

## Individual variability in the perception of cues to place contrasts in initial stops

VALERIE HAZAN and STUART ROSEN  
*University College London, London, England*

Synthetic continua of two minimal pairs, BAIT-DATE and DATE-GATE, closely modeled on natural utterances by a female speaker, were presented to a group of 16 listeners for identification in *full-cue* and *reduced-cue* conditions. Grouped results showed that categorization curves for full- and reduced-cue conditions differed significantly in both contrasts. However, an averaging of results obscures marked variability in labeling behavior. Some listeners showed large changes in categorization between the full- and reduced-cue conditions, whereas others showed relatively small or no changes. In a follow-up study, perception of the BAIT-DATE contrast was compared with the perception of a highly stylized BA-DA continuum. A smaller degree of intersubject and between-condition variability was found for these less complex synthetic stimuli. The amount of variability found in the labeling of speech contrasts may be dependent on cue salience, which will be determined by the speech pattern complexity of the stimuli and by the vowel environment.

Speech contrasts are marked by a multiplicity of cooperating cues. A number of studies have investigated the perceptual importance of single cues by removing or neutralizing other cues to the contrast or by presenting them in conflicting cue experiments. It appears, generally speaking, that better categorization and discrimination are obtained in cooperating cue conditions than they are in single or conflicting cue conditions (e.g., Eilers, Oller, Urbano, & Moroff, 1989; Morrongiello, Robson, Best, & Clifton, 1984). However, the behavior of individual listeners has not been well quantified. Perceptual weighting experiments have typically been carried out by using a group of listeners considered homogeneous with regard to variables such as age range, hearing threshold, linguistic background, and exposure to synthetic speech. Results are usually presented in the form of data averaged over the whole test group. The emphasis, derived from early Haskins Laboratories experiments (e.g., Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Liberman, Harris, Kinney, & Lane, 1961), has often been on general trends in labeling behavior. Results were not claimed to be "normative data on how synthetic speech was perceived by the population in general but to see how subjects who categorize members of a set of stimuli discriminated between them" (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). This desire for uniform results has sometimes led to stringent subject selection, with more than half of the subjects discarded (e.g., Liberman, Harris,

Eimas, Lisker, & Bastian, 1961). Less stringent selection has occurred in other studies, although subjects may still be eliminated if they respond in "an inappropriate manner" to control stimuli (e.g., Walley & Carrell, 1983).

Despite this emphasis on averaged results, the existence of individual variability in the perception of reduced-cue stimuli has long been recognized and has been mentioned in a number of perceptual studies that cover a wide range of contrasts. Zlatin (1974) mentioned considerable individual variation in the identification of initial plosives that differ in voicing and cautioned against the reporting of averaged data. Stevens and Klatt (1974), in their investigation of first formant frequency (F1) onset and burst duration as cues to the voicing contrast, showed individual results for two subjects: For the first, identification was strongly influenced by F1 onset, whatever the voice onset time (VOT) of the stimuli; for the second, identification was based on VOT, whatever the status of F1 onset. Haggard, Ambler, and Callow (1969) found that fundamental frequency could be used as a voicing cue only by some subjects and cautioned that individual differences may be found for all but the most robust cues. Van Tasell, Hagen, Koblas, and Penner (1982) and Kewley-Port (1981) mentioned substantial variability in the perception of truncated stimuli differing in place of articulation, although all listeners could label the full-cue stimuli. Walley and Carrell (1983), in their study on the perception of cues to a place contrast in initial stops by adults and children, stated that, although a majority of subjects were consistent in their use of either a formant- or onset-spectrum-based rule, poorer average identification for the group was attributable to individual differences in the cues used. Repp and Lin (1988) also mentioned substantial variability in the identification of initial burst transients. Fitch, Halwes, Erickson, and Liberman (1980), in their

This work was partly funded by Alvey project MMI/009 on Speech Pattern Algorithmic Representation. The second author is supported by the Medical Research Council of Great Britain. We would like to thank Adrian Fourcin for his helpful comments. We also thank the listeners who participated in this experiment. Correspondence may be addressed to either author at the Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, U.K.

study of cues to stop consonant manner contrasts, noted that individual subjects varied in the extent of the phoneme shift between test conditions, although all showed the same effect. One of the few studies to have focused specifically on the issue of individual differences in perception (Simon & Fourcin, 1976) compared speech intelligibility in noise with performance on identification tests that used synthesized minimal pairs (BU-DU, GOAT-COAT, rise/fall). Results showed individual differences in all tests and greatest variability in the reduced-cue condition of the GOAT-COAT voicing contrast.

Although the issue of listener variability has been mentioned in the context of discussions of experimental results, there has been little attempt to assess, in controlled experiments, the extent of this variability. A finding that listeners can differ significantly in their ability to make use of certain cues would have important implications. First, the existence of true variability between listeners in their use of perceptual cues could go some way toward explaining sometimes contradictory results obtained in cue-weighting studies. It could also be an important factor in explaining gross differences in performance obtained in clinical trials such as those involving cochlear implant patients and profoundly hearing-impaired patients, who are often tested on their perception of specific speech features. Finally, it has theoretical implications, as theories of speech perception and of speech development would need to account and allow for differences in listener strategies.

The main objective of this study, then, was to determine the extent of subject variability in the use of cues to place contrasts in initial plosives. Because it has been argued that this variability might be due to the poor quality of some highly stylized synthetic speech stimuli (Studdert-Kennedy et al., 1970), synthetic stimuli were prepared that were modeled on the utterances of a specific speaker and that were of a similar level of pattern complexity to natural speech tokens. Given that a reduction in identification ability is the norm in reduced-cue stimuli, the question to be asked is whether or not the extent of the cue-reduction effect will differ significantly within a group of normal-hearing listeners. Furthermore, if variability is significant, one will want to establish to what extent it is the result either of listener variability in the cues normally used or of a general ability to make use of reduced information.

## EXPERIMENT 1

### Method

**Subjects.** Sixteen subjects were tested during five 45-min sessions and were paid for their participation in the experiment. They were between 19 and 38 years of age (an average age of 26), had pure-tone thresholds (.25–4 kHz) better than 20 dB Hearing Level at each test frequency, and were native speakers of English. None had previously participated in listening experiments involving synthetic speech.

**Stimuli.** Two phonemic contrasts were chosen: a labial-alveolar place contrast (BAIT-DATE) and an alveolar-velar place contrast

(DATE-GATE). These contrasts were chosen because they are formed of meaningful words and, therefore, are more likely to induce listeners into a "speech mode." Also, their use provided continuity with preliminary investigations that were carried out with adults and children. All stimuli were prepared on a six-formant hardware, parallel-formant synthesizer with an 8-kHz bandwidth (West, 1976) with the use of the copy-synthesis method. Time, frequency, and intensity parameters that were obtained from the analysis of natural tokens produced by an adult female speaker were used in the synthesis of the endpoint stimuli. These syntheses were, therefore, fully synthetic but were modeled on a given natural token by a single speaker and, in this respect, differ substantially from most synthetic speech continua used for similar experiments that are much more stylized (e.g., Lisker & Abramson's, 1970, voicing continua), typically have a more restricted bandwidth, and, usually, an /a/ vowel environment. For each condition, an 8-token continuum was constructed in which the cues were varied in equal steps, all other parameters remaining constant. Spectrograms of the synthetic endpoint stimuli along the place continuum are shown in Figure 1.

The total duration of all stimuli was 360 msec with an initial burst transient of 10 msec. The transient was formed of four noise-excited formants, with F1 and F2 fixed at 495 and 1020 Hz, respectively. F3 and F4 varied throughout the stimulus continua. They were centered at 1000 and 1500 Hz in the labial exemplar [b], at 4000 and 5000 Hz in the alveolar exemplar [d], and at 2000 and 3300 Hz in the velar exemplar [g]. The second and third formant transitions into the vowel also varied. F2 and F3 frequencies at vowel onset were 1896 and 2908 Hz in [b], 2300 and 3150 Hz in [d], and 2600 and 3200 Hz in [g]. Target values of F2 (2300 Hz) and F3 (3250 Hz) were reached for all stimuli after a transition duration of 50 msec. For all stimuli, F1 varied from 502 Hz at the beginning of the transition to 655 Hz at the end of the initial transition period. Vowel formant frequencies varied over the remaining part of the utterance to provide a good match to the original natural vowel. Fundamental frequency varied from 222 Hz at the beginning of the vowel to 75 Hz at the end of the vowel, with the contour copied from the natural utterance.

