53, p < .01; F2(1,80) = 3.55, p < .07$]. There were no significant interactions between the variables.

The miss rate was again high (25.5%). Again, there was a significant difference between the five vowels [$F1(4,80) = 63.11, p < .001; F2(4,100) = 70.05, p < .001$]; Newman-Keuls post hoc analyses showed that schwa was missed significantly more often (58.8%) than the four full vowels, and /ʌ/ was missed significantly more often (27.8%) than the other three vowels, which did not differ significantly (range 10.3%–17.5%). An analysis excluding schwa showed that, again, vowels were missed less often in primary-stressed syllables (10.9%) than in secondary-stressed syllables [23.4%; $F1(1,20) = 70.64, p < .001; F2(1,80) = 27.76, p < .001$]. Furthermore, this stress effect interacted with syllable position, being three times as large in second syllables than in first syllables [$F1(1,20) = 12.52, p < .01; F2(1,80) = 6.05, p < .05$], and with vowel, being large for /ʌ/ and /ɪ/ but small for /a/ and /ɛ/ [$F1(3,60) = 18.71, p < .001; F2(3,80) = 4.35, p < .01$].

Just as in the previous experiment, there was no speed–accuracy tradeoff; vowels that were responded to slowly were also missed more often. There was a significant positive correlation between the two dependent variables [$r(119) = .39, p < .001$]. Correspondingly, measured vowel duration again correlated negatively with both RT

[$r(119) = −.31, p < .001$] and miss rate [$r(119) = −.45, p < .001$], and the correlations also held for the four full vowels alone [RT, $r(95) = −.22, p < .05$; miss rate, $r(95) = −.23, p < .05$]. There was again no correlation with overall item length. An analysis of variance on the vowel duration measurements again revealed a highly significant difference between vowels in first syllables (mean duration, 99 msec; 107 msec without schwa) and vowels in second syllables (mean duration, 145 msec; 163 msec without schwa; $F(1,100) = 134.64, p < .001$).

Thus, the results of Experiment 2 closely replicate those of Experiment 1. Again, there was a high miss rate and an inverse relationship between duration and both RT and miss rate. Again, schwa was the most difficult vowel to detect, and the present result with nonword materials suggests that the difficulty of schwa could be purely acoustic–phonetic rather than due to lexically mediated responding. Again, vowels in the first syllable of bisyllables took longer to detect than vowels in second syllables.

The fact that this syllable-position effect appeared at all in Experiment 2 with nonwords suggests that it is not, or at least not wholly, an effect of increasing lexical predictability across the word. The most obvious candidate nonlexical explanation for a syllable-position effect is that it is an artifact of the tendency in English for word-final syllables to be lengthened combined with the negative correlation that we found between measured vowel duration and RT. If so, then the syllable-position effect should disappear in an analysis of covariance across items in which measured vowel duration is used as the covariate. (If the effect is multiply determined, i.e., is in part due to final lengthening and in part to increasing lexical predictability across the word, then an analysis of covariance could of course succeed in removing the syllable effect with nonwords but not with words.)

Note that another effect in our results could also in principle follow simply from the inverse correlation between vowel duration and RT; in both experiments, RTs were significantly longer to schwa than to any of the other vowels, and measured duration of schwa was shorter than measured duration of any other vowel. Thus, an analysis of covariance in which measured vowel duration is used as the covariate might also remove the main effect for vowel type.

We conducted such an analysis of covariance separately for each experiment. For Experiment 1, the main effect of vowel type remained significant [$F(4,99) = 12.68, p < .001$], as did the syllable-position effect [$F(1,99) = 8.68, p < .005$]. Both of these effects also remained in an analysis without schwa [$F(3,99) = 5.32, p < .01$, and $F(1,99) = 7.84, p < .01$, respectively]. In the latter analysis, however, the main effect of primary versus secondary stress disappeared, although the interaction between stress and syllable position remained [$F(1,99) = 10.2, p < .005$].

For Experiment 2, the main effect of vowel type again remained significant [$F(4,99) = 4.78, p < .001$, in the analysis with schwa, $F(3,99) = 3.25, p < .05$, in the anal-

ysis without schwa]. The syllable-position effect, however, did not reach significance in either analysis, and the stress factor in the analysis without schwa was also not significant.

The syllable-position effect therefore cannot be a unitary effect. Our combined results suggest that it is multiply determined by increased lexical predictability of targets in second syllables (in words) and by final-syllable lengthening leading to vowels with longer duration (in words and nonwords). Note that in both experiments correlations with duration (the longer the vowel, the faster the response time or the fewer the targets missed) were separately significant in first and in second syllables.

What causes vowels to be easier to detect the longer they are, and what can this finding tell us about the processing of vowels in spoken-word recognition? One simple account of the present data would be to assume that information was accumulated and processed throughout the course of a vowel. (Vowels differ here from most consonants in having a relatively even distribution of perceptual information; we would not expect this argument to be directly applicable to the most consonant-like sounds of speech.) Incremental accumulation of evidence would continue until the criterion for a detection response had been reached. However, if the vowel had not been identified by the end of the incoming vocalic information, further perceptual processing could continue for some time, albeit at a reduced rate if subsequent incoming phonetic information was also undergoing processing. According to this suggestion, long vowels would be likely to be identified both quickly and accurately if sufficient evidence was accumulated to permit a response to be initiated either before or shortly after the vowel terminated. Short vowels, on the other hand, would take longer to identify because more of the perceptual processing would take place at the slower rate after the vowel had terminated. Short vowels would also be less accurately identified because less perceptual evidence would be accumulated in the available time. This account raises the possibility that the effects of vowel length might be attenuated, and possibly even eliminated, if we could make the task of discriminating between vowels easier. Subjects might then be able to adopt a lower response criterion, which would permit vowel-detection responses to be made on the basis of less accumulated evidence, with the result that even short vowels could be identified before they terminated.

