Perception & Psychophysics
1983, 34 (4), 338-348

Phonological context in speech perception

DOMINIC W. MASSARO and MICHAEL M. COHEN
University of California, Santa Cruz, California

Speech perception can be viewed in terms of the listener's integration of two sources of information: the acoustic features transduced by the auditory receptor system and the context of the linguistic message. The present research asked how these sources were evaluated and integrated in the identification of synthetic speech. A speech continuum between the glide-vowel syllables /ri/ and /li/ was generated by varying the onset frequency of the third formant. Each sound along the continuum was placed in a consonant-cluster vowel syllable after an initial consonant /p/, /t/, /s/, and /v/. In English, both /r/ and /l/ are phonologically admissible following /p/ but are not admissible following /v/. Only /l/ is admissible following /s/ and only /r/ is admissible following /t/. A third experiment used synthetic consonant-cluster vowel syllables in which the first consonant varied between /b/ and /d/ and the second consonant varied between /l/ and /r/. Identification of synthetic speech varying in both acoustic featural information and phonological context allowed quantitative tests of various models of how these two sources of information are evaluated and integrated in speech perception.

Whorf (1956) claimed that speech is the greatest show people put on and his observation is no less true of perception than of production. Speech perception has consistently amazed its students primarily because of the relatively complex relationship between the acoustic signal and perceptual recognition. A discrete linguistic message is conveyed by a relatively continuous signal. In addition, the acoustic signal specifying a particular linguistic unit is context sensitive; properties of a unit found in one context are significantly modified in another. The listener also functions reasonably well when the speech signal is embedded in noise or other potentially distracting messages. There is considerable debate concerning how informative the acoustic signal actually is (Cole & Scott, 1974; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Massaro, 1975; Stevens & Blumstein, 1978). However, even if acoustic signal proved to be sufficient for speech recognition under ideal conditions, few researchers believe that the listener relies on only the acoustic signal. Most researchers would not disagree with the idea that the listener normally achieves good recognition by supplementing the information from the acoustic signal with information generated through the utilization of linguistic context.

Given this state of affairs, one goal of speech perception research is to assess how information from the acoustic signal is combined or integrated with information from linguistic context. Previous research has been primarily directed at showing a positive contribution of linguistic context rather than at showing how it is integrated with information from the acoustic signal (Cole & Jakimak, 1978; Marslen-Wilson & Welsh, 1978; Pollack & Pickett, 1964). The goal of the present investigation was to study the evaluation and integration of information in the acoustic signal and linguistic context. The experiments manipulated both the acoustic signal and phonological context in a speech-identification task. Synthetic speech was used to vary the information in a given sound segment. This segment was placed in different sequences of sounds to vary the degree of phonological context for a given sound. Phonological context simply corresponds to the degree to which a sound segment is appropriate or likely in the context of surrounding speech sounds.

Brown and Hildum (1956) provided one of the first systematic studies of phonological and lexical context in speech perception. Consonant-vowel-consonant syllables were recorded and presented to listeners for identification. The initial consonant was either an admissible or an inadmissible consonant cluster in word initial position in English. In addition, the admissible clusters either made a word or did not. The vowel-consonant portion of the syllable was always admissible and was the same for each comparison. For example, /glib/, /spib/, and /tlib/ would be instances of words, phonologically admissible pseudowords, and phonologically inadmissible nonwords, respectively. Listeners made more identification errors for the inadmissible syllables than for the admissible syllables. Words were identified better than the admissible syllables. The usual conclusion from these results is that listeners utilize knowledge of lexical and phonological context in their perception of speech. However, one limitation with interpreting the

The preparation of this paper was supported in part by NIMH Grants MH-19399 and MH-35334. The authors' mailing address is: Program in Experimental Psychology, University of California, Santa Cruz, California 95064.

results in terms of context effects is that the different context conditions actually involve different sounds. In the example, the consonant cluster /tl/ may be actually more difficult to recognize than /sp/ regardless of the listener's past experience. This problem may be particularly acute because the utterances were made using natural speech with no possible control for clarity of articulation. In addition, this experiment could not address the issue of how acoustic featural information and phonological context are evaluated and integrated in perceptual recognition.

More recently, Ganong (1980) assessed the contribution of lexical context on the perception of stop consonants in initial position. The voice onset time (VOT) of the initial stop consonant was varied to create a continuum from a voiced to a voiceless sound. The following context was varied so that, in one condition, the voiced stop would make a word and the voiceless stop would not. In the second condition, the reverse would be true. For example, subjects identified the initial stop as /d/ or /t/ in the context *-ash* (where /d/ makes a word and /t/ does not), or *-ask* (where /t/ makes a word and /d/ does not). Positive effects of context were found in that voiced responses were more frequent when /d/ made a word than when /t/ made a word. In addition, the interaction of VOT with context revealed that the contribution of context was largest at the most ambiguous levels of VOT. Massaro and Oden (1980b) extended their fuzzy logical model of speech perception (Massaro & Oden, 1980a; Oden & Massaro, 1978) to describe the quantitative findings of Ganong. The central assumption of the model was that acoustic featural information and lexical context make independent contributions to perceptual recognition. Even with this constraint, the model was able to provide a good quantitative description of the observed results.

The goal of the present paper was to extend the basic paradigm of Ganong (1980) to assess the contribution of phonological context to speech perception. In the first two experiments, the observers listened to and identified the glides /l/ and /r/. The synthetic speech sounds were varied along a continuum between /li/ and /ri/, which can be made by changing the starting frequency of the third formant (F3) transition. Analogous to the study of lexical context, these sounds are placed after different consonants to vary the phonological context. If the sounds are placed after the word initial consonant /s/, then /l/ is phonologically admissible in English but /r/ is not. Listeners should hear /l/ more often than /r/ in this context. Given the initial consonant /t/, however, listeners should be more likely to hear /r/ than /l/. In English, /l/ cannot follow initial /t/. In addition to these two conditions, the contexts /p/ and /v/ were included. Both /l/ and /r/ are phonologically admissible following initial /p/ but neither is admissible following initial /v/. These four context conditions are analogous to the conditions of Massaro's (1979) study of visual featural information and orthographic context in letter recognition. The results of the present experiment provide a test of whether the listener utilizes phonological context in speech perception. If phonological constraints are utilized, the experimental design would allow for quantitative tests of various models of how context and acoustic signal are integrated together in speech perception.