The place contrasts were presented in three conditions: (1) a *full-cue* condition, in which both the initial burst center frequency and formant transitions were varied along a continuum; (2) a *formant-transitions-cue* condition, in which the burst was fixed at an intermediate position (2000–2700 Hz for the BAIT-DATE contrast and 2830–4060 Hz in the DATE-GATE contrast) and the contrast was cued solely by a change in F2/F3 transitions; and (3) a *burst-cue* condition, in which the F2 and F3 transitions were fixed at an intermediate position (2090 and 3026 Hz for the BAIT-DATE contrast, 2445 and 3175 Hz for the DATE-GATE contrast) and the contrast was cued by changes in burst frequency. In constructing reduced-cue conditions, one has the choice of removing a cue (e.g., burst deletion) or neutralizing it, for example, by fixing a burst center frequency at an intermediary value to those appropriate to the endpoints of the continuum. Both techniques have been used in cue-weighting experiments, although cue deletion seems the more frequent choice. Each technique has some advantages. It could be argued that cue deletion, where possible, produces a less ambiguous token than does cue neutralization. It is methodologically simpler, as there is no need to decide what is to be considered a neutral position. However, it can substantially alter the overall complexity of the stimulus, and this may, in turn, affect labeling strategies used by the listener. Additionally, there are problems in deleting formant transitions in this vowel environment, as flat F2/F3 onsets would produce a strong bias toward DATE responses. In this study, therefore, in reduced-cue conditions, a neutralization, rather than a deletion, of the complementary cue was performed.

**Procedure.** The speech synthesizer was linked to a microcomputer, which controlled stimulus presentation, data collection, and data processing. Stimuli were presented on-line in a two-alternative,

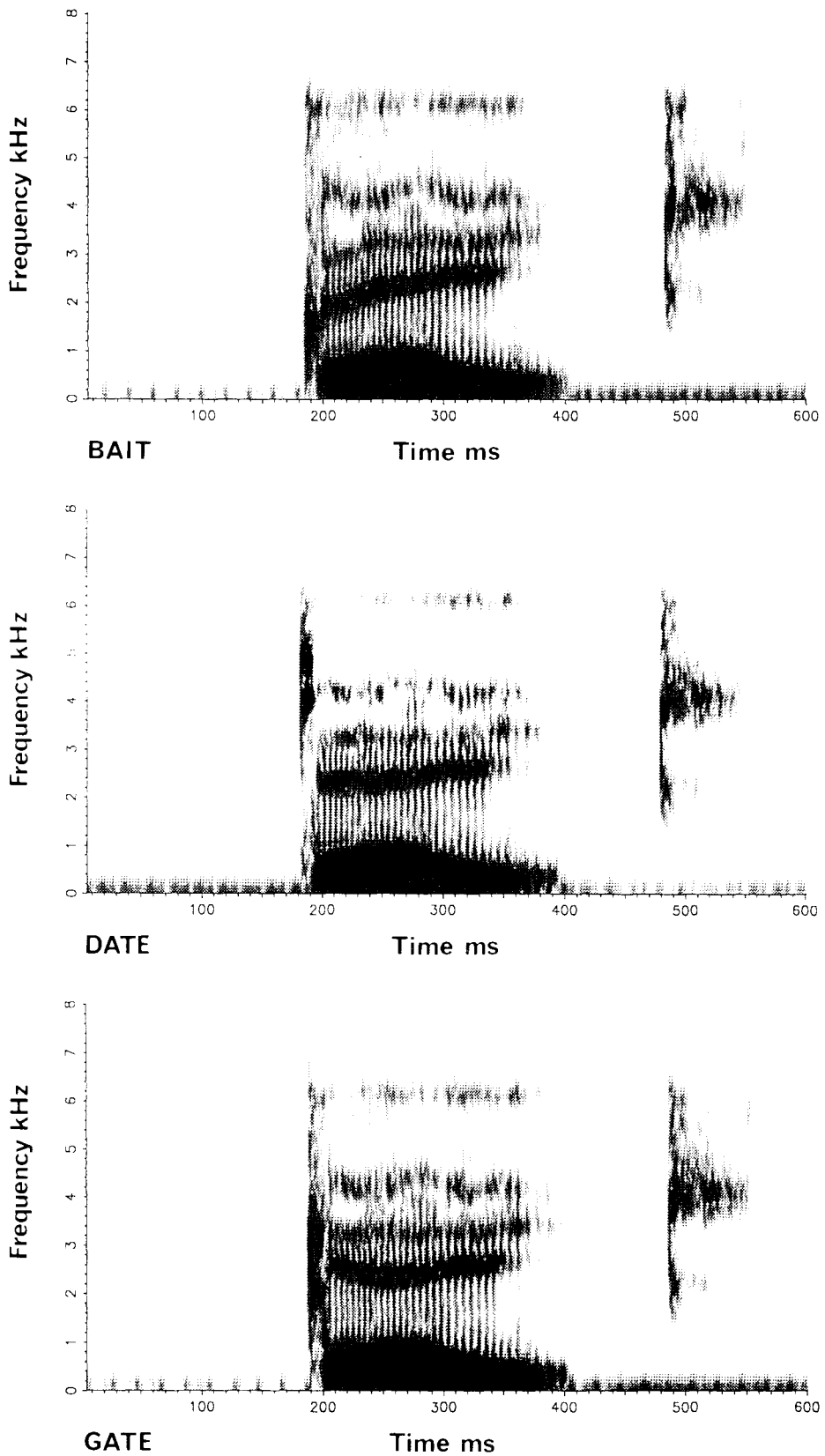


Figure 1. Spectrograms of the endpoints of the synthesized stimulus continua.

forced-choice identification experiment (BAIT-DATE or DATE-GATE). Each of the six test conditions was presented in a separate test. Stimuli were presented at a comfortable level, constant for all listeners, through Sennheiser HD414 headphones, which were linked directly to the synthesizer. The subjects heard three sets of appropriate exemplars at the beginning of each test. Then, 10 tokens of each stimulus were presented, randomized by permuting the order of the eight stimuli in 10 blocks. Each test, therefore, consisted of 80 presentations. The subjects responded by pressing one of two frames on a touch-sensitive response box. The test was self-paced: the recording of a response triggered the presentation of the following stimulus.

At the first testing session, a pure-tone audiogram was obtained for each subject and the full-cue conditions were presented. The subjects were retained for further sessions if they reached a criterion of 90% correct on the endpoints of the range for both contrasts. This selection was done to exclude listeners who were responding unreliably to full-cue stimuli. These subjects would have been of interest in themselves, but their inclusion in group results would have only confused the issue here, as the aim of this study was to measure variability in a group that would have been considered homogeneous in other studies. Two out of 18 subjects failed to reach this criterion. At each of the following four sessions, the six conditions were presented in random order. Different orders of test presentation were used for each subject at each session.

## Results

**Group data.** The mean identification functions, averaged over all sessions and subjects, are presented in Figure 2. The identification functions obtained for the two full-cue conditions are sharp, with well-established plateaus of consistent labeling at the extremes of the continuum and steep gradients around the phoneme boundary. Functions obtained for the single-cue conditions are distinctly shallower so that at least one endpoint of each continuum is not consistently labeled. Functions obtained for both reduced-cue conditions of the place contrasts did not differ significantly from each other with respect to their gradient and phoneme boundaries.

**Variability across subjects for full- and reduced-cue conditions.** These averages, however, conceal the extent of variability across subjects, especially for reduced-cue conditions. Figures 3 and 4 show the individual categorization functions obtained for all test continua. Note the much greater spread of performance in the single-cue case. In the BAIT-DATE tests, for example, the /b/ and /d/ endpoints were labeled correctly between 90%–100% and 75%–100% of the time, respectively, in the full-cue condition but between 55%–100% and 17.5%–100% of the time in the burst-only continuum. Although this is the most extreme example, intersubject variability was always greater in reduced-cue conditions than it was for the full-cue continua.