In Experiment 3, therefore, we explicitly attempted to manipulate subjects' response criterion. In Experiments 1 and 2, we had used target vowels that were sufficiently similar to one another for intrinsic distinctiveness within the experiment to be low; this in itself could have encouraged adoption of a cautious response criterion. Therefore, in this third experiment, we used a five-vowel target set which attempted to maximize distinctiveness within the English vowel repertoire. Of course, the English vowel space is quite densely populated, and this in itself may cause English vowel types to have inherently low distinctiveness. (The results of Mehler et al., 1981,

mentioned above, suggest that vowel detection is difficult in French; the vowel space of French is populated with a density very similar to that of English.) It is then possible that the effective vowel repertoire in the experiment is actually not defined by the target set but by all the vowels in any item (including the filler items) spoken in the experiment. Accordingly, in Experiment 3, we further contrasted detection of the five relatively distinct vowels in two filler contexts, one that ranged over the entire English vowel repertoire and one that was constrained to the relatively distinct set represented by the target vowels. If subjects are sensitive either to the relative distinctiveness of the target set itself or to the relative distinctiveness of the experiment's vowel repertoire, then we should observe that they relax their response criterion in Experiment 3.

EXPERIMENT 3

Method

Materials. Target phonemes were five full British English vowels /a/ ("barb"), /ɛ/ ("bet"), /i/ ("beat"), /ɒ/ ("bob"), and /u/ ("boot"). (The lax vowels /ɛ/ and /ɒ/ were chosen rather than the tense vowels /eɪ/ ["bait"] and /əu/ ["boat"] because the latter two are diphthongs.)

One hundred twenty monosyllabic and disyllabic nonwords were constructed. There were 24 target-bearing nonwords for each target phoneme, of which 8 were monosyllabic, 8 were disyllabic with the target in first syllable position (e.g., for /a/ *FRARdock*), and 8 disyllabic with the target in second syllable position (e.g., *deFARN*). A comparison between levels of stress was not included in this experiment; all targets occurred in stressed syllables. Thus the bisyllables with first-syllable targets were also stressed on the first syllable (SW), and the bisyllables with second-syllable targets were stressed on the second syllable (WS). The 120 experimental items are listed in the Appendix. A further 40 nonwords, 10 per target phoneme, served as dummy target items. About 2,000 nonwords served as fillers. Half of these were constructed with no constraint, so that they could contain any English vowel(s) (unrestricted fillers). The other half contained only the five vowels which were also used as target phonemes (restricted fillers). A practice set was also constructed; in order not to introduce additional vowel stimuli, the target phoneme for the practice set was /l/.

Experimental design. The materials were arranged in 10 blocks, 2 for each target vowel. Each block consisted of 44 lists of two to six nonwords in length; of these, 24 lists contained an experimental nonword in the penultimate (third, fourth, or fifth) position, 10 lists contained a dummy target item in first or second position, and 10 lists contained no occurrence of the target. For each target vowel, 1 block had unrestricted fillers (the unrestricted condition). In unrestricted blocks, the fillers could contain any English vowel (except of course the target vowel for that block). The other block had restricted fillers (the restricted condition); fillers contained only the four vowels from the restricted set remaining once the vowel that served as target phoneme in that particular block had been excluded. For example, in the block where /a/ was the target vowel, the fillers contained only /ɛ/, /i/, /ɒ/, or /u/.

The blocks plus the practice set and a small set of example nonwords were recorded by the speaker used in Experiments 1 and 2. As before, target-phoneme length and target-bearing nonword length were measured using a waveform editor. The measurements revealed that both vowels and nonwords in the restricted blocks were on average slightly longer than their counterparts in the unrestricted blocks. This was probably due to the fact that limiting the speaker to the five vowels in the restricted blocks had caused him

to speak more carefully and, in consequence, more slowly. To ascertain whether this difference would affect response time, a third hybrid control condition was constructed; this contained the blocks from the unrestricted condition with half of the experimental nonwords in each block replaced by their restricted counterparts.

Subjects. Sixty-five students of Jesus College, Cambridge, were paid to take part in the experiment. The data of 5 subjects were lost due to equipment failure. Of the remaining 60 subjects, 20 were assigned to each of the three conditions. Four subjects in each condition heard each order of presentation.

Procedure. The procedure was as in Experiments 1 and 2.

Results and Discussion

RTs shorter than 100 msec or longer than 1,500 msec were again discarded (this resulted in the loss of 0.67% of the data). Mean response times and mean proportion of missing responses are shown in Table 3.

Analyses of variance were conducted as in Experiments 1 and 2. The overall mean response time was 593 msec. Mean response times to the five vowels varied from 535 msec to /u/ to 645 msec to /ɒ/; the main effect of vowel identity was significant [$F1(4,180) = 26.6$, $p < .001$; $F2(4,105) = 13.39$, $p < .001$]. Newman-Keuls post hoc analyses showed that RTs to /u/ and /i/, which did not differ, were faster than RTs to the other three vowels, while RTs to /ɒ/ and /ɛ/, which did not differ, were also slower than RTs to /a/. Mean response time in the unrestricted condition, in which filler items could contain any English vowel, was 624 msec, while in the restricted condition, in which filler items could contain only the five distinct vowels which also served as targets, mean response time was 575 msec. (Mean response time in the hybrid control condition was 580 msec; recall that this condition contained the unrestricted filler set but a mixture of targets from the restricted and unrestricted conditions, so that this result suggests that the apparent RT advantage of the restricted filler set was, in fact, due to small differences between the targets themselves rather than to filler-set characteristics.) The main effect of condition was, in fact, not significant.

The only other main effect that reached significance in both analyses was the effect of nonword item structure [$F1(2,90) = 17.47$, $p < .001$; $F2(2,105) = 4.22$, $p < .05$]. Post hoc analyses of this effect showed that response times to targets in monosyllables (574 msec) were significantly faster than response times to targets in the first syllable of SW bisyllables [609 msec; $t1(59) = 6.41$, $p < .001$; $t2(78) = 2.55$, $p < .05$], but neither differed significantly from response times to targets in the second syllable of WS bisyllables (596 msec).

