Perception & Psychophysics
1988, 43, 21-30

Cues to lexical choice: Discriminating place and voice

PAUL WARREN
University of Cambridge, Cambridge, England

and

WILLIAM MARSLEN-WILSON
Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and MRC Applied Psychology Unit, Cambridge, England

Two experiments used the gating paradigm to investigate the manner in which acoustic-phonetic information is mapped onto the lexical level during the processes of lexical access and selection. The first experiment tested word identification across successive 25-msec gates of monosyllables contrasting in word-final voicing and showed a continuous uptake of durational cues. The second experiment expanded upon earlier research into the uptake of partial cues in the spectral domain and revealed strong effects at vowel closure on the choice between word candidates terminating in different places of articulation. The results were interpreted as suggesting a contrast between symmetric and asymmetric decision processes, with phonological structure being the potential source of asymmetries in the lexical interpretation of acoustic cues.

The process of spoken language comprehension begins with the projection of the speech input onto mental representations of lexical form. The purpose of the research reported here is to explore the properties of the acoustic-phonetic decision space within which the listener conducts this process of lexical access and selection. In earlier research (Warren & Marslen-Wilson, 1987), we showed that the fine-grained detail of variation in the speech signal is continuously projected onto the lexical level. Listeners do not need to wait until the end of a segment, as conventionally defined, in order to guide and constrain lexical choice. In particular, they seem to be able to exploit online the temporal overlap, in the speech signal, of acoustic cues to distinct speech segments (see Fowler, 1984).

In our first study (Warren & Marslen-Wilson, 1987), we looked at the consequences for lexical choice of temporally overlapping cues in the spectral domain. To do this, we used a gating task in which listeners heard successively larger fragments of the initial consonant cluster and vowel of words such as *scoot* or *scoop*, in which the final consonant differed in place of articulation, and of words like *crown* or *crowd*, in which the final consonant differed in manner of articulation. The subjects in this task were able to use coarticulatory changes in the formant structure of the vowel to help them identify the correct word before the final consonant was heard.

We thank Marie Jefsioutine for her help in designing and constructing the stimuli and in running the experiments. We also thank Aditi Lahiri and Uli Frauenfelder for their comments on the manuscript. This research was supported by a programme grant from the Medical Research Council. Address reprint requests to Dr. Paul Warren, University of Cambridge, Department of Experimental Psychology, Downing St., Cambridge CB2 3EB, England.

The research reported here expands on the previous study in two ways: by looking at a more differentiated set of contrasts in place of articulation, and by examining a different kind of partial cue—namely, variations in relative duration for vowels preceding voiced and unvoiced stops. This research sheds light not only on the kinds of information available to the listener to guide lexical choice, but also on the decision procedures he/she uses to assess this information.

In the first part of this study we examined the contrast in vowel duration found before stops differing in their voicing. For example, the vowel in a word like *bad* is typically longer than the vowel in *bat*. The presence or absence of voicing in word-final stops can be marked by a number of cues, which include vowel and closure duration, presence or absence of prevoicing, and the properties of the burst (Denes, 1955; Raphael, 1972, 1981; Raphael, Dorman, & Liberman, 1980; Watson, 1983). In English, the contrast in vowel duration is especially prominent (House, 1961; Peterson & Lehiste, 1960), and it is this cue that we focused on in this study. This differs from the primarily spectral cues we have looked at before, and it is of special interest in the context of a sequential lexical access process because of its essentially temporal nature.

Durational cues differ from spectral cues in the kinds of problems they pose for the listener's decision mechanism. For the latter type of cue—for example, the changes in vowel formants associated with different places of stop articulation—the spectral pattern directly signals the place of articulation of the final stop. Thus, as the last few pitch periods of the vowel are heard, the listener can begin to determine that what he/she is hearing is, for example, a word ending in a velar stop, as opposed to a labial stop.

With the durational cues to voicing, the situation is different: it is not a specific qualitative property of the acoustic spectrum that cues a given distinction, but rather the perceived duration of some stretch of the speech input relative to a given set of phonologically determined criteria for the vowel lengths associated with the voicing contrast. Our question was how the listener handles these criteria, in the processing context of on-line lexical choice, as the vowel for a given word is heard. Is, for example, a vowel treated as short—therefore signaling a voiceless stop—until it exceeds its criterial length? Or does the processing mechanism remain uncommitted until the full length of the vowel is known?

The second part of this study was intended as a refinement and replication of effects found in the earlier study for place contrasts in word-final stops (as in the pair *scoop-scoot*). Over the 80 msec of vowel leading up to closure, we found a gradual advantage emerging for the correct member of the pair—for example, an increasing tendency to produce *scoop* rather than *scoot*. We reexamined this result in this study for two reasons. First, the effect at closure, although significant, was quite small (55% responses for the correct word vs. 31% for the incorrect word), and should be replicated for a new set of stimuli. The second reason was to determine whether anticipatory coarticulation is informative in specific place contrasts. There are three different places of stop articulation in English: labial, alveolar, and velar. Although the previous study included examples of all of these, it did not allow a systematic comparison. Informal analyses of the results suggested, however, that labial consonants were more distinctive at closure than were alveolar or velar ones, and that alveolar and velar stops were less well discriminated from each other. These findings are consistent with results obtained by Pols and Schouten (1981) for Dutch listeners in a study using gated nonwords, and they need to be more systematically investigated in the domain of lexical choice. How far do these very fine-grained differences in the acoustic signaling of different places of articulation percolate through to the lexical level?

Finally, as in the previous study, we took into account the frequency of occurrence of the words whose access and identification we were studying. In the main tests of different place and voicing contrasts, the frequency variable was neutralized by using sets of words matched in frequency. However, we also included additional sets of words that contrasted in frequency to test whether the orthogonality of frequency and acoustic-phonetic cues, as observed for the cues examined in Warren and Marslen-Wilson (1987), continued to hold for the different kinds of phonetic contrast being tested.

Specifically, the frequency effect may differ as a function of the different types of indeterminacy involved in the voicing contrasts, as opposed to the place contrasts. For the place contrasts, the spectral transitions as the vowel approaches closure provide partial cues to the place of articulation of the syllable-final or word-final stop. The interpretation of this information does not interact with frequency biases.