**Individual variability in the effects of cue reduction.** Because subject performance was relatively uniform in the full-cue conditions, and varied widely in the reduced-cue conditions, it follows that the subjects varied in the extent to which their performance was affected by cue reduction. A statistical modeling approach based on *generalized linear models* (GLMs) was used to quantify this. Analyses were done individually for each subject, comparing identification performance for a particular reduced-

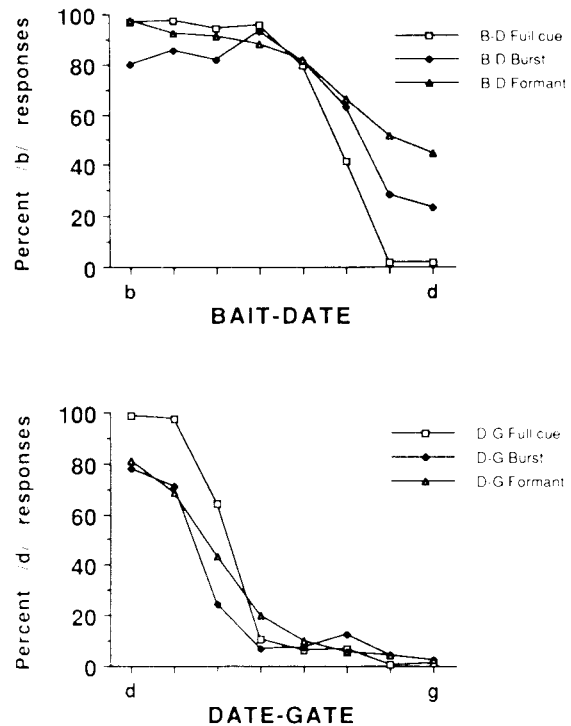


Figure 2. Mean identification functions for the BAIT-DATE and DATE-GATE contrasts (Experiment 1).

cue condition (e.g., BAIT-DATE, with the burst cue alone) with performance in the appropriate two-cue condition (e.g., the two-cue BAIT-DATE continuum). The initial, so-called saturated (or maximal), model assumed that the observed data (the proportion of times a particular stimulus was given a particular label) was a function of stimulus number (1–8), session number (1–4), and condition (full vs. reduced cue). Stimulus number was treated as a continuous variate, whereas session and condition were treated as categorical factors, with the appropriate number of levels. This technique is analogous to probit analysis (in which a sigmoid curve is fitted to a categorization function; Finney, 1952), but, in this instance, all eight categorization functions (4 sessions  $\times$  2 conditions) were fitted at once. As 2 parameters were fitted to each categorization function (a slope and an intercept), 16 parameters needed to be estimated. Also obtained was the “deviance,” a measure of how reasonable the model was for that set of data and estimated parameters.

Following these analyses, another model was fitted in which identical performance across the full- and reduced-cue conditions was assumed. The difference between the deviance obtained for this model (of fewer parameters) and for that of the maximal model can be shown to be asymptotically distributed as chi-square, with the number of degrees of freedom (*df*) given by the difference in the number of estimated parameters (McCullagh & Nelder, 1983). Because only eight parameters were needed for fitting a model under the assumption of no con-

dition effect (i.e., four sessions, each requiring a slope and an intercept), the relevant statistic has  $df = 8$ .

To summarize, the change in deviance for models that do and do not include the effect of condition (hereinafter referred to as the *condition* deviance) across these two models supplies a quantitative, statistically interpretable measure of the extent to which subject performance changed with cue reduction. For further details of the analysis, see the Appendix.

Table 1 gives the condition deviances calculated for each subject and for each reduced-cue condition, along with means and medians by condition. Because of the large number of statistical tests to be performed (64 in this case), a relatively stringent significance criterion ( $p < .001$ ) was chosen to reduce Type I errors. Thus, condition deviances greater than 26.125 indicate instances where the reduction of cues caused a change in performance. As Table 1 indicates, only relatively rarely did a cue reduction appear not to change performance (4 out

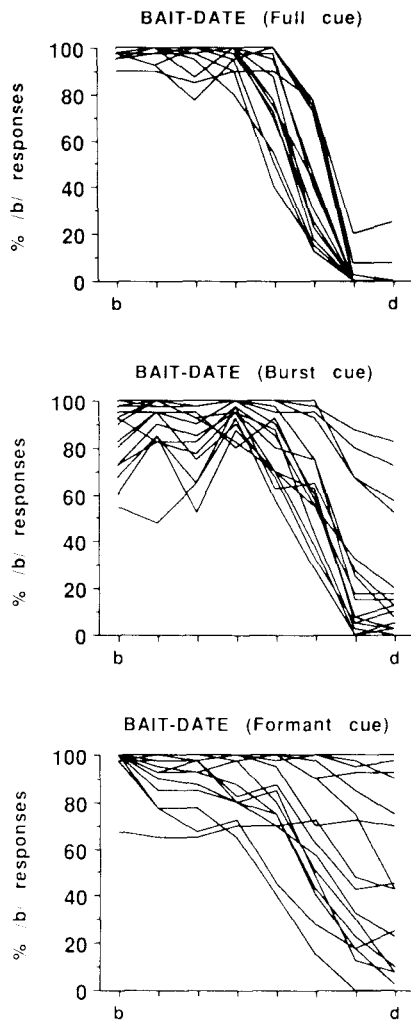


Figure 3. Individual identification functions for the full-cue and reduced-cue conditions of the BAIT-DATE contrast (Experiment 1).

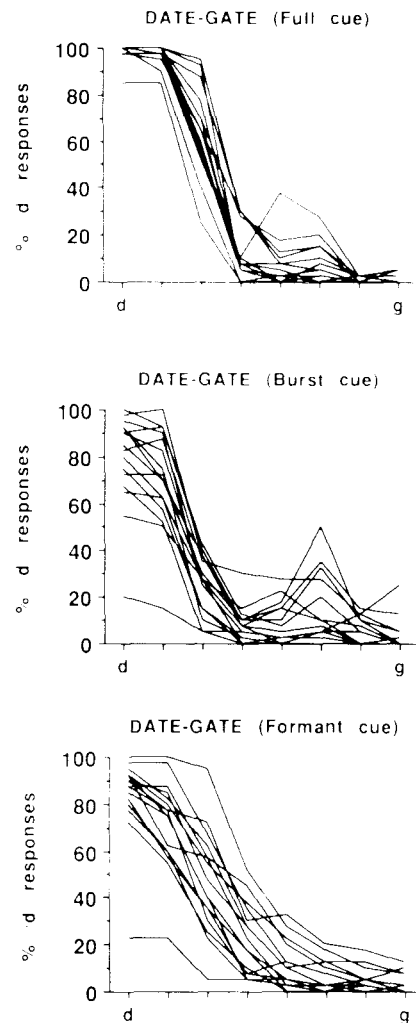


Figure 4. Individual identification functions for the full-cue and reduced-cue conditions of the DATE-GATE contrast (Experiment 1).

of 64 cases, or 6.25%). Even where cue reduction did significantly affect performance, there was a wide range of obtained condition deviances, showing that the magnitude of the condition effect varied quite markedly among subjects. Also, whereas some subjects showed roughly equivalent deviances for both reduced-cue conditions of the place contrasts, a smaller number showed a far greater effect for one condition than for the other. The condition (burst- or formant-cue) showing the most effect is not consistent across subjects.