It is important to demonstrate that it is the phonological context and not the acoustic context that modifies perceptual recognition of the glide in the test syllable. It is possible that the acoustic structure of /t/ provides more acoustic featural information for the glide /r/ than for the glide /l/. It is also possible that the acoustic structure of /t/ modifies the featural analysis of the acoustic information during the glide because of forward masking, assimilation, contrast, or some other auditory process. The first experiment attempted to assess the magnitude of the contribution of the acoustic structure of the initial consonant. The F3 value was either maintained at a fixed value during the initial consonant or it was set to the value of the F3 of the following glide sound. If the acoustic structure of the initial consonant is responsible for differences in perceptual recognition of the glide, then the value of F3 during the initial consonant should have an important influence on perceptual recognition of the glide. If the acoustic structure of the initial consonant is the important variable, the context effect should be much larger for the varying condition than for the fixed one. On the other hand, equivalent context effects for the fixed and varying conditions would provide evidence that the context effect is not simply due to the acoustic structure of the initial consonant.

EXPERIMENT 1

Method

Subjects. Two groups of three subjects each were tested on each of 2 consecutive days. The subjects were students in an introductory psychology course and volunteered to participate for extra course credit.

Apparatus. All speech sounds were produced on-line during the experiment by a formant series resonator speech synthesizer (FONEMA-OVE-IIId) controlled by a DEC PDP-8/L computer (Cohen & Massaro, 1976). Segment durations were always multiples of 8 msec. The stimuli were defined as a series of parameter vectors, each specifying a target value and transition time, with linear, positive or negatively accelerated transitions. Intermediate values were computed and fed to the synthesizer at 8-msec intervals. The output of the synthesizer was amplified (McIntosh MC-50) and bandpass filtered between 20 Hz and 10 kHz (Krohn-Hite 3500R) and presented over headphones (Koss PRO-4AA) at a comfortable listening level (about 72 dB-SPL-A). Four subjects could be tested simultaneously in separate sound-attenuated rooms.

Stimuli. Each speech sound was a consonant cluster syllable beginning with one of the four consonants /p/, /t/, /s/, or /v/,

followed by a glide consonant ranging (in seven levels) from /l/ to /r/, followed by the vowel /i/. Figure 1 gives schematic diagrams of the stimuli used for the first group of three subjects. The formant parameters F1, F2, and F3 for the initial consonants /t/, /s/, and /v/ are plotted in the left panel. Also given are the frication, voicing, and aspiration amplitudes AC, AV, and AH, respectively, as well as the fundamental frequency F0. The parameters for the /pli/ to /pri/ continuum are plotted in the right panel. The /li/ to /ri/ continuum is the segment to the right of point X on the abscissa in the /p/ diagram. This segment was identical for each of the four initial consonants. That is, each of the four consonants was combined with the glide-vowel segment at the point X to produce the synthetic consonant-glide-vowel syllable. The initial values of F3 at the onset of the glide were 2397, 2263, 2136, 2016, 1903, 1796, and 1695 Hz, from the sound most like /l/ to the sound most like /r/. These seven values are illustrated for each initial consonant. For the second group of three subjects, the stimuli were identical except that F3 was fixed at 2016 Hz during the first consonant and did not change until the first consonant was finished (point X in Figure 1). The F3 was then changed immediately to the value designated by one of the seven sounds of the glide continuum. The voicing amplitude (AV) and aspiration amplitude (AH) shown in Figure 1 refer to synthesizer control values only, not amplitudes at the ear. Not shown in Figure 1, the fourth and fifth formants were fixed at 3500 and 4000 Hz, respectively. The fricative pole/zero ratios for the consonants /t/, /s/, and /v/ were 0, 12, and 8 dB, respectively.

Procedure. On each trial, a syllable was randomly selected without replacement from the set of 28 syllables generated from the factorial combination of the four initial consonants and the seven F3 levels of the following glide. The computer waited until each subject responded. The response interval averaged between 1 and 2 sec. An additional 1-sec interval intervened before the next trial. On the first day, subjects responded by pressing one of eight buttons labeled PLE, PRE, TLE, TRE, SLE, SRE, VLE, and VRE. On the second day, subjects responded with one of two buttons labeled L and R.

In order to familiarize themselves with synthetic speech, the subjects first listened to the entire set of stimuli twice. The sounds were presented in a fixed order with the seven labels of F3 defining the /l/-/r/ continuum as the fastest moving variable. The subjects were told that these sounds were a subset of the sounds involved in the experiment and that the stimulus order in the experiment was entirely random. The subjects were told that there were four possible consonants in initial position followed by either /l/ or /r/, followed by /i/. Their task was to identify the syllable on the basis of what they heard. They were told that there was no correct response and simply to make the best judgment they could. The subjects were then given a practice session of 28 trials before the first session of the first day. On both days, there were two sessions of 280 trials, consisting of 10 blocks of the 28 stimuli. However, data from the second session on the second day were lost and do not contribute to the results.

Results

The results of Day 1 with eight responses allow an assessment of how well the initial consonant was identified. The identification of the initial consonant was very good, averaging .98, .94, 1.00, and .98 for /p/, /t/, /s/, and /v/, respectively. Given the very good identification of the initial consonant, the eight responses on Day 1 were summed across identification of the initial consonant and combined to give the proportion of /r/ identifications at each of the 28 experimental conditions. These results were then combined with the proportion of /r/ responses from Day 2. Figure 2 gives the proportion of /r/ identifications for the fixed versus varying acoustic representation of the initial consonant context as a function of initial consonant context and the seven levels of the F3 onset defining the /r/-/l/ continuum.