The mean number of missed responses was 9%. Analysis of the missed responses showed significant differences between the five vowels [$F1(4,180) = 9.03$, $p < .001$; $F2(4,105) = 12.58$, $p < .001$]; Newman-Keuls post hoc analyses showed that there were significantly fewer missed responses to the three tense vowels /a/ (5.1%), /i/ (6.2%), and /u/ (6.9%), which did not differ, than to the two lax vowels /ɛ/ (14.7%) and /ɒ/ (12.3%), which again did not differ. The effect of nonword item structure was again significant [$F1(2,90) = 21.18$, $p < .001$; $F2(2,105) = 11.64$, $p < .001$]; post hoc analyses showed that performance was significantly better for monosyllables (5.5% misses) than for either the initial syllables of SW bisyllables (10.3%) or the second syllable of WS bisyllables (11.3%), which did not differ. Item structure interacted weakly with the vowel-identity variable [$F1(8,360) = 6.46$, $p < .001$; $F2(8,105) = 2.0$, $p <$

Table 3
Mean Response Time (in Milliseconds) and Proportion of Missing Responses for the Five Vowel Targets of Experiment 3, as a Function of Item Structure and Filler Context

Vowel	Monosyllable		First Syllable		Second Syllable	
	Response Time	Responses Missed	Response Time	Responses Missed	Response Time	Responses Missed
			Restricted Fillers			
/a/	565	1.3%	614	7.5%	573	1.3%
/ɛ/	626	8.1%	595	12.5%	589	17.5%
/i/	520	5.0%	566	11.2%	584	5.6%
/ɒ/	614	5.6%	641	12.5%	629	17.5%
/u/	491	2.5%	532	10.0%	492	3.1%
			Unrestricted Fillers			
/a/	601	2.5%	669	10.6%	595	6.9%
/ɛ/	682	18.1%	690	18.8%	691	32.5%
/i/	551	5.0%	570	6.2%	564	12.5%
/ɒ/	630	7.5%	701	17.5%	690	16.2%
/u/	558	4.4%	590	15.6%	579	11.9%
			Hybrid Fillers			
/a/	555	4.4%	616	6.2%	596	5.0%
/ɛ/	594	5.6%	569	5.0%	645	14.4%
/i/	531	2.5%	557	3.1%	560	4.4%
/ɒ/	595	7.5%	670	11.9%	640	14.4%
/u/	500	1.9%	550	6.2%	519	6.9%

.055]; although monosyllables produced fewest missed responses to all five vowels, the small difference between SW and WS bisyllables differed across vowels, with /a/ and /u/ producing slightly fewer missed responses in WS than in SW, the other three vowels slightly more.

In the unrestricted condition, 12.4% of all responses were missed, as opposed to 8.1% in the restricted condition and 6.6% in the hybrid control condition; however, the difference between conditions again did not reach significance, and the condition variable did not interact with the vowel- or item-type variables.

Finally, a correlation analysis (conducted on the RTs from the restricted and unrestricted conditions only, excluding the control) revealed that again the pattern of RT and missed-response results correlated across items [$r(239) = .48$, $p < .001$], and there was an inverse relationship between vowel duration and RT [$r(239) = -.22$, $p < .001$] and consequently also between vowel duration and miss rate [$r(239) = -.20$, $p < .01$]. Moreover, the manipulation of experimental context had no effect on these relationships (the correlation between RT and miss rate was positive and significant at the .001 level in both the restricted and unrestricted conditions, and, indeed, in each of the three item types: monosyllables, SW bisyllables, and WS bisyllables). As expected, the vowel durations both were longer in this experiment than in Experiments 1 and 2 (in part perhaps because of the speaker's care in pronouncing the materials set and in part due to the different vowel target set) and differed significantly across item structure (245 msec for monosyllables, 157 msec for first syllables, and 257 msec for second syllables; $F(2,210) = 97.09$, $p < .001$].

The results of this experiment suggest that vowel-detection performance is not affected by restricting the set of targets to vowels that are distinct from one another. This manipulation did have one effect: the major difference between Experiments 1, 2, and 3 was that in the present study fewer targets were missed. Lax vowels were missed more often than tense vowels; since a majority of target vowels in Experiments 1 and 2 were lax, it would seem likely that the high miss rates in those earlier experiments were in part caused by the particular selection of targets.

Response times in the present experiment were, however, comparable to those in Experiments 1 and 2. Thus, the detection task was not easier in the present experiment. Again, the longest response times were found in the set of items with the highest rate of missed responses, suggesting that there was no speed–accuracy tradeoff. The range of vowels occurring in filler items had no effect on the pattern of response times or on the pattern of misses—in other words, discrimination was apparently not easier when the available repertoire was limited to distinct candidates. Finally, the inverse correlation of target duration and RT, though slightly weaker in this study, was still apparent.

It would appear, therefore, that the manipulations of vowel repertoire in Experiment 3 did not succeed in sig-

nificantly altering subjects' response criterion. Higher distinctiveness of the target set did reduce the number of missed targets from around 11% for the four full vowels of Experiment 1 and 17% for the four full vowels of Experiment 2 to 9% in Experiment 3. This certainly is consistent with a less cautious response criterion, although the inverse correlation of RT and miss rate with vowel duration was still observed. The repertoire restriction manipulation had, however, no effect. This suggests that English listeners cannot usefully restrict their effective available vowel space in such a way as to facilitate target detection.