A better feel for the results can be obtained by examining individual results across the range of condition deviances that were obtained. For each test condition, changes in deviance were rank ordered to select those situations in which cue reduction had its greatest, median, and least effects. One labeling function was chosen at each level (i.e., the largest condition deviance, the smallest, and the ninth largest). The categorization curves (averaged across sessions) for the reduced-cue continuum and its associated

Table 1  
Condition Deviances

Subject	BAIT-DATE		DATE-GATE	
	Burst	Formant	Burst	Formant
1	37.58	41.66	50.61	63.23
2	131.04	84.10	91.96	54.44
3	98.56	124.20	69.83	22.52
4	73.32	217.85	42.41	46.56
5	55.43	68.71	25.09	46.73
6	30.20	19.74	42.31	28.81
7	144.65	178.47	117.49	71.92
8	177.13	171.40	149.93	97.93
9	108.08	142.97	43.70	67.25
10	96.94	161.60	109.51	104.80
11	124.35	119.86	101.57	105.29
12	125.00	40.10	59.93	40.23
13	56.55	244.77	111.69	55.98
14	63.29	112.42	12.39	26.60
15	141.24	236.01	76.65	65.11
16	113.67	56.53	57.67	26.18
Mean	98.56	126.27	72.67	57.72
Median	103.32	122.03	64.88	55.21

Note—Condition deviances are measures of the degree to which subjects change their labeling behavior from full-cue to reduced-cue conditions. These may be compared to a chi-square distribution with  $df = 8$ . Values of 26.125 or greater are significant at the .001 level.

full-cue continuum corresponding to those selected condition deviances are shown in Figure 5. Note that the strongest effects of cue reduction are associated with the BAIT-DATE contrast. The largest condition deviances are usually obtained when the subject is unable to differentially label the two endpoints of the reduced-cue continuum but sharply labels the full-cue continuum (see, e.g., the formant condition results for BAIT-DATE). Median condition deviances (middle column) can be associated with reasonably dichotomous labeling of the two extremes of the reduced-cue continua, but with categorization curves that are distinctly shallower than those that are obtained for the full-cue continua (see, e.g., both DATE-GATE conditions). Finally, the smallest condition deviances are associated, as expected, with categorization curves that are very similar across full- and reduced-cue continua.

**Measures of categorization: Function gradient and phoneme boundary.** An identification function can usefully be characterized by two measures: its gradient and its boundary. A change in deviance between two conditions will usually reflect changes in one or both of these measures. As is apparent in Figure 5, cue reduction typically led to shallower labeling functions. To quantify this further, a single sigmoid curve was fitted to the results from each condition, without taking session into account (again using a GLM). In a sense, this is like fitting a curve to the average of the four sessions. Labeling function gradients were shallower for the reduced-cue continua than they were for the associated full-cue continua in 61 of 64 possible comparisons (16 subjects  $\times$  4 reduced-cue continua). The 3 other comparisons had proportional increases in slope that were the smallest in magnitude for that particular condition.

It is, perhaps, worth emphasizing what is also evident from Figure 5: the proportional change in labeling function gradient with cue reduction varied widely across subjects. For example, comparing the full-cue BAIT-DATE continuum with its burst-cue counterpart, slopes decreased by 2.5% to about 86% (with a mean of 58.9%), depending upon the subject.

Differences in performance between test conditions can also be marked by phoneme boundary shifts. Median boundary (on a scale from 1 to 8) for the BAIT-DATE contrast shifted from 5.41 for the full-cue condition to 5.84 for the burst-cue condition and to 6.63 for the formant-cue condition. For the DATE-GATE contrast, the median boundary shifted from 3.21 to 2.43 and 2.65 for the burst and transitions conditions, respectively. In the reduced-cue conditions, for a number of subjects, the estimated boundary fell outside the stimulus continuum, showing that no contrast had been clearly established between the endpoints. This was the case for 13 out of 64 functions (or 20% of the cases). This extreme effect of cue reduction was found 25% of the time for the burst condition and 43.75% of the time for the formant condition of the BAIT-DATE contrast. For the DATE-GATE contrast, this extreme effect was only found for 6.25% of the cases for both the burst and formant conditions.

**Relationships among different continua.** First, the relationship between the two place contrasts was examined by looking at the median values obtained for the gradient and boundary measures, and at the amount of variability found in the two test continua. The full-cue continuum of the DATE-GATE contrast was more sharply labeled (with a gradient of  $-1.465$ ) than that of the BAIT-DATE contrast ( $-1.031$ ). An examination of the graphs and condition deviances for the reduced-cue conditions of these two contrasts generally reveals much greater variability in the labeling of the BAIT-DATE contrast. Ranges of values obtained for the condition deviances were larger for the two reduced-cue (burst and formant) conditions of the BAIT-DATE contrast than they were for the matching conditions of the DATE-GATE contrast.

Interestingly, even though generally greater effects of cue reduction were found for the BAIT-DATE contrast, on average, the neutralization of the burst or formant cue had very similar effects on the function gradient within each of these contrasts. For the BAIT-DATE contrast, the median gradient was reduced by 69.5% and 67.8% for the burst- and formant-cue conditions, respectively. In the DATE-GATE contrast, the reductions in gradient were 46.4% and 44.7%, respectively.

The relationship among different continua can also be examined across subjects. Given that subjects vary so markedly in the effects of cue reduction, we may ask if there are any relationships in sensitivity to cue reduction across the different continua. There are two main interesting ways this could happen: (1) Listeners could differ in their sensitivity to particular acoustic features; for example, some listeners may place more reliance on the release

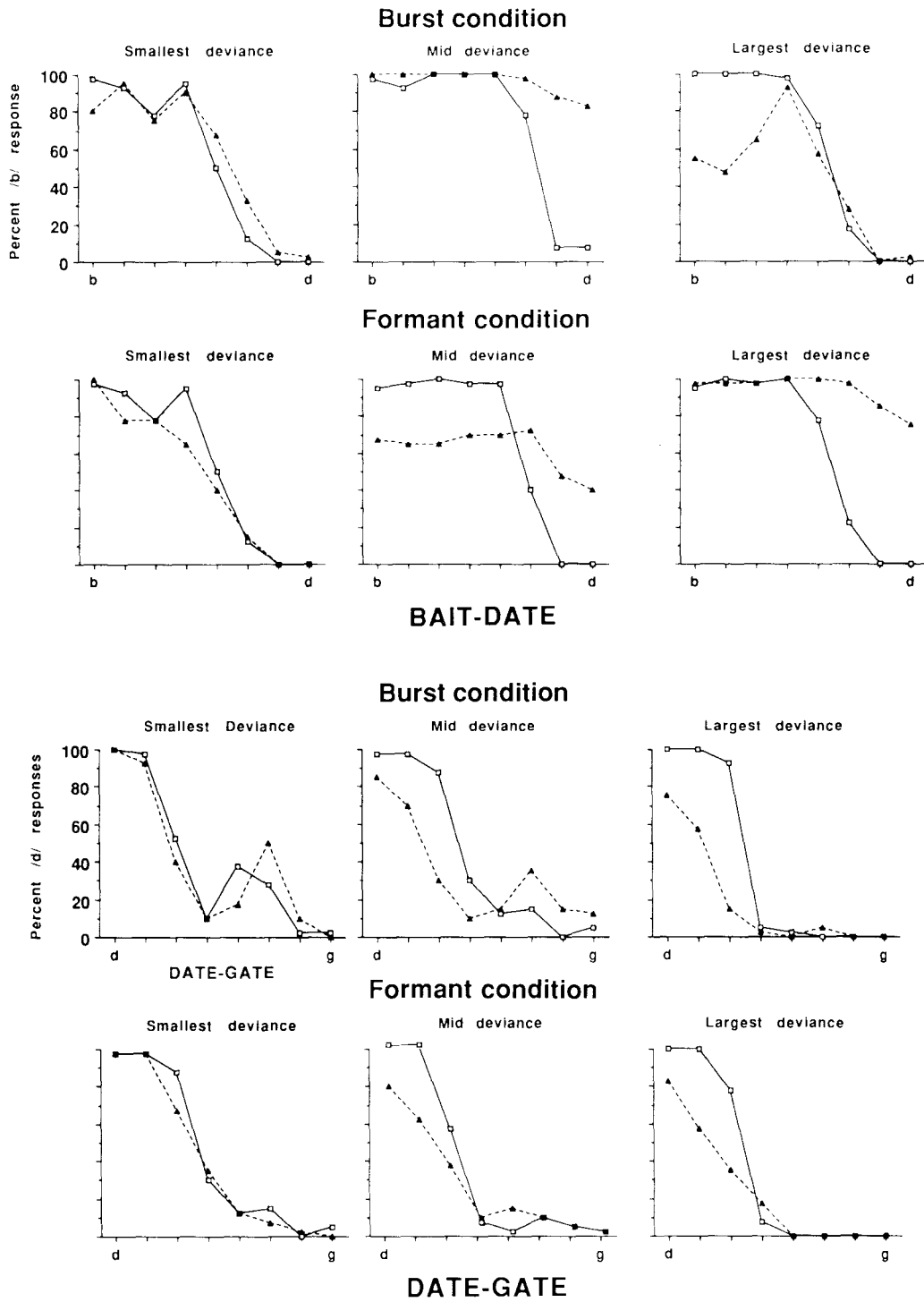


Figure 5. Individual identification functions showing the smallest, median, and largest deviation for each of the test conditions (Experiment 1). The full-cue function (solid line) is given as a reference.

burst in plosives. In this case, we would expect performance on the two formant-transitions-cue place continua to be more severely degraded than on the two burst-cue continua; and (2) listeners could differ in a general ability to use reduced auditory information in labeling tasks. Here, we would expect a listener to show a similar sensitivity to cue reduction of all kinds.