The first question for this study was whether the acoustic structure of the initial consonant modifies the effect of phonological context. For one group of subjects, the F3 value during the initial consonant was equivalent to that given by the following glide. For the other group of subjects, the F3 value during the initial consonant was always set at a fixed value regardless of the F3 of the following glide sound. As can be seen in a comparison between the two panels of Figure 2, the context effects were equivalent for

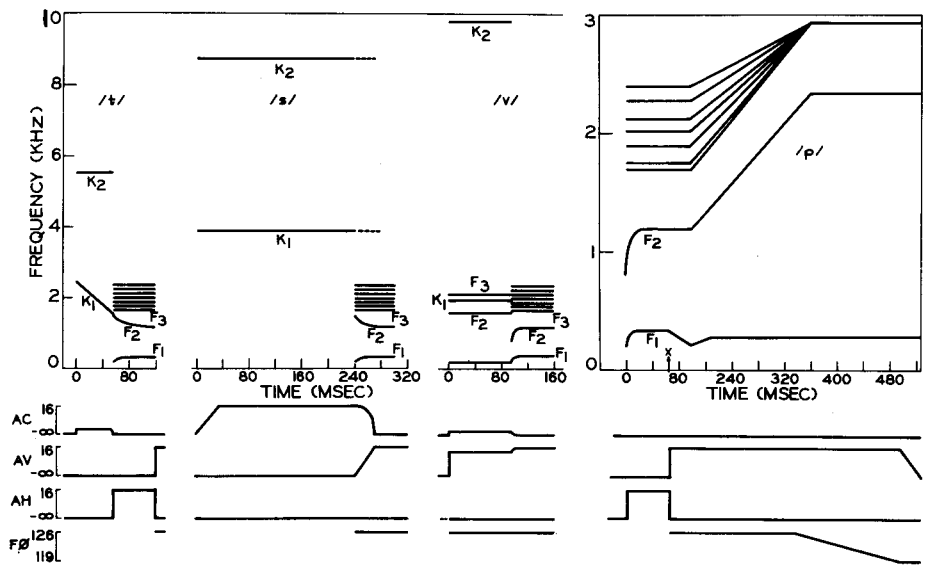

Figure 1. Schematic spectrographs of the speech sounds used in Experiments 1 and 2.

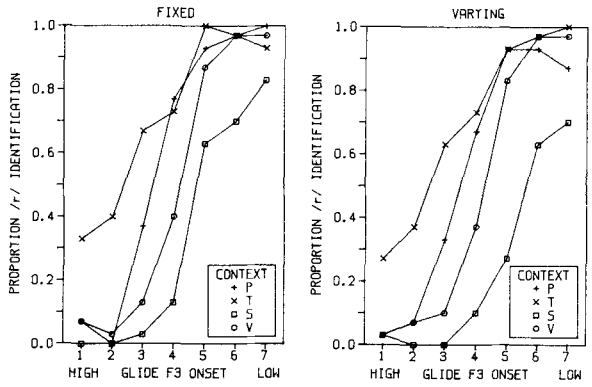

Figure 2. The proportion of /r/ identifications for the fixed versus the varying representation of the initial consonant as a function of the initial F3 transition during the glide; the initial consonant is the curve parameter (Experiment 1).

the two different acoustic representations of the initial consonant. Analyses of variances showed that there were no significant differences (all Fs < 1) between the two acoustic representations of phonological context in terms of response (/r/ or /l/), response as a function of the initial consonant or F3 of the glide, or the triple interaction of these factors. Therefore, the identification of the sounds along the /l/-/r/ continuum and the contribution of phonological context did not depend on whether the acoustic structure of the initial consonant was fixed or varying. Thus, we have evidence that the context effect is not due to simply the acoustic structure of the initial consonant.

The left panel of Figure 3 shows the proportion of /r/ identifications as a function of the F3 transition and initial consonant context for Day 1. An analysis of variance was carried out on the proportion of identifications treating the eight response alternatives, the four consonant contexts, and the seven levels of F3 as factors. There was a strong effect of phonological context; responses were not equally distributed across the eight alternatives [$F(7,104) = 108.04$, $p < .01$]. Overall, the proportion of /r/ identifications was greatest for /t/, smallest for /s/, and intermediate for /p/ and /v/. As expected, the proportion of /r/ identifications increased with decreases in the starting frequency of the F3 transition [$F(42,168) = 31.45$, $p < .001$]. The interaction of F3 transition and phonological context was also significant [$F(126,504) = 33.04$, $p < .001$]. This reflects the fact that the effect of the initial consonant was greatest for intermediate levels of the F3 transition.

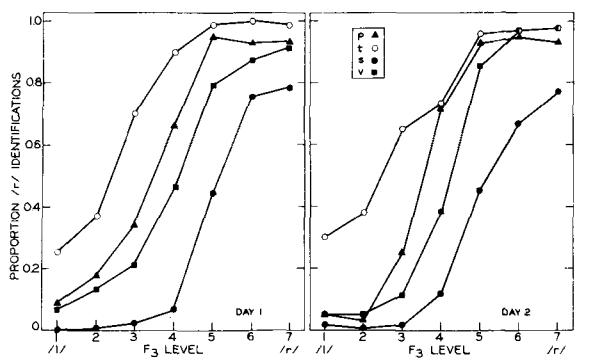

Figure 3. The proportion of /r/ identifications for Days 1 and 2 as a function of the F3 transition during the glide; the initial consonant is the curve parameter (Experiment 1).

The right panel of Figure 3 presents the results of Day 2 of the experiment. With only two possible responses, contextual effects of the initial consonant on the response were balanced, and overall /l/ and /r/ responses did not differ significantly [$F(1,4) = 7.09$, $p < .16$]. As on Day 1, responses varied significantly as a function of the F3 transition [$F(6,24) = 31.09$, $p < .001$], the initial consonant [$F(3,12) = 7.09$, $p < .01$], and the combination of these two factors [$F(18,72) = 3.53$, $p < .001$].

Comparing the two panels of Figure 3 shows that the results are very similar for the 2 days of the experiment. It appears that neither practice in the task nor whether or not the context must be overtly identified is a critical factor for the observation of a strong contribution of phonological context in speech recognition. The next experiment was carried out to provide an independent replication of the first experiment and to provide results to assess quantitative models of how phonological context contributes to speech perception.

EXPERIMENT 2

Method

Subjects. Seven students from an introductory psychology class participated on 2 consecutive days for extra course credit.

Stimuli. The stimuli were essentially the same as those used for Group 2 in Experiment 1 (fixed F3 during the initial consonant) with the following changes: The F2 transition for /t/ was linear rather than nonlinear, and the starting frequencies for the F3 transition of the glide were changed. On the first day, the starting frequencies along the /l/-/r/ continuum were 2851, 2540, 2329, 2198, 2074, 1093, and 1695 Hz. During the initial consonant, the F3 frequency was set at 2198 Hz. On Day 2 the starting frequencies of F3 were 3109, 2770, 2540, 2397, 2263, 2075, and 1849 Hz, with the F3 during the initial consonant set at 2397 Hz.

Procedure and Apparatus. The procedure and apparatus were the same as those used on Day 1 of Experiment 1, with eight response alternatives.