The partial information available in the case of the voicing contrast has a potentially different status. When only the early gates for a given vowel are heard, the information that the listener has about this vowel is fully compatible with two possible interpretations. It can be interpreted simply as a short vowel, and therefore as a cue for the absence of voicing, or, so long as closure is not clearly indicated, it can be interpreted as part of a long vowel, and therefore as a cue to the presence of voicing. The signal itself does not determine what the interpretation should be. Instead, the interpretation depends on the listener's decision criteria. Biases due to frequency may interact with the application of these criteria during the interpretation of partial durational information.

METHOD

This research investigated the uptake of partial cues during word recognition and the relation of the partial cues to word frequency. The two types of cues studied—voicing and place characteristics of final consonants—were included in one set of experimental materials, but will be treated separately as Experiments 1 and 2.

Materials and Design

Experiment 1. All of the stimuli were pairs of monosyllabic words ending in plosives and differing only in the voicing of the final consonant (as in *mop* vs. *mob*). For the frequency-matched pairs, all items had a frequency of less than 50 per million in the Brown corpus, with a mean of 9 (Kučera & Francis, 1967). The structure of the word-initial cohorts from which the word pairs came was also taken into consideration. No other cohort members could be higher in frequency than the test words, and the overall cluster size was kept as small as possible. The mean cluster size was 4.78.

A total of nine frequency-matched pairs of items were used, with three pairs contrasting voicing in each of the three places of articulation of the final consonant: labial (e.g., *mop-mob*), alveolar (e.g., *squat-squad*), and velar (e.g., *flock-flog*).

A second set of eight word pairs, also differing only in the voicing of the final consonant, contrasted in the frequency of occurrence of pair members. For four of these pairs (e.g., *need-neat*), the voiced member had a higher frequency (mean of 249) than did the voiceless member (mean of 24). For the other four pairs (e.g., *bright-bride*) the voiceless member had a higher frequency (mean of 224) than did the voiced member (mean of 20). In each case, the higher-frequency member was the only high-frequency word in the cohort.

Experiment 2. The second set of items consisted of matched and contrasting frequency pairs differing only in the place of articulation of the final consonant (e.g., *sleep-sleet*). For the matched frequency items, the set consisted of four pairs in each of three place contrasts: labial/alveolar (e.g., *slop-slot*), labial/velar (e.g., *flip-flick*), and alveolar/velar (e.g., *pat-pack*). The mean frequency of the 24 items was 21 per million, and the mean cohort size was 4.67.

The frequency-contrasted items involving a place contrast formed two sets of five pairs each. The place opposition was between labial and nonlabial, with the labial items having higher frequency than the nonlabial items in one set (e.g., *top-tot*) and lower frequency in the other (e.g., *stripe-strike*). The mean frequencies of high- and low-frequency items were 157 and 7, respectively. The nonlabial set was not further differentiated into velar and alveolar be-

cause it was not possible to find an adequately controlled set of stimuli with the relevant frequency contrast.

The 39 pairs of items described above were organized into two stimulus sets. Each set contained one item from each pair and was balanced both for the occurrence of high- and low-frequency items from the frequency-contrasted materials and for the number of items from each side of a phonetic opposition.

Procedure

The test items were recorded onto audio tape by one of the experimenters in a soundproof booth using a Revox B77 tape deck. The recordings were then digitized at 20 kHz (after analog low-pass filtering at 15 kHz) and stored on a Winchester computer disk. The stimulus sequences were generated directly from these stored waveforms.

For each item, an *alignment point* was designated that corresponded to the "phoneme boundary" as used, for example, in durational measurements of speech-wave data (Cooper & Paccia-Cooper, 1980; Klatt, 1975; Peterson & Lehiste, 1960; Umeda, 1977; for further discussion see Warren & Marslen-Wilson, 1987).

The alignment points were defined according to the following procedures. For voiceless plosive items, the offset of the vowel before closure (i.e., the cessation of regular voicing with formant structure) is a clearly distinguishable feature. The alignment point for voiced plosives was defined as the point in the waveform at which the vowel formant peaks had died out, leaving vocal murmur or voicing during the closure. The alignment point for voiceless fricative items was defined as the midpoint between clear unfricated vowel and friction noise without voicing; for voiced fricatives, it was defined as the midpoint between unfricated vowel and voiced frication with no vowel formant peaks. All points were measured from the digitized waveform displayed on a high-resolution Hewlett-Packard monitor.

The construction of the gating materials was controlled by the alignment points assigned to each item. The first segment for each sequence included the entire word up to a point 125 msec before the alignment point, and normally included at least part of the vowel.[1] The sequence of test items then proceeded in 25-msec increments from this point until the end of the word, meaning that the sixth gate that the subjects heard for each item terminated at the alignment point. The total number of gates for each item varied from 8 to 13, depending on the length of the word in question. The last 2 msec of each gate was windowed to produce an accelerating attenuation that eliminated audible clicks (cf. Ohde & Sharf, 1981; Pols & Schouten, 1981).

During each experimental session, the sequences of items were generated as required from the Winchester disk, with 5-sec intervals between items within a sequence and 10-sec intervals between sequences. Each sequence was cued by a series of three tones, and each item was cued by a single tone. The test items were preceded by three practice items, after which the procedure was discussed. After each gating fragment was presented, the subjects wrote down on response sheets (one for each sequence) the word they thought they had heard, together with their confidence (on a scale from 1 to 10) in this judgment. Scale point 1 corresponded to the label *pure guess* and point 10 to the label *totally confident*.

Subjects

Twenty-six subjects from the MRC Language and Speech Group subject pool participated in the study. Thirteen subjects were tested on each set of materials. Data from 1 subject were rejected because of too many incomplete response sheets. Subjects were paid for their participation. The experiment was conducted at the Department of Experimental Psychology, Univeristy of Cambridge.

