Perception & Psychophysics
1980, 28 (5), 407-412

Influence of preceding liquid on stop-consonant perception

VIRGINIA A. MANN
Bryn Mawr College, Bryn Mawr, Pennsylvania 19010
and Haskins Laboratories, New Haven, Connecticut 06510

Certain attributes of a syllable-final liquid can influence the perceived place of articulation of a following stop consonant. To demonstrate this perceptual context effect, the CV portions of natural tokens of [al-da], [al-ga], [ar-da], [ar-ga] were excised and replaced with closely matched synthetic stimuli drawn from a [da]-[ga] continuum. The resulting hybrid disyllables were then presented to listeners who labeled both liquids and stops. The natural CV portions had two different effects on perception of the synthetic CVs. First, there was an effect of liquid category: Listeners perceived "g" more often in the context of [al] than in that of [ar]. Second, there was an effect due to tokens of [al] and [ar] having been produced before [da] or [ga]: More "g" percepts occurred when stops followed liquids that had been produced before [g]. A hypothesis that each of these perceptual effects finds a parallel in speech production is supported by spectrograms of the original utterances. Here, it seems, is another instance in which findings in speech perception reflect compensation for coarticulation during speech production.

When an utterance is articulated, the gestures for adjacent phonemes overlap and become interwoven. One consequence of this coarticulation of adjacent phonemes is that stop consonants may have different places of occlusion when they occur in different phonetic sequences. To date, the best known illustration of this point concerns the shift in place of occlusion that is consequent upon a change in the preceding or following vowel. Velar stops receive a more forward place of occlusion when they are adjacent to a front vowel such as [i] than when they are adjacent to a back vowel such as [a] (Öhman, 1966; Gay, Note 1). Another example, which has recently emerged from Repp and Mann's (in press) perceptual and acoustic observations of stops in fricative-stop clusters, is that when [t] or [k] follow [s], these stops can receive a relatively more forward place of articulation than when they follow [ʃ].

Insofar as coarticulation with adjacent phones causes shifts in the place of stop occlusion and, correspondingly, changes in the acoustic signal that reflects stop production, we should suppose that perception of a stop consonant must often require the integration of acoustic cues that are numerous, diverse, and context sensitive. That listeners do, in fact, integrate such cues in the process of stop perception can be seen in the existence of two perceptual "context effects" that reflect perceptual compensation for the particular coarticulatory effects cited above. With regard to the relative fronting of velar stops before vowels such as [i]—which causes release bursts to be relatively higher in frequency—Liberman, Delattre, and Cooper (1952) have shown that when steady-state synthetic vowels are preceded by bursts of various frequencies, listeners require a higher frequency burst to hear [k] before [i] than before [a]. With regard to the fronting of stops following [s], Mann and Repp (in press) report that when stimuli from a [ta]-[ka] continuum are preceded by a fricative noise appropriate to [s] (natural or synthetic), listeners give more "g" responses than when the preceding noise is appropriate to [ʃ].

These and other instances in which perceptual findings parallel the dynamics of speech production have led some investigators (see, e.g., Liberman et al., 1952; Mann & Repp, 1980, in press; Repp, Liberman, Eccardt, & Pesetsky, 1978; Repp & Mann, in press) to the view that speech perception operates with some reference to the dynamics of speech production. According to this view, perceptual context effects, such as those described above, should be found wherever stop production is influenced by production of an adjacent phonetic segment. This prediction is clearly upheld by the above-mentioned findings that stop perception is influenced by an adjacent vowel (Liberman et al., 1952) or fricative (Mann & Repp, in press). The purpose of the present experiment was to determine whether perceived place of stop occlusion could be influenced by a preceding liquid, since it seemed possible that a preceding liquid can influence the production of a following stop.

This research was supported by NICHD Grant HD01994 and BRS Grant R05596 to the Haskins Laboratories and by NICHD Postdoctoral Fellowship HD05667 to the author. Some of the results were reported at the 99th Meeting of the Acoustical Society of America in Atlanta, April 1980. I would like to thank Bruno Repp and Alvin Liberman for their advice at all stages of this project.

0031-5117/80/110407-06$00.85/0

Table 1
Mean Duration (in Milliseconds) and Intensity for Naturally Produced VC Syllables

	Duration								Relative Peak Amplitude (in Decibels, Arbitrary Reference)							
	AL-(DA)		AL-(GA)		AR-(DA)		AR-(GA)		AL-(DA)		AL-(GA)		AR-(DA)		AR-(GA)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V́C-CV	278	24	252	29	287	14	248	22	9.1	.4	9.4	1.0	5.1	2.8	6.4	.8
VC-CV́	240	3	245	9	239	11	243	13	−6.0	1.3	−9.0	1.3	−3.6	.1	−5.3	1.6