The relationship among changes in performance across speech continua was investigated by calculating the linear correlation coefficient between all possible pairs of the condition deviances shown in Table 1. As Table 2 shows, there is little evidence that subjects are sensitive to particular acoustic features. In fact, the largest correlation is between the effect of formant transitions and bursts on the DATE-GATE continuum. Better evidence is found for the supposition that subjects have a general ability to make use of reduced auditory information. Although only two of the eight correlations are significant at the .05 level, two others approach this level, and all are positive.

### Discussion

Two aspects of the results will be discussed. First, the group results obtained in this study will be compared with the findings of other studies of the perception of the place contrast in initial stops. Second, hypotheses will be made as to the likely causes of the sizable amount of individual variability found in this study.

Generally, greater effects of cue reduction were found for the BAIT-DATE contrast than were found for the DATE-GATE contrast, and, on average, cue-reduction effects differed little across bursts and formant transitions. This appears to contradict earlier findings that listeners showed little effect of the removal of burst information when labeling place contrasts (Raz & Noffsinger, 1985; Walley & Carrell, 1983). It also contradicts the findings of Blumstein and Stevens (1980), who considered the static spectrum onset cue as salient and formant transitions as more secondary cues, mainly useful in the absence of burst information. Several factors may be contributing to this apparent contradiction. First, the presence of the burst at a neutral position may be giving more contradictory information than does the mere removal of the burst. However, in their study, Walley and Carrell found that even when contradictory burst information was present, formant information dominated. Second, differ-

ent stimuli were used here, with consonants in an /ei/ vowel environment, rather than in the more typically used /a/ environment. Given that cue weighting for place of articulation varies with vowel environment (Dorman, Studdert-Kennedy, & Raphael, 1977), transition information may be less dominant in this context.

In the reduced-cue conditions, listeners generally had more difficulty labeling the dental stimuli in both contrasts. In their perceptual experiments, which used edited natural speech, Dorman et al. (1977) also found more variability for reduced-cue conditions in the labeling of dental stimuli than they did for labials or velars, even though the experimental conditions and stimuli were quite different from those used here. However, difficulty in labeling the DATE stimuli in this study could also be related to the test continua that were used. Results of the two-cue test conditions show that the continua were not perceptually balanced, with a predominance of BAIT responses for the BAIT-DATE continuum (i.e., a boundary at a point greater than the midrange point) and of GATE responses for the DATE-GATE continuum (i.e., a boundary at a point lower than the midrange point). As the burst frequencies and formant transitions were neutralized at frequency values at the midrange point rather than at the true boundary, listeners who were influenced by the neutralized cue would tend to give a greater proportion of BAIT responses for the BAIT-DATE continuum, and GATE responses for the DATE-GATE continuum. These are indeed the results obtained from the group data.

When individual results were examined, it was found that the magnitude of the cue-reduction effect varied greatly between listeners. Most listeners showed a reduction in categorization-function gradient in reduced-cue conditions. However, some showed quite drastic effects of cue reduction in some test conditions, with no clear contrast established between the endpoints, whereas a small number of listeners showed no significant change in performance when certain speech pattern cues were removed. An examination of individual results also revealed that whereas some subjects were similarly affected by the removal of both cues, a small number were more greatly affected either by the removal of the burst cue or by the removal of the formant cue.

As far as can be ascertained from the sparse data on individual variability presented in other studies, it appears that the individual variability found in this study was more extensive than is typically found. The cause of this variability may be in the stimuli used or in the particular group of subjects tested. The stimuli used here differ from the great majority of test continua in a number of ways. First, the syntheses were based on female voice. Second, the stimuli were closely copied on natural tokens and, therefore, are more complex than are the highly stylized stimuli used in most laboratories. Third, a diphthong was used, rather than a monophthong. An investigation of the effect of each of these factors is necessary and will be the scope of thorough investigation. It must first be established

Table 2  
Correlations Between the Condition Deviances of Table 1

	BDBU	BDFF	DGBU
BDFF	0.3085		
DGBU	0.6458*	0.4610	
DGFF	0.4584	0.3855	0.6968*

Note—These correlations measure the extent to which a subject was similarly affected by cue reduction in the two cases. Values of .497 or greater (with \*) are significant at the .05 level, while those at .623 or greater are significant at the .01 level. BDBU = BAIT-DATE (burst cue); BDFF = BAIT-DATE (formant cue); DGBU = DATE-GATE (burst cue); DGFF = DATE-GATE (formant cue).



that the increased variability obtained here is not a function of the subject group instead of a difference in stimulus characteristics. It was therefore decided to compare, in a single group of listeners, the labeling of the BAIT-DATE contrast, for which most variability was found, with that of stimulus continua that would be fairly typical of those used in cue-weighting experiments and for which results and measures of subject variability had been published.

**EXPERIMENT 2**

**Method**

**Stimuli.** A new BA-DA continuum, based on the parameters listed by Raz and Noffsinger (1985), was constructed (Figure 6). In their study, Raz and Noffsinger obtained very uniform results with normal-hearing adults; performance was not affected by the removal of the burst information. In the original study, a 12-step BA-DA-GA continuum was used. Here, Steps 1 to 8 were replicated. The initial burst was composed of two formants: F2, which varied