RESULTS AND DISCUSSION

Experiment 1: Voicing Contrast

To evaluate the way the listeners responded to the voicing contrasts, we need to look first at gating responses for the nine matched-frequency pairs. The overall effects of the voicing distinction on the timing of word identification is measured by the mean isolation points and recognition points for each type of stimulus. The *isolation point* is defined (as in previous research) as the mean gate at which subjects start to select the correct word candidate without subsequently changing their minds. This is typically at a point in the speech signal at which the evidence for the correct word is not yet conclusive, and other candidates may still be possible. The *recognition point* sets a stronger criterion, requiring not only that subjects should have identified the word, but that they should also be at least 80% confident in this judgment. Previous research has shown the recognition point to correspond to the point in the signal at which the identity of the stimulus has become completely clear. The mean isolation points and recognition points are given in Table 1.

On these two measures, voiced and voiceless stimuli behave in effectively the same way. Mean isolation point for both types of final stop is at vowel closure, very close to the alignment point for these stimuli. The recognition point differences for the two types simply reflect differences in the way information accumulates after closure. For the voiced stops, the closure is followed by some degree of prevoicing or vocal murmur, with the release of the burst falling an average of 59 msec after closure. For the voiceless stops, the closure is followed by a period of silence, with the onset of the plosive release falling an average of 91 msec after closure. For both voiced and unvoiced stops, therefore, mean recognition point falls about 20 msec before the release, reflecting the fact that, for some subjects, the presence or absence of prevoicing following closure was sufficient to bring their confidence up to criterion, whereas other subjects waited until they heard the gate containing the release.

The detailed pattern of responses to voiced and unvoiced items is given in Figure 1, which plots the correct and incorrect responses for the 12 gates covering the span from 125 msec before the alignment point to 150 msec after the alignment point. We scored as correct only those cases in which the subjects produced the actual word being heard—for example, responding *mob* when *mob* was be-

Table 1
Mean Isolation Points and Recognition Points for Voiced and Unvoiced Matched-Frequency Pairs (in Milliseconds from Alignment Point)

	Isolation Points	Recognition Points
Voiced	−3.4	+39.2
Unvoiced	−2.7	+67.9

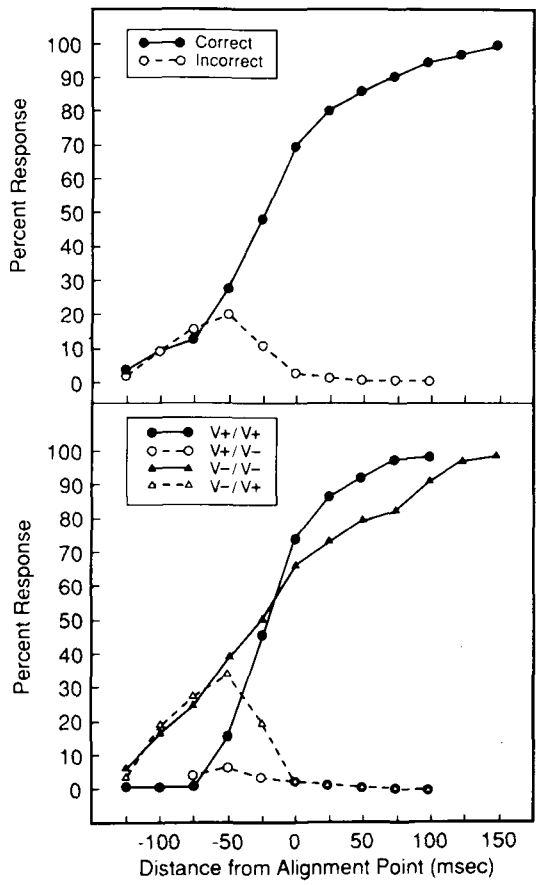

Figure 1. Percent responses to matched-frequency voiced and unvoiced stimuli. The upper panel gives the overall correct (producing the correct word) and incorrect (producing the paired word with incorrect voicing) responses. The lower panel gives the responses broken down by voice, plotting correct voiced (V+/V+) and unvoiced (V−/V−) responses and incorrect voiced (V+/V−) and unvoiced (V−/V+) responses as a function of gate. The alignment point is at 0 on the horizontal axis.

ing presented. In the incorrect cases, the subjects produced the opposite member of a voicing contrast (e.g., responding *mop* to *mob*). The upper panel of the figure plots the combined responses for voiced and unvoiced stimuli, and the lower panel breaks down the responses according to voicing category. We can see here how the subjects' abilities to discriminate between voiced and unvoiced stimuli changed depending on the amount of stimulus information available.

The response data for the first six gates, covering the crucial period for the accumulation of information about duration, were entered into a three-way analysis of variance (ANOVA), with response (correct or incorrect), voicing, and gate as the three fixed effects. Separate ANOVAs were run on the subject and item means, and combined to yield minF' ratios (all values are significant at the .05 level unless otherwise stated). There were strong main effects of response [min$F'(1,14) = 31.39$] and of gate [min$F'(5,53) = 22.26$], with a significant interaction between them [min$F'(5,67) = 21.45$]. These results reflect the dominant pattern in the upper panel of Figure 1. Correct and incorrect responses increase together over the first four gates. The two curves then diverge sharply, with incorrect responses decreasing to near zero at the alignment point, while the proportion of correct responses increases to 70%.

The pattern of voiced and unvoiced responses is not, however, uniform over gates, as reflected in a marginally significant interaction of voicing with response [min$F'(1,9) = 3.51, p < .10$] and in a significant three-way interaction with gate and response [min$F'(5,57) = 2.89$]. The source of this interaction is the listeners' strong biases over the early gates to respond with the voiceless member of each pair. The lower panel in Figure 1 shows that very few voiced responses were made over the first four gates—that is, up to about 50 msec before the end of the vowel. If the subjects responded with any member of the stimulus pair in question, it was almost always the unvoiced member. For example, if the stimulus was either *mob* or *mop*, then *mop* was produced. Over the last 50 msec preceding the alignment point, however, voiced responses accelerated rapidly, achieving equality with unvoiced responses at closure (74% vs. 66% correct at the sixth gate for the voiced and unvoiced candidates, respectively). This is reflected in the mean isolation points (Table 1), which fall within 5 msec of the end of the vowel for both stimulus types.