Results

The proportion of identifications was analyzed as a function of the 28 experimental conditions. The results of Experiment 1 were replicated exactly. As in Experiment 1, the recognition of context was very good, averaging about 95% correct. The points in Figure 4 represent the probability of an /r/ identification as a function of both the F3 transition of the glide and the initial consonant context for each of the 2 days of the experiment. Replicating the results of Experiment 1, identifications were an orderly

function of the F3 transition and the initial consonant. Phonological context effects were largest at the more ambiguous levels of the F3 transition. These effects were statistically significant. For Day 1, the proportion of responses differed significantly [$F(7,42) = 8.46$, $p < .001$], as did response as a function of the F3 value of the glide [$F(42,252) = 27.72$, $p < .001$], the initial consonant [$F(21,126) = 72.60$, $p < .001$], and the combination of these two factors [$F(126,252) = 27.39$, $p < .001$]. Similarly, on Day 2, the proportion of responses differed significantly [$F(7,42) = 4.79$, $p < .01$], along with response as a function of F3 transition [$F(42,252) = 29.80$, $p < .001$], the initial consonant [$F(21,126) = 62.70$, $p < .001$], and the combination of these two factors [$F(126,252) = 29.61$, $p < .001$].

Discussion

The results of the first two experiments showed large effects of acoustic featural information and phonological context on the identification of the test consonant. The significant interaction of these two variables revealed that the magnitude of the context effect was largest at the more ambiguous levels of stimulus information. The context effect did not appear to decrease with experience in the experiment. In the following discussion, several models will be quantified within the framework of the fuzzy logical model (Oden & Massaro, 1978), and tested against the results of the experiment. The models will be formulated to predict the likelihood of an /r/ identification, since recognition of the context was nearly perfect.

Contextual feature models. In the first set of models, we assume that two independent sources of information are available: featural information from the glide segment and featural information representing the phonological context. The first source of information can be represented by T_i, where the subscript i indicates that T_i changes only with the F3 transition. For the /l/-/r/ identification, T_i specifies how much R-ness is given by the critical F3 transition feature. This value lies between 0 and 1 and is expected to increase as the starting frequency of the F3 transition is decreased. With just two alternatives along the continuum, it is reasonable to assume that the amount of L-ness given by the featural information is simply 1 minus the amount of R-ness given by that same source (see Appendix). Therefore, if T_i specifies the amount of R-ness given by the F3 transition, then $(1 - T_i)$ specifies the amount of L-ness given by that same transition.

The phonological context provides independent evidence for R and L. The value C_j represents how much the context supports the consonant R. The subscript j indicates that C_j changes only with changes in phonological context. The value of C_j lies between 0 and 1 and should be large when R is admissible and small when R is inadmissible. The degree to which the phonological context supports the consonant L is indexed by D_j and is independent of the value of C_j. The value of D_j also lies between zero and one and should be large when L is admissible and small when L is inadmissible.

It is assumed that the featural information derived from the glide segment is independent of the information derived from the phonological context. In addition, the listener is assumed to have access to these two independent sources of information. During the feature evaluation operation, the amount of R-ness and L-ness is evaluated from each of these two sources. The amount of R-ness and L-ness for a given syllable can therefore be represented by the conjunction of the two independent sources of information:

$$\text{R-ness} = (T_i \wedge C_j) \tag{1}$$

$$\text{L-ness} = [(1 - T_i) \wedge (D_j)] \tag{2}$$

At the prototype matching operation, the sources of information are conjoined, using a multiplicative combination rule:

$$\text{R-ness} = T_i \times C_j \tag{3}$$

$$\text{L-ness} = (1 - T_i) \times (D_j) \tag{4}$$

The outcome of prototype matching is made available to the pattern classification operation. A choice of R is assumed to be made by evaluating the degree of R-ness relative to the sum of R-ness and L-ness values. In this case, the probability of an R response, P(R), can be expressed as:

$$P(R) = \frac{T_i C_j}{T_i C_j + (1 - T_i)(D_j)}. \tag{5}$$

General context model. In the general form of the contextual feature model, unique C_j and D_j values are required for each of the four different initial consonant contexts. Seven values of T_i are also required for the seven starting frequencies of the F3 transition of the glide. Fitting the model to the observed data therefore requires the estimation of 15 parameters. The model was fit to the proportion of /r/ identifications from Experiment 2 as a function of the initial context and the F3 transition. The predictions of the model were obtained by estimating parameters using the iterative routine STEPIT (Chandler, 1969). The parameter values are adjusted to minimize the squared deviations between the observed and predicted values. The model was fit to each subject's data individually for each of the 2 days of the experiment. Table 1 gives the root mean squared deviation (RMSD) for this general contextual feature model for each subject on each of the 2 days. The

RMSD was obtained by summing the squared deviations between the predicted and observed values across each of the 28 conditions, dividing by 28, and taking the square root of this value. The average RMSD over subjects for this model was .073 for Day 1 and .043 for Day 2.

Complement model. In a second form of the contextual feature model, the general contextual feature model is modified so that the degree to which the context supports the inadmissible alternative is 1 minus the degree to which the context supports the admissible alternative. The contextual information is represented by C_j, where $0 \leqslant C_j \leqslant 1$, and the subscript signifies that the value of C can change with the context j. The value of C_j represents the degree to which the context is compatible with the admissible alternative; 1 minus this value represents the degree to which the context is compatible with the inadmissible alternative. This model is identical to the general contextual model, except that it is assumed that D_j is equal to $1 - C_j$. The fit of this model to the results is identical to the fit of the general model, with four fewer parameters (see Table 1). As shown in the Appendix, the present experiment cannot discriminate between the general context and complement models. Thus, the complement model will be preferred, since it is the more parsimonious of the two models.

Admissible-inadmissible model. Massaro (1979) applied a special form of the contextual feature model to a similar study of letter perception in reading. In this form of the model, the context is considered to be either admissible or inadmissible. A given alternative is supported to the degree x by an admissible context and to the degree y by an inadmissible context, where $1 \geqslant x > y \geqslant 0$. The values x and y do not have subscripts, since they depend only on the admissibility of the context. Therefore, C_j is equal to x when R is admissible in a particular context and equal to y when R is inadmissible. Analogously, D_j is equal to x when L is admissible in a particular context and equal to y when L is inadmissible. Given this assumption, the derivation of P(R) for the four phonological contexts in the present experiment is analogous to that given by Massaro (1979).