It is possible to investigate the effect of vowel repertoire from a different angle, however. We may ask, for instance, whether vowel-detection responses pattern differently in a language with a dense vowel space than they do in a language in which the vowel space is more sparsely populated. Spanish is such a language: it has only five vowels, and they occupy highly distinct positions in the vowel space. The restricted condition of Experiment 3 was essentially an analogue of the situation in Spanish; but the listeners in Experiment 3 were native speakers of the vowel-rich language of English. Listeners whose language has accustomed them to only a few, distinct vowels may produce a quite different pattern of vowel-detection responses. Our final experiment, therefore, was analogous to Experiment 1, except in that it involved Spanish materials and Spanish listeners.

EXPERIMENT 4

Method

Materials. The five target vowels used were the five vowels of Spanish: /a/, /e/, /i/, /o/, /u/. One hundred and twenty bisyllabic words (nouns, verbs, and adjectives) were chosen, 24 for each of the five target vowels. For each vowel, each set of 24 items consisted of six quartets. Per quartet, the target vowel occurred in two words with initial stress, once in the first syllable (example for /a/: *nave*), once in the second syllable (example for /a/: *mina*), and in two words with final stress, again once in the first syllable (example for /a/: *nariz*) and once in the second syllable (example for /a/: *reinar*). Within the quartets, the words were matched for frequency (i.e., they differed by no more than 10 occurrences per half a million words; Juilland & Chang Rodriguez, 1964) and, where possible, given the restrictions imposed by frequency, for the phonemic environment. It was not possible to match word frequency of occurrence across the five vowel sets; the words containing the vowel /a/ had the highest mean frequency, and the words with the vowel /u/ had the lowest frequency. A further 50 bisyllabic words, 10 for each vowel set, were chosen as dummy target items. An additional 1,000 words of one, two, or three syllables were used as filler items. No filler items in any set contained an occurrence of the target vowel for that set. The 120 experimental items are listed in the Appendix.

Experimental design. The experimental design was as in Experiments 1 and 2, except that the word immediately preceding each target item was always a bisyllable. The materials were recorded by a male native speaker of Castilian Spanish at the same rate as in Experiments 1 and 2. Also recorded was a short practice set, with the target sound /l/, as in Experiment 3.

Subjects. Thirty students of the Universidad Complutense, Madrid, ranging in age from 20 to 24 years, participated in the ex-

periment as part of the requirements of a course in perception. All subjects were native speakers of Castilian Spanish and had normal hearing. The responses of 6 subjects were lost due to equipment failure. Of the remaining 24 subjects, 4 heard one order of presentation of the experimental tapes and 5 heard each of the four other orders of presentation.

Procedure. The subjects were tested individually in a quiet room. They listened to the recorded materials over headphones; they were provided with written instructions in which they were asked to press a button as soon as they heard the specified vowel sound. Prior to each experimental block, the subjects were given examples of words containing the target sound for that block. Timing and data collection was the same as in Experiments 1–3. As in the previous experiments, the 120 experimental words were digitized, and word length, target-vowel duration, and the time from target-vowel onset to timing mark were measured using a waveform editor. As in Experiments 1–3, the measurements were carried out in the Cambridge laboratory by the second author, though here jointly with the fourth author, using the same software and decision criteria as had been used for Experiments 1–3.

Results and Discussion

The overall mean response time was 580 msec. No responses in this experiment were discarded, since all fell within the 100–1,500-msec range. Table 4 shows the mean response times and percentage of missed responses for each vowel in each position.

Analyses of variance were conducted in the same fashion as in Experiments 1–3. Mean response times for the five vowels varied from 550 msec (/i/) to 607 msec (/a/); the main effect of vowel identity was, however, not significant. Response times were significantly faster when the target vowel occurred in the second syllable of a word (535 msec) than when it occurred in the first [625 msec; $F1(1,23) = 103.12, p < .001; F2(1,100) = 16.81, p < .001$]. There were no effects of stress pattern.

Miss rates were lower than in Experiments 1 and 2; the overall mean percentage of missed responses was 7.7%. Significantly more targets were missed in second syllables of words (10.1%) than in first syllables [5.2%; $F1(1,23) = 30.39, p < .001; F2(1,100) = 8.82, p < .01$]. There was again no effect of stress pattern. The main effect of vowel identity was marginally significant in this analysis of missed responses [$F1(4,92) = 5.29, p < .001; F2(4,100) = 2.4, p < .055$] and post hoc analyses showed that error rates were significantly higher to /a/ targets (12.5%) than to the other four vowels, which did not dif-

fer significantly among themselves (range 5.7% to 8.2%). The higher miss rate for /a/ is puzzling, given that the English vowel /a/ was missed *least* often in Experiments 1 and 2, and that in perceptual confusion studies in Spanish, /a/ is least often misperceived (Romero, 1988).

The faster RTs but higher miss rates in second syllables suggest a speed–accuracy tradeoff. A correlation of mean response time against mean number of missed responses across all items revealed that there was indeed a negative correlation [$r(119) = -.21, p < .05$]. Thus, the parallel patterns in RT and missed-response analyses that we found in Experiments 1–3 were *not* repeated in Experiment 4.

An analysis of variance on the duration measurements showed that, just as in the English experiments, vowels in first syllables (at an average duration of 119 msec) were significantly shorter than vowels in second syllables [mean = 172 msec; $F(1,100) = 83.52, p < .001$]. Correlation analyses were again also conducted on the durational measurements and the response times and miss rate. As in the English experiments, there was an inverse correlation between vowel duration and RT [$r(119) = -.33, p < .001$]; however, no significant correlation appeared between vowel duration and miss rate [although, as expected, the relationship was in the opposite direction, that is, positive; $r(119) = .16$].

As in Experiments 1 and 2, we carried out an analysis of covariance on the response times, taking out vowel duration as a covariate. The only significant effect in the original RT analysis, the syllable-position effect, remained significant in the analysis of covariance [$F(1,99) = 6.68, p = .01$], exactly in accordance with Experiment 1, the other experiment in which the targets occurred in real words.