Trend analyses of the results over the first six gates confirm this early contrast between voiced and voiceless responses. For the correct voiceless responses, the only significant component across gates is linear [min$F'(1,10) = 31.54$]. This contrasts with the curve for voiced correct responses, which has both linear [min$F'(1,10) = 42.91$] and quadratic components [min$F'(1,10) = 17.83$], and with the purely quadratic effect for the incorrect voiceless responses [min$F'(1,10) = 8.39$].

The picture that emerges from these results is straightforward. The subjects were unwilling to interpret a stimulus as voiced until the length of the vowel exceeded some criterial amount. If they did produce any member of the test pair before this amount was reached, then it was normally the unvoiced member. This pattern is illustrated in Figure 2, which plots correct and incorrect responses to voiced stimuli as a function of the amount of the vowel that the subject had heard when the response was made as established by measuring from vowel onset to the end of the current gate. The voiceless responses to voiced stimuli show especially clearly the operation of a durational criterion. These responses peak at around 130 msec, corresponding quite well to the average vowel duration of 141 msec in the sample of voiceless stimuli to which the subjects were exposed. Voiced responses, in contrast, did not begin to predominate until 150 msec or more of the vowel had been heard.

This pattern of results, with the listeners' early responses controlled by a durational criterion, recurs essentially unchanged for the contrasted-frequency results.

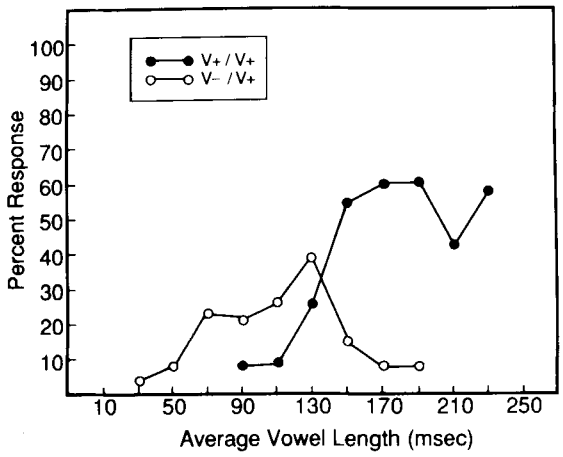

Figure 2. Percent voiced and unvoiced responses to voiced stimuli, plotted as a function of vowel length at the gate where the response was produced.

These results are plotted in Figure 3, which shows the correct and incorrect responses across gates for the voiced and unvoiced stimuli, plotted according to type of response.[2] Voiced responses (to both voiced and unvoiced stimuli) are plotted in the upper panel, and the corresponding unvoiced responses are plotted in the lower panel.

As in the analysis of the matched-frequency sets, the data for the first six gates were entered into two ANOVAs (on subjects and on items) with the three factors of response (correct or incorrect), voicing, and gate, plus the fourth factor of frequency (high or low). Paralleling the analysis of the matched-frequency data, there were main effects of response [min$F'(1,16) = 18.60$] and gate [min$F'(5,108) = 13.42$], and an interaction between them [min$F'(5,115) = 7.55$], reflecting the increase in correct candidates over gates and the simultaneous decrease in incorrect candidates. The three-way interaction with voicing [min$F'(5,97) = 2.69$] reflects the differential distribution over gates of correct and incorrect responses for voiced and unvoiced stimuli, closely paralleling the pattern previously observed for the matched-frequency sets (see Figure 1).

Frequency does not significantly affect this basic pattern, as we can see by comparing the effects in Figure 3 with the pattern of responses to matched-frequency stimuli plotted in the lower panel of Figure 1. There is no main effect of frequency [min$F'(1,13) = 1.39, p > .10$], but there is a strong interaction with response [min$F'(1,16) = 9.55$]. For high-frequency stimuli, there is a much greater disparity over the first six gates between correct and incorrect responses (45% vs. 4%) than for low-frequency stimuli (21% vs. 15%). But there is no evidence for any interaction of frequency with the voice variable. Frequency can shift up or down the level of response to voiced and voiceless stimuli, but it does not change the shape of the curves over gates. In particular, the listeners' responses are still strongly controlled by the durational criterion. Thus, in the lower panel of Figure 3, high-frequency voiceless responses to low-frequency voiced stimuli follow a pattern identical to that of the comparable errors in Figure 1: They climb steeply over the first few gates, and then drop off as vowel length starts to exceed criterion. Conversely, for the high-frequency voiced responses in the upper panel of Figure 3, incorrect responses do not start to appear until the fourth gate.

In general, as in our previous study (Warren & Marslen-Wilson, 1987), word frequency operates orthogonally to the effects of sensory variables. Frequency can change the likelihood that a given response will be produced, but it does so within the limits imposed by the sensory input. Its strongest effects, therefore, are on the most ambiguous stimuli. For both voiced and unvoiced stimuli, the differential between high- and low-frequency stimuli is largest for the first five gates, and diminishes as closure approaches and additional cues become available. For voiced stimuli, the differential disappears completely at closure, as prevoicing begins. For unvoiced stimuli, the

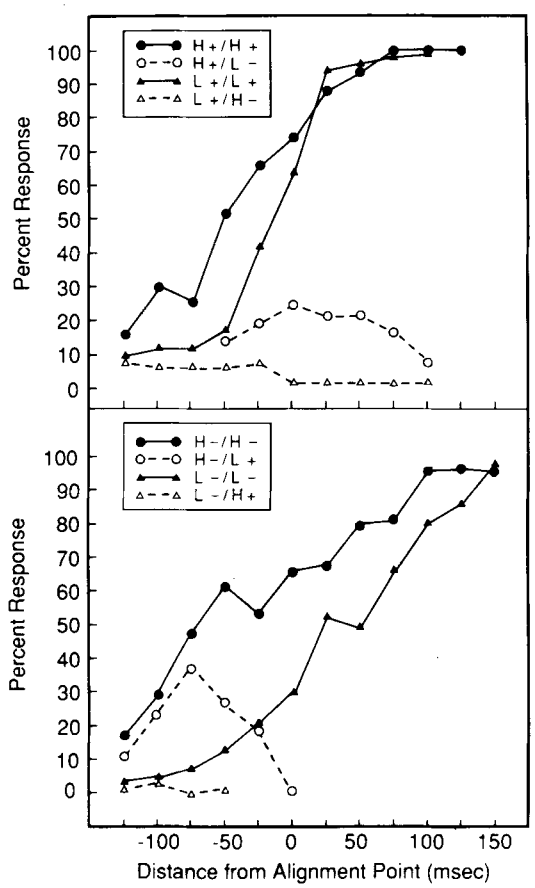

Figure 3. Percent responses to contrasted-frequency voiced and unvoiced stimuli. The upper panel gives the correct high-frequency voiced responses (H+/H+), the correct low-frequency voiced responses (L+/L+), and the incorrect voiced responses to the corresponding frequency-contrasting unvoiced items (H+/L− and L+/H−). The lower panel gives the equivalent information for the correct and incorrect unvoiced responses. The alignment point is set at 0.

differential remains apparent until the stops start to be released (75 to 100 msec after closure).