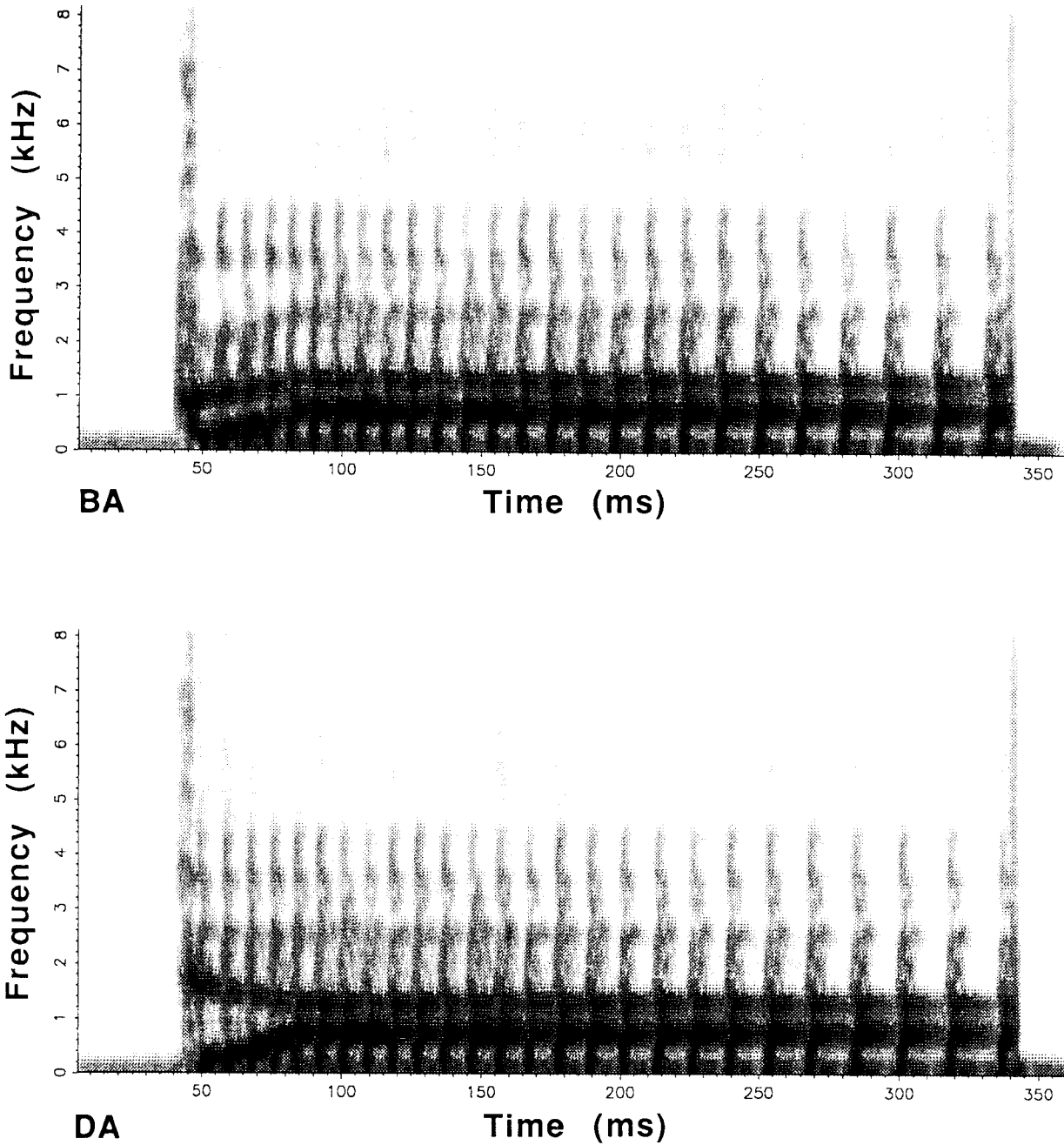


Figure 6. Spectrograms of the endpoints of the synthesized BA-DA stimulus continua.

in frequency, and F4, which was fixed at 3600 Hz. Amplitudes of these two formants varied throughout the continuum to produce a shift in spectral balance. At vowel onset, F1 was set at 200 Hz, F4 at 3600 Hz, F5 at 4500 Hz, and F2/F3 onset frequencies varied throughout the continuum: F2 was set at 942 Hz and F3 at 1903 Hz at the /b/ endpoint; for the /d/ endpoint, F2 was set at 1691 Hz and F3 at 2613 Hz. Transitions lasted over 40 msec, after which formants reached the following steady-state values: F1 at 720 Hz, F2 at 1240 Hz, F3 at 2500 Hz, F4 at 3600 Hz, and F5 at 4500 Hz. Fundamental frequency increased from 103 to 125 Hz during the 40-msec transition period, then decreased linearly to 50 Hz over the next 260 msec. As in the original study, two conditions were prepared: with and without initial burst. Although the stimuli were prepared here on a parallel, rather than a cascade, synthesizer, it was thought that they would not be significantly different from those used in the original study.

Note that the two sets of continua differ in the way that the cue reduction was effected: In the BAIT-DATE continuum, the burst cue was neutralized, whereas it was deleted in the BA-DA continuum. It is, therefore, important to ascertain that cue neutralization, by introducing ambiguity to the stimuli, is not leading to greater listener variability than is cue deletion. Seven listeners who participated in Experiment 1 were recalled and tested on the endpoints of five BAIT-DATE continua. The first three were those used in Experiment 1. In a fourth, the initial burst was deleted; in the fifth, F2/F3 transitions were "deleted" by setting onset frequencies equal to those at the end of the transition period in the full-cue version. Ten tokens of each of these 10 endpoints were randomized and presented to listeners for identification at two consecutive sessions. The results for the burst-cue conditions will be discussed here, as they are relevant to Experiment 2. In both the burst-neutralized and burst-deleted conditions, very little variability was found in the labeling of the /b/ endpoints. All were labeled 100% correctly by all listeners, except for one who obtained a score of 95% for the burst-neutralized endpoint. Greater variability was obtained for the /d/ endpoint. Here, scores of 55%-90% correct labeling were obtained for the burst-neutralized /d/ endpoint, as opposed to scores of 10%-95% correct labeling for the burst-deleted endpoint. Therefore, it appears that cue neutralization does not necessarily lead to greater variability than does cue deletion.

**Subjects.** Nine new subjects were tested during five 45-min sessions and were paid for their participation; they were between 18 and 20 years of age, had pure-tone thresholds better than 15 dB hearing level at each test frequency between .25 and 4 kHz, and were native speakers of English.

**Procedure.** The subjects were tested over five sessions. At the first session, a pure-tone audiogram was obtained, and the subjects were tested on the two full-cue continua (BAIT-DATE, BA-DA). As in the first experiment, subjects who failed to reach a criterion of 90% correct labeling on all endpoints of the full-cue continua at this initial testing session were not included in the final results. Three subjects were thus eliminated from the group results—interestingly, all failed to label the /d/ endpoint of the BA-DA contrast appropriately, although they obtained 100% correct labeling on the endpoints of the BAIT-DATE contrast. The data from these subjects is discussed below. At the following four sessions, the subjects were tested on all continua. The test procedure was identical to the one used in Experiment 1.

## Results

**Group results.** The mean identification functions averaged over all sessions are presented in Figure 7. The mean functions obtained with the new subjects for the BAIT-DATE contrast are similar to those obtained in Experiment 1, except that better labeling was obtained at the /d/ endpoint for the burst-cue condition. Functions obtained

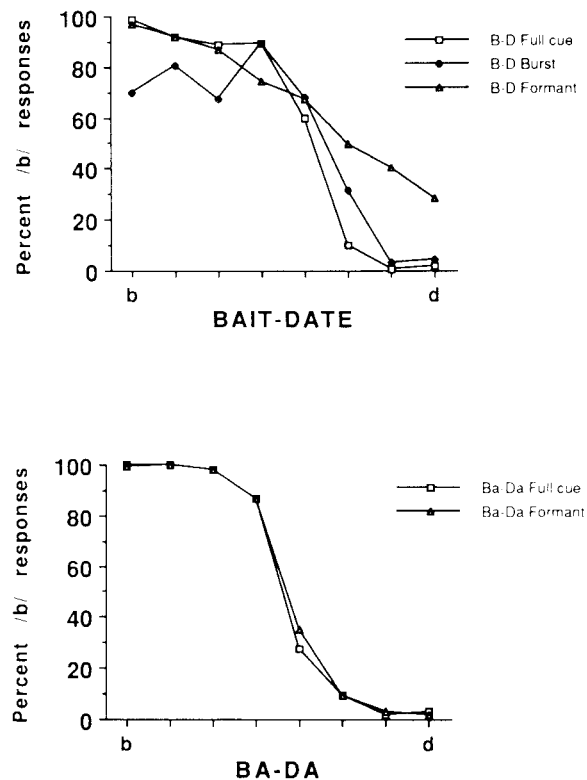


Figure 7. Mean identification functions for the BAIT-DATE and BA-DA contrasts (Experiment 2).

for the BA-DA stimulus continua are more monotonic than those obtained for the BAIT-DATE contrast. Virtually identical functions were obtained for the full-cue and formant-alone conditions, showing that there was, on average, little effect of burst deletion. The 7 subjects recalled from Experiment 1 were also tested on the two conditions of the BA-DA range. Mean labeling functions obtained for these subjects are similar to those obtained for the new group of 6 subjects.

Lower median gradients were obtained for the BAIT-DATE contrast (full cue,  $-0.822$ ) than for the BA-DA contrast (full cue,  $-1.800$ ). Shallower labeling was obtained for reduced-cue conditions in all cases, but the extent of the difference in gradient was greater for the BAIT-DATE than it was for the BA-DA contrast. Shifts in median phoneme boundary in reduced-cue conditions were found for the BAIT-DATE contrast (full cue, 4.65; burst cue, 4.14; formant cue, 6.11) but not for the BA-DA contrast (full cue, 4.41; formant cue, 4.49).