Experiment 4 has thus shown that vowel detection in Spanish patterns quite similarly to vowel detection in English, with the exception that missed responses in Spanish do not correlate positively with RTs, as they do in English. The replication here of the inverse relationship between vowel duration and RT observed in the three preceding experiments with English listeners and English word and nonword materials suggests that this is a language-independent effect. It is not dependent upon the existence of a large repertoire of candidate vowel sounds in the listener's language. The failure to replicate the parallel pattern of RT and missed responses, on the other hand, suggests that language-specific factors such as size of the vowel repertoire may indeed play a role in vowel detection. Just as restriction of the target set to five distinctive vowels in Experiment 3 reduced the miss rate, so did we find here a lower miss rate (in comparison with, for instance, the directly analogous English Experiment 1) in the situation in which the language itself provided only a distinctive set of possible targets.

GENERAL DISCUSSION

Four experiments with words and nonwords in two different languages have revealed a remarkably consis-

Table 4
Mean Response Time (in Milliseconds) and Proportion of Missing Responses for the Five Vowel Targets of Experiment 4 as a Function of Position in First Versus Second Syllable

	First Syllable		Second Syllable	
Vowel	Response Time	Responses Missed	Response Time	Responses Missed
/a/	643	10.8%	565	14.2%
/e/	606	4.5%	576	11.8%
/i/	579	2.1%	520	9.4%
/o/	636	4.2%	483	7.3%
/u/	645	4.5%	547	8.0%

tent picture—the speed with which a vowel target can be detected is a function of the vowel's duration. Longer vowels are responded to more rapidly. Shorter vowels are responded to more slowly.

This effect appears in a language with a large and confusable vowel inventory (English), but also in a language with a small and highly distinct vowel inventory (Spanish). It cannot be removed in English by restricting either the set of targets or the entire experimental repertoire of vowels to easily distinguishable tokens. Although listeners miss fewer targets when the targets used in the experiment are themselves quite distinct (Experiments 3 and 4) than when the experimental target set is more confusable (Experiments 1 and 2), the speed of detection remains sensitive to the duration of the vowel.

It is unlikely that the effect is located at a level at which vowel tokens are auditorily discriminated one from another. Fox, Flege, and Munro (1995) have recently observed, in a multidimensional scaling analysis of vowel discrimination performance by speakers of English and Spanish, that the English listeners relied upon more perceptual dimensions than did the Spanish listeners (unsurprisingly, a central–noncentral dimension played an important role in the English discriminations but not in the Spanish). Furthermore, there was an apparent role of duration for the English but not for the Spanish listeners. We, however, observed comparable durational effects in these two languages. We propose, instead, that the effect is located at a strategic level and may, therefore, offer us information about listeners' processing of vowels per se. We suggest that an important component of the knowledge that listeners bring to bear upon the task of speech processing is the realization that vowels are intrinsically variable. This realization can affect the strategy adopted in tasks requiring explicit manipulation of vowel phonemes. Moreover, such effects do not occur only when the language includes many competing vowel candidates, as in English; listeners treat even the distinct vowel repertoire of Spanish as consisting of intrinsically variable items.

Our explanation invokes characteristics of the task we employed in our studies; it relies upon the notion of reaching criterion for the making of an explicit response. Vowels with longer duration will enjoy the greatest likelihood that a response will be made to them, since sufficient evidence will accrue for even a high response criterion actually to be reached. Shorter vowels either will not be responded to or will be responded to with some delay as the response criterion is reached only after additional processing, which we assume will occur at a reduced rate due to concurrent processing of subsequent phonetic information. Note that when there is no subsequent phonetic information coming in—that is, when the target vowel occurs in utterance-final position—processing of the vowel should continue at the initial rate and the conditions for appearance of the correlation between response pattern and duration may no longer be met. Indeed, in five experiments comparing detection of vowel and consonant targets in isolated words, van Ooijen (1994) found that the negative correlation of RT with duration regularly appeared with vowel targets in medial position but was always absent for vowel targets in final position.

Van Ooijen's consonant/vowel experiments also serve as proof that the high response criterion for vowels does not appear only in experiments in which only vowel targets are used; in general, RTs were longer to the vowel targets than to the consonant targets. In a further study by Cutler and Otake (1994), English listeners also responded more slowly to vowel than to consonant targets in a language foreign to them, namely Japanese; since Japanese, like Spanish, has only five, relatively distinct, vowels, this result is comparable to the maintenance of a cautious response criterion in the restricted condition of Experiment 3 of the present study.

The particular relationship of vowel duration to response patterns that we observed in these experiments may, of course, be an effect specific to the phoneme-detection task; it may be that it is exactly the difficulty of the target-detection task, with its concurrent requirement to respond as fast as possible without missing occurrences of the target sound or making false-alarm responses, that brings out the prerequisite conditions: listeners respond to the imposition of a target that they realize to be intrinsically variable by adopting a cautious response strategy. This question can be answered only by designing other on-line experiments, with different response measures, in which a correlation of RT with vowel duration could potentially be observed. But the fact is that vowels have now been shown to constitute harder detection targets than consonants, and we believe that, task-specific or not, the results we have observed tell us something about the processing of vowels in general, namely that the intrinsic variability of vowel phonemes is something of which listeners are well aware.