For a final look at how the voicing contrast is handled during the perception of a word, we turn to a more complete analysis of the listeners' responses to the gating sequence. The discussion up to now has been based on a subset of the responses—namely, those in which listeners responded with one of the two members of a voicing contrast pair. This analysis is expanded here to include all responses with the correct initial phoneme and the correct vowel—in effect, all responses that belong to the same word-initial cohort as the target word. Each of these responses was analyzed according to whether or not the following consonant agreed with that of the target word in each of three basic phonological categories: voicing, place, and manner. Figure 4 displays the results of this analysis, graphically illustrating how the subjects moved through the acoustic-phonetic decision space over time, narrowing down their choices as more of a word was heard. The analysis of responses to voiced stimuli is given in the upper panel, and the analysis for voiceless stimuli in the lower panel.

As shown in Figure 4, cues to manner became available earliest. Candidates in which the following consonant had a different manner of articulation were already becoming disfavored 100 msec before vowel closure. Place of articulation began to be discriminated somewhat later, with correct and incorrect responses starting to separate at about 50 msec before closure. For voicing, the timing with which this was discriminated depended on whether the word being heard terminated in a voiced or an unvoiced stop. The correct responses to the unvoiced stimuli and the incorrect (i.e., unvoiced) responses to the voiced stimuli exhibited the same strong initial bias toward unvoiced responses that we observed in the earlier analyses (see Figure 1). For both types of stimulus, unvoiced responses by far exceeded voiced responses over the first four gates, with correct voiced responses starting to predominate only at the fifth gate, 25 msec before closure. In effect, the subjects treated a vowel as short—and therefore as signaling a voiceless consonant—until the signal proved otherwise.

Before moving on to Experiment 2, we should consider an alternative account of the voicing results, which claims that the effects we found here do not primarily reflect durational cues, but are based instead on differences in spectral quality for the last few pitch periods of vowels preceding voiced or unvoiced stops.

There are a number of reasons for rejecting this theory. First, there is ample evidence from previous research (see Watson, 1983) that vowel duration *is* an important cue to voicing in English, and there is no reason to suppose that this cue was not operative for these stimuli as well. Second, the preference for voiced over unvoiced responses started to emerge for our stimuli some 70–80 msec before vowel offset, which was before any vowel-final spectral cues to voice were likely to have been heard. Third, the pattern of error responses clearly shows the operation of a durational criterion. Voiced responses

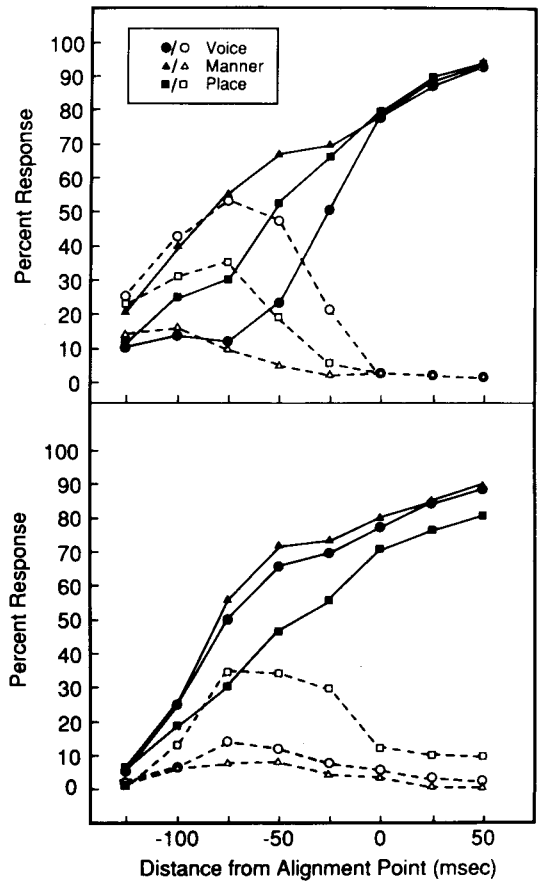

Figure 4. Percent responses to matched-frequency voiced and unvoiced stimuli as a function of type of deviation from the word actually being heard (see text). Filled symbols represent responses matching in voice, manner, and place; unfilled stimuli represent the corresponding incorrect responses. The upper panel plots responses to voiced stimuli, and the lower panel gives responses to unvoiced stimuli. The alignment point is set at 0.

to voiceless stimuli, for example, emerged only for the longer voiceless vowels, and for the later gates (e.g., Figure 3, upper panel). Here there can be no question of spectral cues to the presence of voice, since what the listeners were hearing was the vowel preceding an unvoiced stop. By the same token, the predominance of unvoiced responses to the early gates for vowels preceding voiced consonants (see Figure 2) cannot also have been based on spectral cues.

In summary, although we accept the possibility that spectral cues to voice may become available late in the vowel, and may contribute to the discrimination of voiced from unvoiced postvocalic consonants, it is clear that duration was the primary cue determining the listeners' performance over the time period critical for our durationally based analysis.

Experiment 2: Place Contrast

The second part of this study looked at the discrimination of place of articulation for voiceless word-final plo-

sives, contrasting labials, alveolars, and velars. Of the 12 matched-frequency pairs, 2 had to be discarded.[3] The results for the remaining 10 pairs were entered into ANOVAs on subjects and on items, with the factors place (labial, velar, or alveolar), response (correct or incorrect), and gate. We concentrated here, as elsewhere, on the six gates leading up to closure. There was a strong main effect of gate [min$F'(5,135) = 35.93$], a marginally significant main effect of response [min$F'(1,27) = 3.90$, $p < .10$],[4] and an interaction between them [min$F'(5,142) = 8.63$]. There were no significant effects of place.