**Individual results.** Condition deviances were calculated for all subjects (Table 3). Greater median deviance was found for the BAIT-DATE contrast than was found for the BA-DA contrast. For the BAIT-DATE contrast, all the new subjects showed significant deviances in both test conditions. For the BA-DA continuum, only 2 subjects showed a significant ( $p < .001$ ) effect of the removal of burst information. These 2 subjects showed the second and fourth

Table 3  
Condition Deviances

Subject	BAIT-DATE		BA-DA
	Burst	Formant	Formant
1	131.24	185.25	22.23
2	61.39	42.50	6.79
3	55.84	149.54	26.65
4	127.59	62.19	32.14
5	45.26	48.84	13.60
6	51.92	71.49	12.98
mean	78.87	93.30	19.07
median	58.61	66.84	17.92

Note—Condition deviances are measures of the degree to which subjects change their labeling behavior from full-cue to reduced-cue conditions. These may be compared to a chi-square distribution with  $df = 8$ . Values of 26.125 or greater are significant at the .001 level.

largest deviances for the formant-transition condition of the BAIT-DATE contrast. Most functions for the BA-DA contrast show evidence of sharp labeling, even in the reduced-cue conditions. This is reflected in the generally smaller deviance values obtained. The greatest deviance for this contrast was 32.14, compared with 131.2 for the burst condition and 185.25 for the formant condition of the BAIT-DATE contrast.

An examination of individual functions obtained for the BA-DA contrast (Figure 8) revealed that labeling of the endpoints was highly consistent for both the full- and formant-cue conditions. The interquartile ranges of the phoneme boundaries were much smaller for the BA-DA contrast (full cue, 0.59; formant cue, 0.73) than they were for the BAIT-DATE contrast (full cue, 0.87; burst cue, 1.33; formant cue, 4.79).

**Results for listeners not reaching testing criterion.** Three subjects were not included in the group results, because they had failed to score 90% correct on all full-cue endpoints in the initial screening session. This criterion was set in order to focus particularly on variability in the labeling of reduced-cue continua for the subjects who were labeling full-cue continua in a categorical fashion. It is still of interest to examine results for these subjects, as they only had difficulty in labeling one specific endpoint, namely the /d/ endpoint of the BA-DA contrast, and were highly consistent in their labeling of the BAIT-DATE full-cue continuum. This result is unexpected as little subject variability had been found for the labeling of the BA-DA continuum in the Raz and Noffsinger study. Group labeling functions for the 3 'failed' subjects (Figure 9) show that results obtained for the BAIT-DATE contrast were similar to group results obtained by the old and new subjects, but that shallower labeling was obtained for the BA-DA contrast. Also, there was a marked difference between the labeling of the two test conditions, which show an effect of burst-cue removal for these subjects. Variability in labeling behavior was, therefore, found for stimulus continua for which highly consistent labeling has been claimed and, indeed, been demonstrated for most listeners.

## Discussion

An important aim of Experiment 2 was to check whether the extensive variability obtained in the first experiment was caused by the subject group that was used or was a factor of particular characteristics of the stimulus continua that were used. When more stylized stimuli are used, with an /a/ vowel environment, a different outcome is seen. Here, group results do not show any effect of cue reduction (burst removal), and the extent of between-subject variability is reduced. It could be hypothesized that the effect of burst removal in the BA-DA stimuli might be minimal because of the abrupt onset (short rise time) of the vowel in the reduced-cue stimuli, with formants reaching their full intensity over a short time period.

The BAIT-DATE and BA-DA stimulus continua differ in a number of ways. Formants in the BAIT-DATE contrast varied throughout the duration of the vowel to provide a good match to a natural utterance, whereas the formants in BA-DA remained constant after the initial transition period. Transitions will, therefore, be more salient in the BA-DA stimuli. Furthermore, if onset frequencies are examined relative to the 'steady-state' portions, it appears that there are much clearer pattern differences in the BA-DA continua, as rising F2 and F3 transitions in the BA endpoint are contrasted with falling F2 and F3 transitions in the DA endpoint. In BAIT-DATE, one finds steeply rising F2 and F3 transitions in BAIT, as opposed

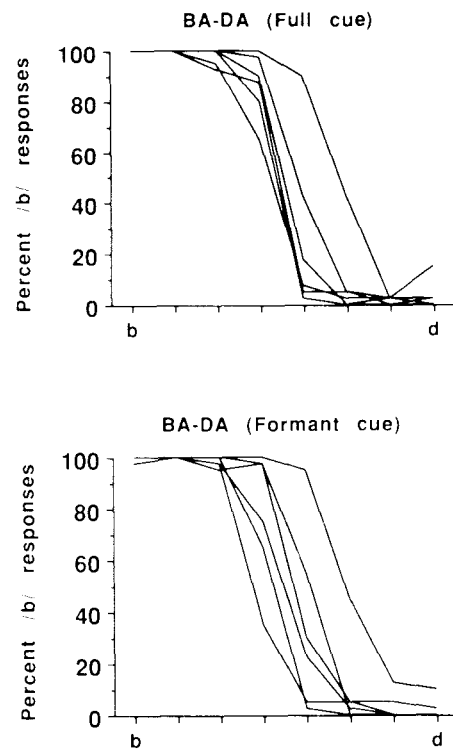


Figure 8. Individual identification functions for the full-cue and formant-cue conditions of the BA-DA contrast (Experiment 2).

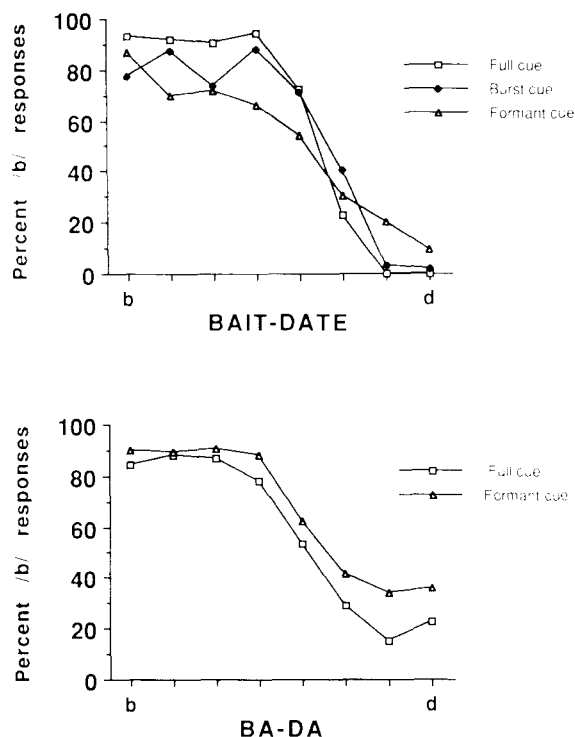


Figure 9. Mean identification functions for the full-cue and formant-cue conditions of the BA-DA contrast for the three listeners failing the endpoint labeling criterion.

to a flat F2 and slightly rising F3 transition in DATE. Therefore, both the vowel environment and the inherent degree of acoustic complexity of the syntheses contribute in making the /b/-/d/ contrast more salient in the BA-DA continua than it is in the BAIT-DATE continua. One hypothesis might be, therefore, that the amount of variability in labeling could be related to cue salience, which would be determined by vowel environment and overall stimulus complexity. When cues are highlighted and speech pattern elements clearly differentiated, less variability in their use by individual listeners will be observed. Further investigations into the effect of stimulus complexity are under way.

### GENERAL DISCUSSION

A first question to be addressed is whether variability in speech-pattern use found in this and other cue-weighting experiments reflects an important aspect of speech-pattern perception or whether it is merely an artifact of the experimental techniques that were used. In some studies, in which both computer-edited natural stimuli and copy syntheses have been used, greater variability has been found in categorizing synthetic stimuli (Hazan, Holden-Pitt, Revoile, & Edward, 1987), even though individual variability in the effect of cue removal was found for both types of stimuli. However, results of early experiments involving highly stylized stimuli are more equivocal.

Some of the continua produced at Haskins Laboratories, such as the voicing continua of Lisker and Abramson (1970) and the BA-DA-GA place continuum, lack many of the secondary cues present in speech but are nevertheless quite consistently categorized by large numbers of subjects. It may be hypothesized that it is their very simplicity, marked especially by steady-state formants and highly focused burst transients, that makes the cues that are present more salient. Even in reduced-cue conditions, the perceptual salience of the remaining cue may make it highly intelligible to most listeners. In copy syntheses, which are more closely matched to natural speech, cues will be less salient because of the presence of a more complex set of patterns. When cue redundancy is further reduced, greater individual differences in categorization may become more apparent. Another point that must not be overlooked is that the greater the complexity of the synthesis, the greater the possibility that some frequency or intensity parameters may be inappropriately set. The resulting anomalies in the synthesized stimuli may affect some listeners more than others. Overall, however, it could be argued that results obtained with copy syntheses, because of the similarity in the level of pattern complexity to natural speech, would be more representative of natural speech processing than those that are obtained with highly simplified synthetic stimuli.