The study of vowel perception via classic identification and discrimination tasks, as described in the introduction, is a lively area of speech-perception research. As summarized by recent contributions to this debate (e.g., Andruski & Nearey, 1992; Nearey, 1989; Rosner & Pickering, 1994; Strange, 1989a), current theories contrast the extraction of underlying target values via compensation for contextual effects (e.g., Kuwabara, 1985; Miller, 1989) with perception of dynamic specification of vowel identity, for example, in transitional information (Strange, 1989b; Verbrugge & Rakerd, 1986) or in vowel-inherent spectral change (Nearey & Assmann, 1986). We do not claim that an on-line task with natural vowel tokens addresses the same issues as arise in these debates. Moreover, the relationship between response patterns and vowel duration could in principle be claimed as supporting evidence by competing parties. It may seem to argue against a crucial role for transitional information (Strange, 1989b) and to be, instead, more consistent with models that rely to a greater extent on the spectral dynamics of the vowel (e.g., Nearey & Assmann,

1986), since the greater the duration of the vowel, the more fully these will be realized. However, consider the finding of Strange, Edman, and Jenkins (1979) that vowel identification is easier when English vowels conform to their prototypical length (i.e., short vowels are actually short and long vowels are actually long); it could be argued that, in naturally spoken words such as we presented to our subjects, the short vowels were more likely to be those which deviated from prototype values to a greater extent.[2] Yet again, recent automatic recognition studies (Harrington & Cassidy, 1994) have shown that the most informative portion of a vowel in a naturally spoken isolated word is a slice from the midpoint; this finding was interpreted by the authors as supporting a target-value model of vowel perception, and again it could be argued by proponents of such models that the longer the vowel token, the more likely it is that the underlying target value will be reached. Thus, our finding is likely to be interpretable within a number of different frameworks.

What we do believe our findings to underline is the practical importance of vowel variability for the human listener. In most identification and discrimination experiments, the natural variability present in spoken vowels is deliberately avoided; phonetic context, for example, is rigorously controlled, as is the vowel's position in the utterance token. When, as in the present experiments, we look at actual spoken words in which vowels occur in a range of phonetic contexts and in different positions in the word and we use an on-line task, we find that variability apparently exercises a significant effect upon listeners' behavior. In fact, we expect that identification tasks with similarly variable tokens would be likely to demonstrate similar effects. In a recent study, van Son and Pols (1995) measured error rates for identification of Dutch vowels extracted from a naturally produced corpus of continuous speech; they found that identification improved when additional speech context beyond the vowel kernel itself was made available, and continued to improve even when the context crossed the boundaries into neighboring phonemes.

We further believe that this effect carries through to listeners' strategy in processing natural speech under normal listening conditions. Listeners characteristically behave as if the processing of vowels is intrinsically a hard task. This even leads them to treat vowels differently from consonants in spoken-word recognition. The most striking evidence for this is a recent finding by van Ooijen (1994; in press) that when listeners are presented with mispronounced words and asked to restore them to their correctly pronounced form, they find it much easier to alter vowels than to alter consonants. One way in which this asymmetry manifests itself is in the relative speed of vowel versus consonant changes: given an input such as *shevel* or *eltimate* and instructed to turn it into a real word by changing only one sound, listeners can much more rapidly find a word that involves a vowel change (*shovel*, *ultimate*) than a word that involves a

consonant change (*level*, *estimate*). Another is in the relative accessibility of each type of change; listeners are far more likely to make an erroneous vowel change if specifically instructed to make a consonant change than vice versa. The apparent readiness of listeners in this word-recognition task to treat vowels as inherently more mutable objects than consonants is, we would argue, further evidence that listeners have adjusted their speech-processing procedures to take explicit account of the intrinsic variability of vowels.

Listeners may not be conscious of spoken-word recognition as a phoneme-by-phoneme process. However, our results suggest that listeners adjust their processing of naturally spoken input such as to differentiate between types of phonemes. One such adjustment is the setting of a high response criterion in a phoneme-detection task when the phoneme target is a vowel. This suggests that fine-tuning of the input to computational models of spoken-word recognition—Shortlist, TRACE, and their successors—to reflect differences in the relative confidence assigned to vocalic versus consonantal segments would be likely to improve the models' ability to simulate accurately the performance of human listeners with naturally spoken words in laboratory studies.

This is not to say that listeners are bad at processing vowels; on the contrary, the response strategies they adopt to deal with vowel variability are highly effective. In many ways, vowel processing appears more efficient than consonant processing. For example, studies of slips of the ear (e.g., Bond & Garnes, 1980) report that the segments least likely to be misreported are vowels in stressed syllables. Likewise, studies of perception of non-native phonological contrasts (Polka & Werker, 1994; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969) have revealed that discrimination of non-native vowel distinctions is not as difficult for adults as is discrimination of non-native consonant contrasts.

Interestingly, when we look across languages, we see that the average number of vowels in a language is typically half the number of consonants (Maddieson, 1984). All languages have both vowels and consonants, though languages differ widely in the size of their inventories of each (total segment inventories range from 11 to 141 segments, the total number of vowels varies from 3 to 46, and the total number of consonants from 6 to 95). The vowel/consonant ratio, however, although it varies from .065 to 1.308, has a mean of .402 and a median of .36. This may reflect adjustment on the part of language inventories to vowel/consonant differences in informational reliability. Certainly we believe that our results indicate adjustment on the part of listeners. The most vowel-like sounds of speech are, in short, different from the most consonant-like sounds; processing of the most vocalic sounds is different from processing of the most consonantal sounds; listeners act in ways consistent with this; and psycholinguists ought certainly to take account of it in modeling spoken-word recognition.

REFERENCES

ADES, A. E. (1977). Vowels, consonants, speech and nonspeech. *Psychological Review*, **84**, 524-530.

ANDRUSKI, J. E., & NEAREY, T. M. (1992). On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables. *Journal of the Acoustical Society of America*, **91**, 390-410.

BOND, Z. S., & GARNES, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 115-132). Hillsdale, NJ: Erlbaum.

CUTLER, A., & BUTTERFIELD, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication*, **9**, 485-495.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory & Language*, **31**, 218-236.

CUTLER, A., MEHLER, J., NORRIS, D. G., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.

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

CUTLER, A., & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 113-121.

CUTLER, A., NORRIS, D. G., & VAN OOIJEN, B. (1990). Vowels as phoneme detection targets. In *Proceedings of the International Conference on Spoken Language Processing* (Vol. 1, pp. 581-584). Kobe: Acoustical Society of Japan.

CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory & Language*, **33**, 824-844.