These overall results are summarized in Figure 5. The upper panel of Figure 5 shows the pooled correct and incorrect responses, collapsed across place and plotted for the six gates up to closure (the alignment point) and the two subsequent gates, taking the responses 50 msec into the closure. The lower panel contains the discrimination curve for these pooled responses, derived from the overall data by subtracting incorrect from correct plosive responses. This curve shows directly the timing with which anticipatory coarticulatory information about place of articulation became available to affect lexical choice. There is no discriminatory information in the first four gates, up to 50 msec before closure. Listeners were equally likely to opt for the incorrect or correct place of final articulation. Over the next 50 msec, place of articulation became increasingly better discriminated, giving a robust effect at closure of 64% correct, as opposed to 14% for the incorrect responses.

These overall results give a clear positive answer to the first question we asked in this second part of the study—whether the relatively weak place effects in our earlier research could be replicated for a new set of stimuli in which contrasts between places of articulation were better controlled. To answer the second question—whether these places of articulation differ in discriminability—we need to turn to the detailed results for each place contrast. These are illustrated in Figure 6, which plots the discrimination curves over the first eight gates for each of the three place contrasts. These discrimination curves are calculated by subtracting the incorrect plosive scores at each gate from the corresponding correct plosive scores. For example, in the leftmost panel of Figure 6, the discrimination curve for the labials is calculated by subtracting the number of labial responses to alveolar stimuli from the labial responses to labial stimuli. Conversely, the alveolar curve is derived by taking alveolar responses to alveolar stimuli, and subtracting from these the alveolar responses to labial stimuli. This gives us the proportion of correct responses due to the presence of the appropriate vowel transitions, while canceling out effects due to other causes.

Two points can be made on the basis of these results. First, they clearly follow the trend we observed in our earlier study. The place contrasts involving labials tended to be discriminated better and earlier than those involving only velars and alveolars. For the labial/alveolar and

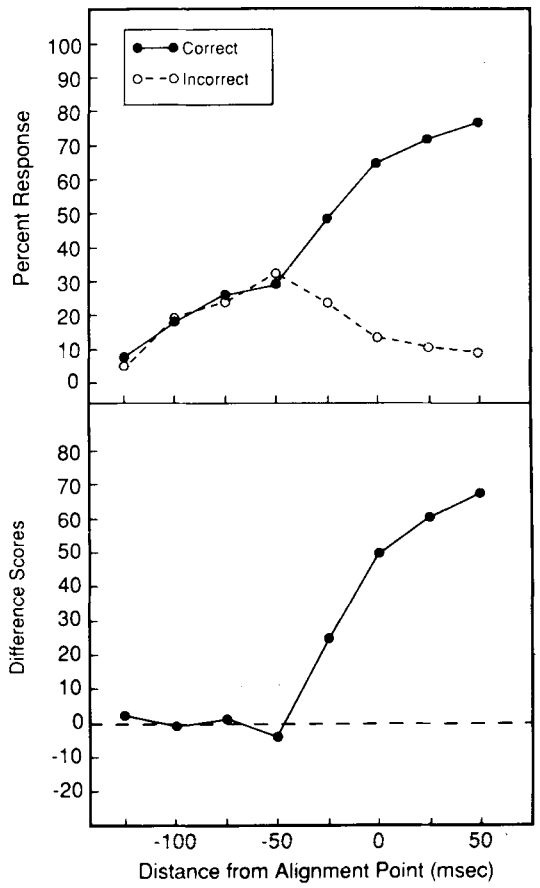

Figure 5. Overall responses to matched-frequency stimuli contrasting in place. The upper panel gives the overall percent correct (producing the correct word) and incorrect (producing the paired word differing in place) responses. The lower panel gives the overall discrimination curve (see text). The alignment point is set at 0.

the labial/velar contrasts, the average effect at the fifth gate is 36.2, and at closure it is 62.8. For the alveolar/velar contrasts, the corresponding scores are 2.5 and 27.0. However, in ANOVAs carried out on the discrimination curves, these differences were significant only in the ANOVAs based on the subject means, and did not reach significance on the item ANOVAs.

The trend for labials to be more distinctive before closure remains, therefore, no more than a trend.[5] Furthermore, it may be that the reduced discriminability of our velar and alveolar stimuli primarily reflects the vowel that was used. All four velar/alveolar pairs contained a front vowel. This may have led to the production of a "fronted" /k/, articulated farther forward than the standard velar /k/. Such a /k/ is articulated with the tongue moving into a position closer to the tongue position for a /t/, so that the transitional cues in the vowel would be more difficult to discriminate for the two cases.

The second point to be made about the place discrimination curves concerns their implications for the listeners' decision strategies, and for the way that different types

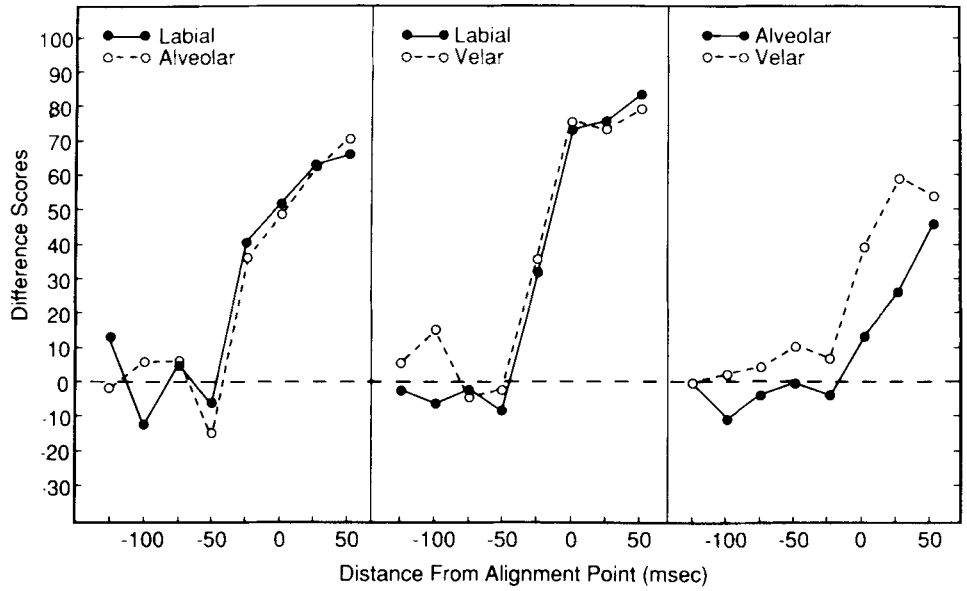

Figure 6. Discrimination curves for matched-frequency place stimuli, plotted as a function of type of place contrast. The alignment point is set at 0.