This model predicts a context effect to the extent that an admissible context gives more evidence for a particular test consonant than does an inadmissible context, that is, to the extent $x > y$. A second feature of this model is that P(R) is entirely determined by the F3 transition information when the context supports either both or neither of the test alternatives; accordingly, P(R) is predicted to be identical for the consonant contexts /p/ and /v/.

This form of the contextual feature model was tested against the observed results. In order to fit the model to the data, it was necessary to estimate nine parameters: x and y and seven values of T_i, one for each level of the F3 transition. The average RMSD value and the parameter estimates are also given in Table 1. The average RMSD across subjects for the model was .112 for Day 1 and .096 for Day 2. The admissible-inadmissible model gives a significantly poorer description of the results than does the complement model.

The relatively poor description of the admissible-inadmissible model is due to the differences observed for the /v/ and /p/ contexts. Although both /r/ and /l/ are inadmissible in the context /v/ and both are admissible in the context /p/, these contexts do not have equivalent effects. This result contrasts with Massaro's (1979) finding that contexts that were admissible or inadmissible for both alternatives gave equivalent results for letter perception. There are two reasons why the contexts /v/ and /p/ could have different influences on identification of the following glide. First, it could be that intial consonants /p/ and /v/ differed with respect to some auditory property that differentially affected perceptual recognition. As an example, the initial consonant /p/ may have provided slightly more coarticulatory evidence for /r/ than for /l/ relative to that provided by /v/. Hence, we would expect slightly more /r/ responses following initial /p/ than following initial /v/. Second, in natural English, initial /p/ is more likely to be followed by /r/ than by /l/ (Roberts, 1965). Therefore, subjects may be somewhat biased to hear /r/ after /p/ even though both /r/ and /l/ are admissible. Since this same bias would not be present for the context /v/, the result would be slightly more /r/ responses following /p/ than following /v/. Therefore, the relatively small differences that were observed for the contexts /p/ and /v/ might have

Table 1
Root Mean Squared Deviation (RMSD) for Each Subject for Each of the Two Days of Experiment 2 for Three Contextual Feature Models

		Model	
	Subject	General or Complement	Admissible-Inadmissible
Day 1	1	.048	.069
	2	.055	.102
	3	.124	.139
	4	.072	.079
	5	.043	.170
	6	.134	.155
	7	.035	.071
	Mean	.073	.112
Day 2	1	.035	.045
	2	.050	.153
	3	.059	.139
	4	.023	.055
	5	.015	.076
	6	.100	.104
	7	.020	.097
	Mean	.043	.096
	Number of Parameters	15 or 11	9

been due to some auditory differences between the two contexts or to differences in the frequency of occurrences of sound sequences in English.

Modifier models. Phonological context may have its influence through modifiers on the featural information available for a given sound. In the prototype, the feature corresponding to the F3 transition of the glide could contain a modifier when it occurs in phonologically inadmissible contexts. The requirement would be for a better match of F3 when the alternative is phonologically inadmissible than when it is admissible. It follows that a higher starting F3 frequency is necessary to perceive /l/ in the context /t/ than to perceive it in the context /p/. Equivalently, a lower value of F3 is necessary to perceive /r/ in the context /s/ than in the context /p/.

General modifier model. The formalization of this assumption involves adding modifiers in the prototype descriptions of the evidence for R and L in the phonological contexts in which the alternatives are inadmissible. In this case,

$$\text{R-ness} = T_i^{E_j} \tag{6}$$

when the context is inadmissible for /r/ and

$$\text{L-ness} = (1 - T_i)^{E_j} \tag{7}$$

when the context is inadmissible for /l/. The j subscript on the exponent signifies that the value of the exponent can change for different contexts.

It is assumed that a better match of the appropriate F3 transition is required for R or L to be heard in an inadmissible phonological context. For example, whereas $(1 - T_i)$ gives the amount of evidence supporting L following the stop /p/, $(1 - T_i)^{E_j}$ gives the amount of evidence supporting L following the stop /t/. If $E_j > 1$, then, for a given F3 transition, the probability of an /l/ identification will be less in the context /t/ than in the context /p/.

The general modifier model was fit to the identification results by estimating seven parameters for T_i and four parameters for the modifiers in the prototypes descriptions. The RMSD values are given in Table 2. The average RMSD values were .081 and .062 for Days 1 and 2, respectively.

Specific modifier model. In a more specific form of the prototype modifier model, no exponents for R and L are assumed in the context /v/. This assumption might be justified on the grounds that both R and L are inadmissible in the context /v/. All other aspects of this more specific model are equivalent to that of the general modifier model. Only two parameters are needed for the modifiers in addition to the seven parameters for the seven levels of F3. Table 2 gives the RMSD values; the average values are .112 and .093 for Days 1 and 2, respectively. This model gives a significantly poorer description than does the general modifier model that has exponents for the alternative /v/.

Table 2
Root Mean Squared Deviation (RMSD) for Each Subject for Each of the Two Days of Experiment 2 for Two Modifier Models

		Model	
	Subject	General Modifier	Specific-Modifier
Day 1	1	.057	.080
	2	.041	.072
	3	.110	.178
	4	.068	.074
	5	.123	.162
	6	.115	.152
	7	.053	.068
	Mean	.081	.112
Day 2	1	.045	.056
	2	.064	.122
	3	.073	.130
	4	.040	.063
	5	.067	.071
	6	.120	.123
	7	.058	.085
	Mean	.062	.093
	Number of Parameters	11	9

The best contextual feature model gives a better description of the results than does the best prototype modifier model. The comparison between the complement model and the general modifier model is straightforward, since the same number of parameters is used in both models. Using average RMSD as a metric, the complement model gives about a 25% better description of the results than does the modifier model. Figure 4 gives the average observed results and the predicted values for the complement

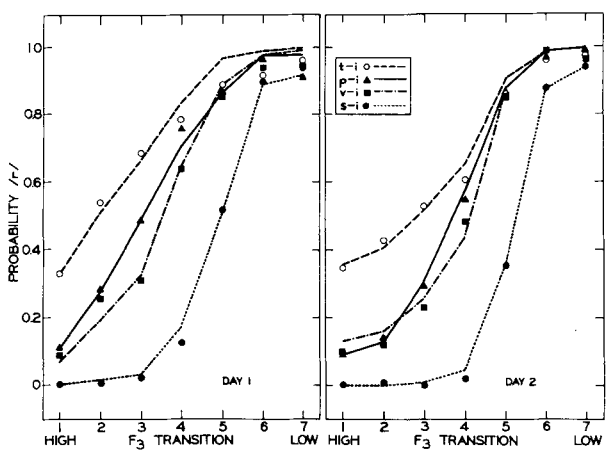

Figure 4. The observed (points) and predicted (lines) probabilities of an /r/ identification for Days 1 and 2 as a function of the F3 transition onset during the glide; the initial consonant is the curve parameter. The predictions are given by the complement feature model (Experiment 2).