DIEHL, R. L., KLUENDER, K. R., FOSS, D. J., PARKER, E. M., & GERNSBACHER, M. A. (1987). Vowels as islands of reliability. *Journal of Memory & Language*, **26**, 564-573.

DIJKSTRA, T., SCHREUDER, R., & FRAUENFELDER, U. H. (1989). Grapheme context effects on phonemic processing. *Language & Speech*, **32**, 89-108.

DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception. I: Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.

FOSS, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning & Verbal Behavior*, **8**, 457-462.

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

FOX, R. A., FLEGE, J. E., & MUNRO, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, **97**, 2540-2551.

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, T. (1990). Lexical effects in phonemic processing: Faciltatory or inhibitory? *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 77-91.

GOTTFRIED, T. L., JENKINS, J. J., & STRANGE, W. (1985). Categorical discrimination of vowels produced in syllable context or in isolation. *Bulletin of the Psychonomic Society*, **23**, 101-104.

HAKES, D. T. (1971). Does verb structure affect sentence comprehension? *Perception & Psychophysics*, **10**, 229-232.

HARRINGTON, J., & CASSIDY, S. (1994). Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language & Speech*, **37**, 357-373.

JOHANSSON, S., & HOFLAND, K. (1989). *Frequency analysis of English vocabulary and grammar*. Oxford: Oxford University Press, Clarendon Press.

JUILLAND, A., & CHANG RODRIGUEZ, E. (1964). *Frequency dictionary of Spanish words*. The Hague: Mouton.

KEWLEY-PORT, D. (1991). Detection thresholds for isolated vowels. *Journal of the Acoustical Society of America*, **89**, 820-829.

KUWABARA, H. (1985). An approach to the normalization of coarticulation effects for vowels in connected speech. *Journal of the Acoustical Society of America*, **77**, 686-694.

LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 307-333). Washington, DC: V. H. Winston.

MADDIESON, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

MARSLEN-WILSON, W. D., & WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

MARTIN, M. (1977). Reading while listening: A linear model of selective attention. *Journal of Verbal Learning & Verbal Behavior*, **16**, 453-463.

McCLELLAND, J. L., & ELMAN, J. L. (1986a). Exploiting lawful variability in the speech wave. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360-385). Hillsdale, NJ: Erlbaum.

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

MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, **20**, 298-305.

MILLER, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, **85**, 2114-2134.

MORTON, J., & LONG, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning & Verbal Behavior*, **15**, 43-51.

NEAREY, T. M. (1989). Static, dynamic and relational factors in vowel perception. *Journal of the Acoustical Society of America*, **85**, 2088-2113.

NEAREY, T. M., & ASSMANN, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, **80**, 1297-1308.

NEWMAN, J. E., & DELL, G. S. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning & Verbal Behavior*, **17**, 359-374.

NORRIS, D. G. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 87-104). Cambridge, MA: MIT Press.

NORRIS, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.

NORRIS, D. G., & CUTLER, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, **43**, 541-550.

PISONI, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, **13**, 253-260.

PITT, M. A., & SAMUEL, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory & Language*, **29**, 611-632.

POLKA, L., & WERKER, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 421-435.

RAKERD, B., VERBRUGGE, R. R., & SHANKWEILER, D. P. (1984). Monitoring for vowels in isolation and in a consonantal context. *Journal of the Acoustical Society of America*, **76**, 27-31.

ROMERO, J. (1988). Campos de disperción auditivos de las vocales del castellano: Percepción de las vocales [Auditory dispersion fields for Castilian vowels: Vowel perception]. *Estudios de Fonética Experimental III*. Universidad de Barcelona, Laboratorio de Fonética, Facultad de Filología.

ROSNER, B. S., & PICKERING, J. B. (1994). *Vowel perception and production*. Oxford: Oxford University Press.

RUBIN, P., TURVEY, M. T., & VAN GELDER, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics*, **19**, 394-398.

SAVIN, H. B., & BEVER, T. G. (1970). The non-perceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, **9**, 295-302.

SCHOUTEN, M. E. H., & VAN HESSEN, A. J. (1992). Modeling phoneme perception: I. Categorical perception. *Journal of the Acoustical Society of America*, **92**, 1841-1855.

STEVENS, K. N., LIBERMAN, A. M., STUDDERT-KENNEDY, M., & ÖHMAN, S. E. G. (1969). Crosslanguage study of vowel perception. *Language & Speech*, **12**, 1-23.

STRANGE, W. (1989a). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, **85**, 2135-2153.

STRANGE, W. (1989b). Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, **85**, 2081-2087.

STRANGE, W., EDMAN, T. R., & JENKINS, J. J. (1979). Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 643-656.

VAN OOIJEN, B. (1994). *The processing of vowels and consonants*. Doctoral dissertation, University of Leiden.

VAN OOIJEN, B. (in press). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*.

VAN SON, R. J. J. H., & POLS, L. (1995). The influence of local context on the identification of vowels and consonants. In *Proceedings of EUROSPEECH* (Vol. 2, 967-970). Madrid: European Speech Communication Association.

VERBRUGGE, R. R., & RAKERD, B. (1986). Evidence of talker-independent information for vowels. *Language & Speech*, **29**, 39-57.

NOTES

1. In a pilot study for the present experiments (Cutler, Norris, & van Ooijen, 1990), we varied the instructions given to the subjects in an attempt to ascertain whether the subjects were constructing orthographic representations of the input. Half of the subjects were given instructions, as in the present study, in writing, which included examples of possible target items; the other half of the subjects were given spoken instructions. This manipulation had no effect at all on the pattern of responses.