of cues are interpreted during speech processing. These implications derive from the fact that the discrimination curves in each contrast are effectively identical, especially for the two contrasts involving labials (see the left and center panels in Figure 6). How is the system arranged such that both members of a pair are equally well discriminated? We will return to this and related questions in the concluding discussion.[6]

CONCLUSIONS

The results of the two experiments reported here confirm our claims, based on earlier research, for the immediate uptake of accumulating acoustic information during lexical access and selection. In the expanded treatment of place contrasts we found strong effects of partial cues during the vowel, with listeners beginning to discriminate place of articulation of the final consonant some 25–50 msec before closure. In the investigation of the durational contrasts in vowels before consonants differing in voicing, we found even stronger anticipatory effects, with voiced and unvoiced stimuli starting to be discriminated 50–75 msec before closure. On the basis of these and earlier results, what conclusions can we draw about the manner in which the acoustic-phonetic input is projected onto the lexical level during the process of spoken word-recognition?

Continuous uptake. There do not appear to be any discontinuities in the projection of the speech input onto the lexical level. The speech signal is continuously modulated as the utterance is produced, and this continuous modulation is faithfully tracked by the processes responsible for lexical access and selection. As the spectrum of a vowel starts to shift toward the place of articulation of a subsequent consonant, the shift is reflected in a shift in the listener's lexical choices. As the duration of a vowel increases, the listener produces lexical choices that reflect these changes in duration, shifting from voiceless to voiced as the durational criterion is reached and surpassed. There is immediate use of partial durational cues, just as there is immediate use of partial spectral cues.

This property of continuous uptake does not, per se, discriminate between segmental and nonsegmental theories of lexical representation and access—that is, between theories that assume some form of segmental coding of the signal prior to entry into the lexicon, and theories that do not. As we showed in our earlier study (Warren & Marslen-Wilson, 1987), in which we took McClelland and Elman's (1986) TRACE model and Klatt's (1979) LAFS model as exemplars of the two approaches, both types of model could, in principle, continuously exploit partial information in the spectral domain. This also holds true for the exploitation of durational information. TRACE, for example, might keep track of the durational properties of speech by using a time pointer aligned to the "processing cycles" of the model (McClelland & Elman, 1986). After a specified number of cycles, only the voiced-consonant context would be compatible with the input, which could presumably be realized in the model in the activation relationships between phoneme nodes. The LAFS model uses a sequence of spectral analyses to construct a pathway that defines the correct lexical representation. This model could also exploit the durational contrast, since the pathways through the lexical network will diverge after a certain number of repetitions of similar spectra (assuming that the model can normalize for speech-rate and speaker differences). Both types of model, therefore, predict that the voiced candidate will be selected at

the point at which the vowel becomes too long to be compatible with the voiceless candidate. Neither model, however, seems to predict the early bias toward voiceless candidates.

Data-driven access and selection. If a system continuously tracks its bottom-up input, then the input will seemingly play a major role in determining its behavior. The patterning of frequency effects in the present research shows that lexical access and selection are indeed primarily controlled by the bottom-up input to the process. For both place and voicing contrasts, frequency affects performance only when the sensory input permits—that is, when the available acoustic information is sufficiently ambiguous or indeterminate to allow a choice between two or more alternatives. Under such conditions, strong frequency effects are observed. But these dissipate as soon as more determinate bottom-up information becomes available—for example, the presence of voicing after closure for word-final voiced stops.

Symmetric and asymmetric decision processes. The priority of the acoustic-phonetic input in lexical access and selection does not mean that all the information in the signal has an equal status for the decision process. In fact, the overall pattern of results suggests a contrast between what we can call *symmetric* and *asymmetic* decision processes in the interpretation of partial acoustic cues.

We see symmetric processes applying in the case of simple coarticulatory effects, such as the vowel transitions signaling place of articulation. The parallel curves in Figure 6 show, for example, how the accumulation of discriminating information occurs at the same rate for each member of a specific contrast in place of articulation. Evidence for any one place of articulation is interpreted symmetrically, both as positive evidence for that articulation and as negative evidence against any of the alternatives. If the formant transitions indicate a velar place of articulation, this is directly interpreted as evidence for a word ending in a velar stop and as evidence against competitors terminating either in labial or alveolar stops—and similarly for each place of articulation. This symmetry in the use of the available information is consistent with the cohort model of spoken word recognition, with its emphasis on the contingency of perceptual choice.

This type of decision process can be contrasted with the apparently asymmetric decision processes observed for effects such as vowel lengthening (in the present study) and nasalization (studied in Warren & Marslen-Wilson, 1987). We described the asymmetries in the interpretation of durational cues—namely, the strong bias to treat partial cues as evidence for unvoiced stops, even though the available information also allowed the voiced interpretation[7]—earlier in this article.

The asymmetries in the interpretation of vowel nasalization emerged in our earlier study, in a contrast between words ending in nasal consonants versus oral consonants (such as *crown* and *crowd*). These were asymmetries in the signal value of the presence of nasalization as opposed to the absence of nasalization. When the vowel was nasalized, there was an early and strong effect on performance, directing listeners toward word candidates ending in nasal consonants and excluding oral consonants. The absence of nasalization, however, had much weaker effects, and did not prevent listeners from selecting words ending in nasal consonants, right up to the point at which the signal became unambiguous.