Another question to be asked is whether the removal of cues to a contrast mirrors any situation found with natural speech or whether cue removal or neutralization is a highly artificial procedure. The necessity of using one cue in isolation, in the absence of possibly more prominent cues to a contrast, is not restricted to the laboratory. Indeed, it is often encountered in listening to speech in noise, where spectral cues can often be masked (e.g., Wardrip-Fruin, 1985). Listeners with normal hearing thresholds vary significantly in their ability to understand speech in noise (Simon & Fourcin, 1976). Some of these differences are due to the varying ability in making use of contextual information. When, however, the context is controlled, as in tests using word lists that are provided to the subjects (Erber, 1969), individual differences in intelligibility are still found. It seems possible that these experiments also reflect how, at a low level of processing, listeners may differ in the way in which they make use of speech-pattern information.

Even in clear speech, some flexibility in the use of cues to a contrast may be necessary because of speaker differences in speech production. In their study of cues to place contrasts in different vocalic contexts, Dorman et al. (1977) found that the perceptual weighting of burst and formant cues varied, not only according to vowel context but also according to speaker. Individual listeners who show less effect of cue reduction may be at an advantage in adapting to different speakers.

The existence of individual variability in speech-pattern perception has implications for past and future cue-weighting studies. The sole presentation of group results gives a false impression of universality in speech-pattern-

cue use. Some measure of individual variability (such as the identification index proposed by Van Tasell et al., 1982, in their cue-weighting studies) is essential. Many factors make cross-comparisons of speech perceptual studies on a specific speech contrast difficult: for example, differences in types of stimuli (synthetic or edited natural stimuli), in vowel environment, in details of intensity relations (e.g., burst vs. vowel), in test methods, and in measures that are used in analyzing results. However, the existence of individual variability in the use of speech-pattern cues may also be an important factor and one that may go some way toward explaining some seemingly contradictory results obtained in existing cue-weighting experiments, both with normal-hearing and hearing-impaired listeners.

Individuals' varying ability to make use of minimal speech-pattern information may also have implications for the intelligibility of synthetic speech, in which redundancy of speech-pattern information is typically much reduced. It may well be that some listeners have greater difficulty in perceiving synthetic speech than do others. Although, at word- and sentence-level testing, individual variability in cue processing may be masked by contextual information effects, it may come to the fore in more analytic tests; a quantification of listener variability is essential to the development of standardized assessment methods for synthetic speech.

#### REFERENCES

- AITKIN, M., ANDERSON, D., FRANCIS, B., & HINDE, J. (1989). *Statistical modelling in GLIM*. Oxford: Clarendon Press.
- BLUMSTEIN, S. E., & STEVENS, K. N. (1980). Preceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, **67**, 648-662.
- BOCK, K. D., & JONES, L. V. (1968). *The measurement and prediction of judgment and choice*. San Francisco: Holden Day.
- COOPER, F. S., DELATTRE, P., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, **24**, 597-606.
- DOBSON, A. J. (1983). *Introduction to statistical modelling*. London: Chapman & Hall.
- DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, **22**, 109-122.
- EILERS, R., OLLER, D. K., URBANO, R., & MOROFF, D. (1989). Conflicting and cooperating cues: Perception of cues to final consonant voicing by infants and adults. *Journal of Speech & Hearing Research*, **32**, 307-316.
- ERBER, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech & Hearing Research*, **12**, 423-425.
- FINNEY, D. J. (1952). *Probit analysis*. Cambridge: Cambridge University Press.
- FITCH, H. L., HALWES, T., ERICKSON, D. M., & LIBERMAN, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, **27**, 343-350.
- HAGGARD, M., AMBLER, S., & CALLOW, M. (1969). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613-617.
- HAZAN, V., HOLDEN-PITT, L., REVOILE, S., & EDWARD, D. (1987). Perception of cues to a stop voicing contrast by normal-hearing children and adults. *Proceedings of XIth International Congress of Phonetic Sciences*, 358-361. Tallinn, Estonia: Academy of Sciences of Estonia.
- KEWLEY-PORT, D. (1981). *Representations of spectral change as cues to place of articulation in stop consonants*. Unpublished doctoral dissertation, Indiana University, Bloomington.
- LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L., & BASTIAN, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language & Speech*, **4**, 175-295.
- LIBERMAN, A. M., HARRIS, K. S., KINNEY, J. A., & LANE, H. (1961). The discrimination of relative onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, **6**, 379-388.
- LISKER, L., & ABRAMSON, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.
- MCCULLAGH, P., & NELDER, J. A. (1983). *Generalized linear models*. London: Chapman & Hall.
- MORRONGIELLO, B. A., ROBSON, R. C., BEST, C. T., & CLIFTON, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, **37**, 231-250.
- PAYNE, C. D. (1985). *The GLIM system release 3.77 manual - Edition 2*. Oxford, U.K.: Numerical Algorithms Group Ltd.
- RAZ, I., & NOFFSINGER, D. (1985). Identification of synthetic, voiced stop-consonants by hearing-impaired listeners. *Audiology*, **24**, 437-448.
- REPP, B. H., & LIN, H. B. (1988). Acoustic properties and perception of stop consonant release transients. *Haskins Laboratories Status Report on Speech Research*, **SR95/96**, 17-50.
- SIMON, C., & FOURCIN, A. J. (1976). Differences between individual listeners in their comprehension of speech and perception of sound patterns. *Speech & Hearing: Work in Progress UCL*, **2**, 94-125.
- STEVENS, K. N., & KLATT, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, **55**, 653-659.
- STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, **77**, 234-249.
- VAN TASSELL, D. J., HAGEN, L. T., KOBLAS, L. L., & PENNER, S. G. (1982). Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects. *Journal of the Acoustical Society of America*, **72**, 1771-1780.
- WALLEY, A. C., & CARRELL, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **51**, 1309-1317.
- WARDRIE-FRUIT, C. (1985). The effect of signal degradation on the status of cues to voicing in utterance-final stop consonants. *Journal of the Acoustical Society of America*, **77**, 1907-1912.
- WEST, J. (1976). A new speech synthesizer. *Speech & Hearing: Work in Progress UCL*, **2**, 4-8.
- ZLATIN, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, **56**, 981-994.

#### APPENDIX

##### Statistical Techniques

All statistical analyses were performed using GLIM 3.77 software (Payne, 1985), which allows a wide range of generalized linear models to be fitted easily to data with the use of a maximum likelihood criterion. One important decision that preceded the analysis was the choice of appropriate link function. In applying GLMs to the categorization of speech continua, this is equivalent to the choice of sigmoid function with which to fit the labeling function. GLIM allows three primary choices of link for modeling data that are binomially distributed, an appropriate model for the two-alternative forced-choice judgments

required here (e.g., BAIT or DATE). The *probit* link assumes the use of a cumulative normal distribution and leads to analyses identical to probit analysis and the maximal likelihood techniques described in Bock and Jones (1968). The *logit* link is commonly used by statisticians, because it is more easily interpreted with respect to an odds ratio. In fact, there are usually few differences in the conclusions arrived at with the use of these two links. Finally, the *complementary log-log* link differs from the previous two in that it is not symmetric around its 50% point (here, the phoneme boundary). Also at issue is the choice of appropriate scale for the stimulus continua, in particular, whether they should be linear or logarithmic. To arrive at an objective choice of link, each of the 64 individual labeling functions (16 subjects  $\times$  4 sessions) for each of the 8 stimulus continua in

the main experiment were fitted with all the combinations of stimulus scale (linear and log) and link function (logit, probit, and complementary log-log), and the overall goodness of fit was assessed by the sum of the Pearson's generalized chi-square. The complementary log-log link, with a linear stimulus scale, gave the best fit generally and so was used for all further analyses.

For further information on the theory and practice of generalized linear models, see Aitkin, Anderson, Francis, and Hinde (1989), Dobson (1983), McCullagh and Nelder (1983), and Payne (1985).

(Manuscript received April 27, 1990;  
revision accepted for publication September 16, 1990.)