Table 3
Average Parameter Estimates for Days 1 and 2 of Experiment 2 for the Complement Model

F3 Level	Context	Day 1	Day 2
1		.025	.028
2		.092	.046
3		.219	.150
4		.612	.363
5		.892	.854
6		.981	.989
7		.985	.996
	/p/	.688	.694
	/t/	.823	.800
	/s/	.095	.075
	/v/	.590	.649

Note–The parameter values represent the degree of R-ness, which can vary between zero and one. F3 level 1 = /l/; F3 level 7 = /r/.

feature model. Figure 4 shows that the model provides a good description of the results. Table 3 gives the average parameter estimates for the fit of the complement model. The parameter estimates of the model shown in Table 3 are meaningful. The T_i values, representing the degree of R-ness, increase systematically with decreases in the starting frequency of F3. The C_j values change systematically with phonological context; the degree of R-ness given by context is much larger for initial /t/ than for initial /s/. Relative to the context /v/, the context /p/ is somewhat more supportive of /r/ than of /l/.

EXPERIMENT 3

In the first two experiments, a distinction was made between the phonological context and the test sound. The contribution of phonological context was studied with an unambiguous speech sound specifying the context. The results showed that the contribution of phonological context was largest when the test sound was most ambiguous. The goal of Experiment 3 was to evaluate the contribution of phonological constraints between two adjacent speech sounds when each speech sound is independently varied between two alternatives. Consider the stop consonant /b/ or /d/ in initial position and the glide /l/ or /r/ in second position. The consonant clusters /bl/, /br/, and /dr/ are admissible, whereas the cluster /dl/ is inadmissible in word-initial position in English. Subjects should be less likely to hear /dl/ relative to the other three alternatives. The question of interest is how the information about appropriate phonological sequences is combined with the auditory information in the recognition of these alternatives. To answer this question, a continuum of sounds between /b/ and /d/ in initial position was factorially combined with a continuum between /l/ and /r/ in second position. The results will be used to evaluate quantitative models of the integration of auditory information from adjacent segments as a function of different degrees of phonological constraint between the segments.

Method

Subjects. Eight subjects from an introductory psychology class served on 2 consecutive days for extra course credit.

Stimuli and Apparatus. The speech sounds used in this experiment were consonant-cluster syllables beginning with a voiced stop ranging between /b/ and /d/ in five steps, followed by a glide consonant ranging between /l/ and /r/ in five steps, followed by the vowel /a/. Figure 5 gives a general schematic diagram of the syllables used. Each of the five levels of initial stop consonant could occur in combination with each of the five levels of the following glide consonant, for a total of 25 different syllables.

The initial values of F2 used for the stop consonant were 1425, 1600, 1796, 2016, and 2263 Hz, from most /b/-like to most /d/-like. The initial value of F1 for the stop was 200 Hz, and F3 during the stop was fixed at 2397 Hz. The negatively accelerated stop transitions took F1 and F2 to 317 and 1234 Hz, respectively, over a 50-msec period. The amplitude of the stop went linearly from silent to full intensity in 10 msec. At the beginning of the glide consonant, F3 was initially set to 2770, 2614, 2397, 2198, or 2016 Hz, from the most /l/-like to most /r/-like sound. During the first 30 msec of the glide, F3 was fixed. Then F3 followed linear transition to 2397 Hz over a 120-msec period. During the first 20 msec of the glide, F1 followed a linear transition to 275 Hz, where it remained for 10 msec. Next, F1 followed a linear transition to 777 Hz over a 120-msec period. Following this transition, the vowel remained on for 220 msec, followed by a 20-msec transition to silence. During the final 120 msec of the vowel, the F0 went from its initial value of 126 Hz to 119 Hz, following a linear transition.

The apparatus used was the same as in Experiments 1 and 2, except that segment durations were always multiples of 5 msec in the speech synthesis.

Procedure. On each trial, a syllable was selected randomly without replacement from a set of 25 syllables generated from the factorial combination of the five initial consonants and the five glides. The computer waited until each subject responded. An additional 1-sec interval intervened before the next trial. The subjects responded by pressing one of four buttons labeled BL, BR, DL, and DR. The subjects were given a practice session of 25 trials before the first session on the first day. On each of 2 days, there were two sessions of 350 trials consisting of 14 blocks of 25 stimuli. Unknown to the subjects, each experimental session was preceded by five unscored trials.

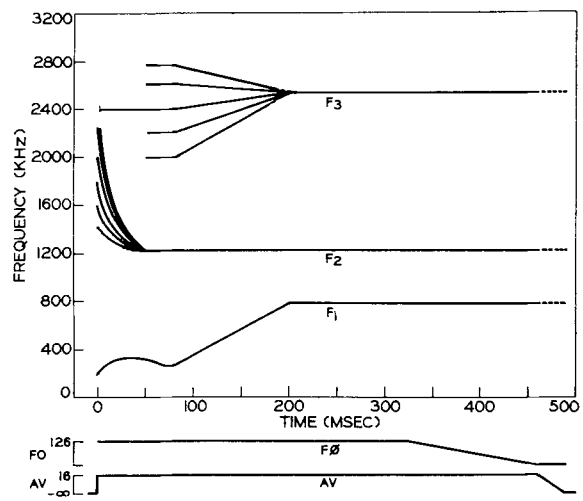

Figure 5. Schematic spectrograph of the speech sounds used in Experiment 3.