What causes these asymmetries in the interpretation of partial durational and nasalization cues? One possibility is that the interpretation of these cues, unlike that of the formant transition cues, depends on the abstract representation of the sound system of the language (for a fuller discussion, see Lahiri & Marslen-Wilson, 1987; Frauenfelder & Lahiri, 1987). The vowel fomant transitions, which occur as the vocal tract moves from one articulatory state to another, are biomechanically obligatory consequences of the physics of the system, whereas the variations in duration, or in the presence or absence of nasalization, are not. They are in some sense optional processes that fall under phonological control. The durational cue to voicing varies considerably across languages, and although some degree of vowel nasalization may be obligatory, the timing of its onset also varies across languages. In English, in which vowel nasalization is allophonic, onset is much earlier than it is in languages such as Dutch, in which vowel nasalization is not allophonic.

For these two cases—durational cues to voicing, and nasalization—the implications of the incoming acoustic information can be assessed only in the context of the phonological system in which they are represented. Presumably, it is the properties of this system of representation that induce asymmetries in the interpretation of these cues. For example, if we assume a notion of *markedness* in phonological representations, then the bias toward voiceless stops may reflect the possibility that the voiceless case is the default, or unmarked case, and that this default will be overridden only when information becomes available that actively excludes it—for example, when vowel length exceeds the appropriate durational criterion.

The asymmetry in the signal value of the presence or absence of nasalization may reflect the status of the nasal feature for vowels in English. In particular, because English has no nasal vowels, it is likely that the abstract specification of English vowels does not include the feature [± nasal]. If so, then the abstract form representation of lexical items (whether terminating in oral or nasal consonants) does not specify whether the vowel is nasal. This means that when an un-nasalized vowel is heard (i.e., a vowel followed by an oral consonant), there is nothing in the abstract representation of lexical items ending in nasals that could exclude these as possible responses. If a vowel has no nasality feature, then the absence of nasalization cannot be a discriminant property of the input. In contrast, a vowel that *is* nasalized provides a positive cue to the status of the following consonant. The acoustic consequences of lowering the velum are readily detectable, and, furthermore, are unambiguous in their signal value.

Since English has no nasal vowels, the nasalization of a vowel can only mean that the following consonant is a nasal.

These are, of course, only preliminary speculations as to the source of some apparent asymmetries in the interpretation of partial acoustic cues. But if we are correct in suggesting that the formal properties of phonological representations can help determine the lexical interpretation of the speech input, then this has important potential implications for how we should investigate the properties of the acoustic-phonetic decision space within which the listener conducts the processes of lexical access and selection. In particular, it suggests that future research should look more closely at the role of phonological structure in mediating between acoustic-phonetic analysis and the process of lexical choice.

REFERENCES

COOPER, W. E., & PACCIA-COOPER, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

DENES, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, **27**, 761-764.

FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

FRAUENFELDER, U. H., & LAHIRI, A. (1987). *Understanding words and word recognition: Can phonology help?* Unpublished manuscript, Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

HOUSE, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, **33**, 1174.

KLATT, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129-140.

KLATT, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, **7**, 279-312.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

LAHIRI, A., & MARSLEN-WILSON, W. D. (1987). *The mental representation of lexical form: A phonological approach to the recognition lexicon*. Unpublished manuscript, Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

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

OHDE, R. N., & SHARF, D. J. (1981). Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study. *Journal of the Acoustical Society of America*, **69**, 297-300.

PETERSON, G. I., & LEHISTE, I. (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693-703.

POLS, L. C. W., & SCHOUTEN, M. E. H. (1981). Identification of deleted plosives: The effect of adding noise or applying a time window (A reply to Ohde and Sharf). *Journal of the Acoustical Society of America*, **69**, 301-303.

RAPHAEL, L. J. (1972). Preceding vowel duration as a cue to the voicing characteristics of word-final consonants in English. *Journal of the Acoustical Society of America*, **51**, 1296-1303.

RAPHAEL, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, **38**, 126-147.

RAPHAEL, L. J., DORMAN, M. F., & LIBERMAN, A. M. (1980). On defining the duration that cues voicing in final position. *Language & Speech*, **23**, 297-307.

UMEDA, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, **61**, 846-858.

WARREN, P., & MARSLEN-WILSON, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, **41**, 262-275.

WATSON, I. M. C. (1983). Cues to the voicing contrast: A survey. *Cambridge Papers in Phonetics & Experimental Linguistics*, **2.**

NOTES

1. In the previous study, the first gate began 80 msec before the alignment point. We used an earlier starting point in this study because some of the effects we found in the first study were already present 80 msec before the alignment point.

2. The pair *cold-colt* had to be discarded because a large number of subjects did not correctly identify *colt*, responding instead with *cult*.

3. The pairs *chap-chat* and *knit-nick* were discarded because of the high proportion of *trap* responses to *chap* and of *net* responses to *knit*.

4. The response factor was significant on both subject [$F1(1,24) = 17.33$] and item [$F2(1,17) = 5.04$] analyses.

5. The trend toward improved discriminability of unvoiced labials also occurs in the first part of this study, in which voicing contrast was tested. The matched-frequency voicing stimuli were assigned equally to the three places of articulation. Over the first six gates, the unvoiced labials averaged 53% correct, as opposed to 34% for the alveolars and 16% for the velars.

6. We do not report in detail here the results for the contrasted-frequency place stimuli, since the pattern was very similar to that obtained both for the voicing stimuli and in our earlier research (Warren & Marslen-Wilson, 1987). The effects of frequency were again statistically orthogonal to the effects of stimulus variables. Frequency changes the height of the response curves but does not change their shape.

7. We cannot, given the current data, exclude a slightly different explanation for the bias to unvoiced stops for early gates. For these gates, what the listener hears is a truncated vowel. This may function, perceptually, as a short vowel, rather than as part of a vowel of uncertain length (as we are claiming). And since short vowels mean unvoiced post-vocalic stops, the listener will give as a response the lexical item containing the unvoiced stop—not because voicing is marked or unmarked, but simply because a truncated vowel is heard as a short vowel. Preliminary evidence from other research suggests, however, that truncated vowels are not always interpreted as evidence for unvoiced stops.

(Manuscript received February 26, 1987;
revision accepted for publication July 6, 1987.)