2. We thank Terry Nearey for pointing this out.

APPENDIX

Experimental Words and Nonwords Used in These Studies

Experiment 1

For the full vowels, the words are listed in quartets, within which they were matched as well as was possible for phonetic context and frequency. The words in each quartet occur in the order: (1) primary stress on first syllable, vowel target in first syllable; (2) primary stress on first syllable, vowel target in second syllable; (3) primary stress on second syllable, vowel target in first syllable; (4) primary stress on second syllable, vowel target in second syllable. For /ə/, the words are listed in pairs, one word with the target vowel in the first syllable and one with the target vowel in the second syllable.

/a/
barter; rhubarb; cartel; embark
carton; placard; cartoon; discard
sharpen; massage; sardine; disarm
harbour; forearm; harpoon; charade
parcel; ramparts; partake; surpass
marker; postmark; marquee; impart

/ɪ/
tissue; captive; dictate; forgive
sickle; festive; sixteen; assist
liver; classic; withdraw; outlive
wither; anguish; withhold; acquit
mitten; comic; mistake; commit
mistress; vanish; mislay; remiss

/ə/
bassoon; bulbous
confuse; deacon
malign; camel
lament; emblem
velour; ravel
shallot; nuptial
convince; falcon
relapse; kestrel
manure; salmon
vacate; havoc
relent; squirrel
renew; apron

/ɛ/
fester; access; festoon; confess
settle; sunset; settee; upset
gesture; digest; genteel; reject
tenant; nutmeg; technique; condemn
melon; bluebell; lessee; propel
beckon; stipend; welloff; rebel

/ʌ/
sulphur; insult; sublet; consult
rubber; bankrupt; trustee; erupt
rusty; breadcrumbs; truncate; instruct
dumpling; sawdust; dumbfound; deduct
funnel; refund; pulsate; defunct
cupboard; hiccup; mundane; robust

Experiment 2

The nonwords are arranged in quartets/pairs, as for Experiment 1.

/a/
tarbnees; keektarb; tarbeel; eentarb
zardosh; pozzard; gardonce; moggard
garkipe; stigark; garkites; nigark
tarcelott; shontarce; tarsoss; nofftarce
lartome; drolart; lartoace; polart
pardue; stupard; parduse; cupard

/ɪ/
tissore; goretiss; tisvore; thoretiss
jickdof; mofgick; jicksot; porjick
kibloon; boolkib; kibbroon; oopkiv
skisbal; wabskiss; skisstaf; rasskiss
dibcone; scodib; dibtose; trodib
simtave; kaysim; simbane; jaysim

/ə/
besteet; beethbous
benfeece; zeaton
mentape; flayment
nekoon; stoonek
vetoss; sovven
shempash; gatshem
penzine; klypen
stetooze; croostet
terrayf; snayster
vastoant; trovas
teclowt; thonetec
gadeen; teefgad

/ɛ/
tefgop; ponteff; tefcog; norteff
tessoo; booftess; tessool; nootess
kestobe; sokess; kessoat; bokess
kettane; blaykett; kettague; vaykett
bennife; stibenn; bentite; pribenn
peshoo; proopesh; peshoon; coopesh

/ʌ/
gucksote; jopeguck; guckthoaf; hoaguck
pugneet; zeempug; pugmese; sleempug
sutclow; flosutt; suttow; dosutt
muptane; zalemup; mupsate; vaymupp
dustyste; syluss; duskyle; blyduss
guthlite; kyguth; guthdibe; diguth

Experiment 3

For each vowel, the nonwords form sets of eight monosyllables, eight bisyllables with initial stress, and eight bisyllables with final stress. In bisyllables, the target vowel always occurred in the stressed syllable.

/a/
parl; sharce; braft; snart; skarl; glarm; yarst; karch
frardock; slarvess; crarmoon; blardlee; harmost; martude; carbeel; bargrest
defarn; preparge; entarve; elsharb; opthart; honlarm; coorarth; noovarps

/ɛ/
bez; fenk; relt; sletch; treb; gred; pless; demf
clesson; smeckard; fleddee; skessood; pecklarn; zepplee; leckop; shentoo
prehetch; bewex; obneft; arkelk; dooguent; fonresh; barvell; roopess

/i/
deeb; smeek; cleast; trean; sweeve; jeeth; meach; teadge
prelest; tweemarn; shreevart; breeboon; neenoff; veenesh; beetrotch; cheedooce
toodeeze; mooteel; exfeen; elbeece; artheet; sarmeed; obseeve; concheem

/ɒ/
spog; todge; crom; gonce; noss; flott; drob; sont
londuce; stoddark; bloglee; dompeen; cothlarn; hondess; grodleck; plocktoon
rookotch; begoz; harboft; endob; empoft; rezond; barnost; tewdoss

/u/
poot; goove; prook; quoon; spoost; drooth; wooch; zooct
gloopon; thootar; stoobart; froobest; swooveen; toofon; goobess; yookly
extoon; embool; degoove; resootch; conjood; darchoom; gosmoot; arlootch

Experiment 4

The words are arranged in quartets, as for Experiment 1.

/a/

sabio; rosa; salud; pesar
santo; prisa; salón; cesar
carcel; chica; calor; local
mando; broma; matiz; normal
lazo; burla; latín; solar
nave; mina; nariz; reinar

/i/

signo; tesis; sillón; fusil
tinta; fértil; timar; hostil
lino; caliz; limón; desliz
cima; dócil; citar; zurcir
pila; lápiz; picar; tapiz
viga; movil; vigor; hervir

/u/

busto; album; burdel; bambú
turba; cactus; turrón; betún
mudo; femur; mugir; vermú
pugna; campus; pudor; champú
nuca; venus; mudar; menú
duque; arduo; durar; gandul

/e/

seda; fase; sector; coser
reja; area; rezar; laurel
cera; trance; cenar; placer
celo; bronce; cegar; torcer
tela; chiste; tenaz; cortés
presa; lumbre; presión; ciprés

/o/

norte; reino; notar; menor
torre; fruto; total; pintor
ropa; hierro; rogar; arroz
cola; pico; collar; precoz
bola; verbo; bordar; fervor
dote; mudo; doblar; bidón