Results

Figure 6 gives the proportion of identifications of each of the four alternatives as a function of the F2 onset level during the stop segment and the F3 onset level during the glide segment. The figure shows that both variables had a significant effect on performance. Although the figure is informative, the exact interaction of these two variables is easier to see in Figure 7. This figure gives the proportion of /d/ identifications and /r/ identifications separately. The left panel of Figure 7 shows the proportion of /d/ identifications as a function of stop and glide levels. There was a significant increase in the proportion of /d/ responses, from .157 to .875 with increases in the onset frequency of the stop F2 transition [$F(4,16) = 80$, $p < .001$]. There was also a significant increase in the proportion of /d/ responses, from .451 to .613, with decreases in the onset frequency of the glide F3 transition [$F(4,16) = 11.473$, $p < .005$]. The interaction between the stop level and glide level was significant [$F(16,64) = 2.755$, $p < .005$]

The right panel of Figure 7 shows the results in terms of the proportion of /r/ identifications as a function of stop and glide. The proportion of /r/ identifications increased significantly, from .140 to .885, with decreases in the onset frequency of the glide F3 transition [$F(4,16) = 22.837$]. The proportion of /r/ responses did not differ significantly with changes in the onset frequency of the stop F2 transition [$F(4,16) = 1.615$, n.s.], although the interaction of stop and glide was significant [$F(16,64) = 2.550$, $p < .01$]. Figure 7 shows that there was a somewhat larger increase in /r/ responses with increases in the onset frequency of the stop F2 transition at the second and third levels of the glide.

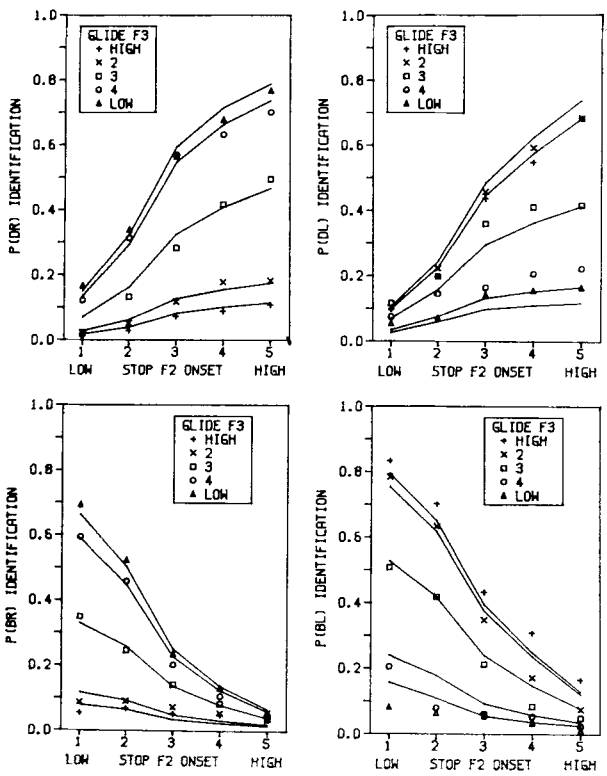

Figure 6. Observed results (points) and predictions (lines) given by the contextual feature model in Experiment 3.

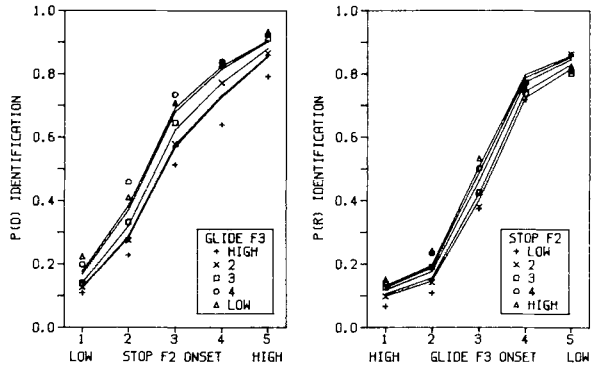

Figure 7. Left panel: The observed (points) and predicted (lines) probability of /d/ identifications as a function of the F2 onset during the stop consonant; the F3 onset during the glide is the curve parameter. Right panel: The observed (points) and predicted (lines) probability of /r/ identifications as a function of the F3 onset during the glide consonant; the F2 onset during the stop is the curve parameter.

Discussion

Fuzzy logical models. Variants of the fuzzy logical model can be constructed to account for the results of Experiment 3. We assume that the listener has established prototypes corresponding to the four alternatives /bla/, /dla/, /bra/, and /dra/. Each prototype contains a cue to the stop and a cue to the glide portion of the sound in addition to the other cues, such as the vowel portion. The latter cues are assumed to be constant for all four alternatives so that it is sufficient to represent the prototypes as:

bla: (low stop F2) and (high glide F3), (8)

dla: (high stop F2) and (high glide F3), (9)

bra: (low stop F2) and (low glide F3), (10)

dra: (high stop F2) and (low glide F3), (11)

where F2 and F3 values refer to the respective stop and glide segments of the sound. This simple model can be fit to the results to provide a baseline for evaluation of other, more complex models. The simple model cannot be expected to provide a good description of the results, since a high F3 not only biased the judgment towards /l/ rather than /r/ but also towards /b/ rather than /d/. Also, a high F2 not only biased the judgment towards /d/ rather than /b/, but also biased the judgment towards /r/ rather than /l/. Thus, the judgment /dl/ is made less often than it should be according to the simple model. The mean

RMSD of the fits of each of the eight subjects for the simple model with 10 parameters was .080 (see Table 4).

A contextual feature model involves the inclusion of the contextual knowledge that /dla/ does not occur in word initial position in English. In this case, the prototype would be:

dla: high F2 and high F3 and not likely. (12)

In this case, an additional parameter is necessary for the knowledge "not likely." The fit of this model gave an average RMSD of .064, a significant improvement over the simple model (see Table 4).

A second modification of the simple model is to include prototype modifiers for the high F2 and high F3 cues for /dla/:

dla: very(high F2) and very(high F3). (13)

The two "very" modifiers mean that /dla/ requires a higher F2 than /dra/ and a higher F3 than /bla/, since /dla/ is inadmissible in word initial position in English. That is, for a given goodness of match, a better match of the acoustic features is required for an inadmissible cluster than for an admissible cluster. In terms of the quantitative model, the modifiers are instantiated as exponents on the fuzzy F2 and F3 values. This model adds two additional parameters to the simple model and gives an average RMSD of .064, significantly better than the description given by the simple model, and equivalent to that given by the contextual feature model.

The contextual feature model and the prototype modifier models give very similar descriptions of the results, although the contextual feature model requires one less parameter. For this reason, and because the contextual feature model did a better job for the results of Experiment 2, we prefer the contextual feature model. Figures 6 and 7 present the observed results and the predictions given by the contextual feature model. Table 5 presents the average parameter values used in the description of the results.

Table 4
Root Mean Squared Deviations (RMSD) for Each Subject in Experiment 3 for the Simple Model, the Contextual Feature Model, and the Prototype Modifier Model

	Model		
Subject	Simple	Contextual Feature	Prototype Modifier
1	.089	.074	.082
2	.111	.085	.069
3	.067	.061	.066
4	.081	.053	.061
5	.082	.057	.064
6	.073	.049	.062
7	.047	.047	.042
8	.087	.087	.068
Mean	.080	.064	.064
Number of Parameters	10	11	12

Table 5
Average Parameter Estimates for the Contextual Feature Model for Experiment 3

	Onset Level				
Parameter	1	2	3	4	5
			F2		
b-ness	.809	.598	.295	.165	.087
			F3		
l-ness	.196	.293	.611	.859	.904

Note—Parameter value for "not likely" is .555. Onset level 1 = low; onset level 5 = high.

As in our first two experiments, it has been shown that a preceding consonant may affect the perception of one that follows. Of greater interest, however, is the finding that the characteristics of a following consonant may affect the perception of one that precedes it. This result seems to be inconsistent with theories that postulate linear, unit-by-unit recognition of consonant phonemes. That is, recognition of the stop consonant could not have occurred before some processing of the glide segment of the syllable. It is more reasonable to assume that the prototype descriptions in the fuzzy logical model are larger than a single phoneme. Given this result and the results reviewed by Massaro (1975), there is a growing amount of evidence that the prototypes are syllables.

GENERAL DISCUSSION

The results of these experiments are relevant to contemporary issues in psychology, phonology, and artificial intelligence. One persistent issue in psychological theory is whether or not context modifies lower level feature analysis processes (Broadbent, 1967; Morton, 1969). The description of the results given here and research in other domains provide strong evidence that context effects occur independently of the lower level processes. That is, there is no evidence that context modifies lower level sensory processing in speech perception. The featural information is not modified by context; context simply provides additional information.

Recent theories of phonology (Chomsky & Halle, 1968; Ladefoged, 1975) have begun to give more weight to actual psychological performance, and the present results indicate that phonological constraints are psychologically real. One important question concerns the way in which knowledge about phonological context is stored. Do listeners have information about relative frequency of occurrence of sound sequences or are the phonological constraints stored

in terms of rules? One possible approach to studying this question is to attempt to separate these two kinds of information in the construction of test sequences. There has been some success in taking this tack in the study of orthographic constraints in reading (Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980).

Finally, with respect to artificial intelligence, it is now generally agreed that automatic speech recognition cannot be completely bottom-up but must involve the utilization of linguistic context in perception and recognition of the message (Klatt, 1977). One advantage of using phonological constraints is that these constraints operate among adjacent sound segments and, therefore, this information can be used early in the processing of the message. Other constraints, such as syntactic and semantic constraints, do not necessarily constrain adjacent sound segments and, therefore, do not offer much help in making decisions at the segment level early in processing. The present results suggest that phonological context might be successfully utilized in automatic speech recognition by machine.

REFERENCES

Broadbent, D. E. Word-frequency effect and response bias. *Psychological Review*, 1967, **74**, 1-15.

Brown, R. W., & Hildum, D. C. Expectancy and the perception of syllables. *Language*, 1956, **32**, 411-419.

Chandler, J. P. Subroutine STEPIT finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, **14**, 81-82.

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Cohen, M. M., & Massaro, D. W. Real-time speech synthesis. *Behavior Research Methods & Instrumentation*, 1976, **8**, 189-196.

Cole, R. A., & Jakimik, J. Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information-processing*. London: Academic Press, 1978.

Cole, R. A., & Scott, B. The phantom in the phoneme: Invariant cues for stop consonants. *Perception & Psychophysics*, 1974, **15**, 101-107.

Ganong, W. F., III. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, **6**, 110-125.

Klatt, D. H. Review of the ARPA speech understanding project. *Journal of the Acoustical Society of America*, 1977, **62**, 1345-1366.

Ladefoged, P. *A course in phonetics*. New York: Harcourt, Brace, and Jovanovich, 1975.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. *Perception of the speech code*, 1967, **74**, 431-461.

Marslen-Wilson, W., & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 1978, **10**, 29-63.

Massaro, D. W. (Ed.). *Understanding language: An information processing analysis of speech perception, reading and psycholinguistics*. New York: Academic Press, 1975.

Massaro, D. W. Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 595-609.

Massaro, D. W., & Oden, G. C. Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, 1980, **67**, 996-1013. (a)

Massaro, D. W., & Oden, G. C. Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3). New York: Academic Press, 1980. (b)

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. *Letter and word perception: Orthographic structure and visual processing in reading*. Amsterdam: North-Holland, 1980.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165-178.

Oden, G. C., & Massaro, D. W. Integration of featural information in speech perception. *Psychological Review*, 1978, **85**, 172-191.

Pollack, I., & Pickett, J. M. The intelligibility of excerpts from conversation. *Language and Speech*, 1964, **6**, 165-171.

Roberts, A. H. *A statistical analysis of Americann English*. The Hague: Mouton, 1965.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 1978, **64**, 1358-1368.

Whorf, B. L. *Language, thought and reality: Selected papers*. New York: Wiley, 1956.

APPENDIX

It can be shown that a factorial design manipulating one test stimulus variable and one context variable with two response alternatives cannot test between the general model and the complement model. In the general model,

$$P(R) = \frac{T_i C_j}{T_i C_j + (1 - T_i) D_j}, \tag{1a}$$

while in the complement model,

$$P(R) = \frac{T_i C_j}{T_i C_j + (1 - T_i)(1 - C_j)}. \tag{1b}$$

Dividing the numerator and denominator of Equations 1a and 1b by $(1 - T_i)C_j$ gives

$$P(R) = \frac{T_i/(1 - T_i)}{T_i/(1 - T_i) + D_j/C_j} \tag{2a}$$

and

$$P(R) = \frac{T_i/(1 - T_i)}{T_i/(1 - T_i) + (1 - C_j)/C_j}. \tag{2b}$$

for Equations 1a and 1b, respectively. The identity of Equations 2a and 2b rests on the identity of D_j/C_j and $(1 - C_j)/C_j$. Given that each of these ratios is indexed by a single subscript j, a single parameter is sufficient to specify each of their values. Therefore, one parameter is all that is needed, and, therefore, the D_j value adds nothing to the predictive power of Equation 1a relative to Equation 1b.

(Manuscript received June 28, 1982;
revision accepted for publication April 15, 1983.)