There are two circumstances under which a liquid may precede a stop: The liquid and stop may either occur as a syllable-final cluster or be separated by a syllable boundary. Here I have focused on liquid-stop sequences of the latter type, since in that case, a finding that liquids influence stop perception would have the additional implication that listeners are able to integrate perceptual information across a syllable boundary. One might expect the preceding liquid to influence perception of the following stop in disyllables such as [al-da], [al-ga], [ar-da], and [ar-ga], since articulation of the liquid most probably overlaps that of the stop. Although the literature does not provide any systematic observations on liquid-stop clusters, it seems at least possible that stops that follow [l] may receive a more forward place of articulation than those that follow [r], considering the fact that coarticulatory effects tend to be assimilatory in nature. It furthermore seems highly likely that the place of stop occlusion is reflected in the portion of the utterance immediately preceding the closure (i.e., in the portion commonly associated with the liquid). Thus, there might be coarticulatory effects in both directions, with appropriate acoustic and perceptual consequences.

The present experiment addressed these possibilities by excising naturally produced VC syllables from utterances of [al-da], [al-ga], [ar-da], and [ar-ga] and following them with stimuli from a [da]-[ga] continuum. Two questions were of interest: First, would a preceding [l] lead to more "g" responses than a preceding [r]? If so, it would suggest that listeners compensate in perception for a "left-to-right" coarticulatory influence of the liquid on the stop. Second, would liquids that had been coarticulated with [ga] lead to more "g" percepts than those coarticulated with [da]? If so, it would suggest that listeners are sensitive to a "right-to-left" coarticulatory influence of the stop on the liquid. In addition, as a means of obtaining more direct evidence for the coarticulatory phenomena underlying the two proposed perceptual effects, acoustic measurements were made of the utterances from which the stimuli were constructed.

METHOD

Subjects

The subjects included the author, a research assistant, and eight paid volunteers. Since experience with listening to synthetic speech did not seem to influence the pattern of results, all data were pooled.

Materials

A male, phonetically trained native speaker of English (L.J.R.) produced six repetitions each of [al-da], [al-ga], [ar-da], and [ar-ga]. These disyllables were produced according to a random sequence in which, as a control for any effects of stress pattern, half received syllable-initial stress and half received syllable-final stress. All utterances were recorded onto magnetic tape, using a Shure dynamic microphone in a soundproof room, before being digitized at 10,000 Hz using the Haskins Laboratories Pulse Code Modulation (PCM) system. Subsequently, separate files were created for the VC and CV portions of each disyllable, that is, the signal portions preceding and following the stop-closure interval. The VC syllables were stored for later use in constructing "hybrid" disyllables. Their durations and relative peak amplitudes are listed in Table 1. The natural CV syllables were analyzed, using the "Convert" program in conjunction with the Haskins Laboratories OVE IIIc synthesizer. (See Kuhn, 1977, for details of the Convert procedure.) Their duration, pitch contour, amplitude contour, and average formant frequencies were taken as guidelines for constructing two seven-member [da]-[ga] continua. The stimuli along each continuum differed only in the onset of F_3, which ranged from 2,690 to 2,104 Hz in approximately equal steps. Onset values for F_1 and F_2 transitions were fixed at 310 and 1,588 Hz, respectively. Steady-state values for the first three formants were 649, 1,131, and 2,448 Hz, respectively, and all formant transitions were stepwise linear and 100 msec in duration. For stimuli along the "stressed" continuum, stimulus duration (240 msec), amplitude contour, and pitch contour were those of a syllable (chosen at random from the several tokens) that had received primary stress. For those along the "unstressed" continuum, duration (180 msec), amplitude contour, and pitch contour were those of a syllable (also chosen at random) that had not been stressed. The relative peak amplitude was 3 dB below that of the "stressed" syllables. The two continua were otherwise identical, with the stimulus at each position from the stressed continuum having the same formant structure as that at the corresponding position from the unstressed one.

The actual test materials were constructed by combining the previously stored natural VC syllables with the stimuli along the two synthetic continua. All synthetic stimuli were first digitized at 10,000 Hz; those stimuli along the stressed continuum were then preceded by tokens of [al] and [ar] that had not received primary stress, whereas stimuli along the unstressed continuum were preceded by VC tokens that had received primary stress. In all cases, a 50-msec silent gap separated VC offset from the onset of the synthetic CV syllable. This value, although slightly shorter than the mean closure duration of the original natural utterances (80 msec), was still within the range of closure durations found in those utterances. As there were 12 tokens of [al] and 12 of [ar] (3 tokens, 2 contexts, and 2 stress conditions), combination of each token with the seven stimuli from along the appropriate synthetic continuum resulted in a total of 168 hybrid disyllables. These disyllables were recorded onto a test tape (the VC-CV tape) in two randomized sequences, with interstimulus intervals of 3 sec and longer pauses between sets of 56 stimuli. A second test tape (the CV tape) contained a randomized sequence of the stimuli along the two [da]-[ga] continua, repeated 12 times.

Procedure

Each subject participated in a single 80-min session during

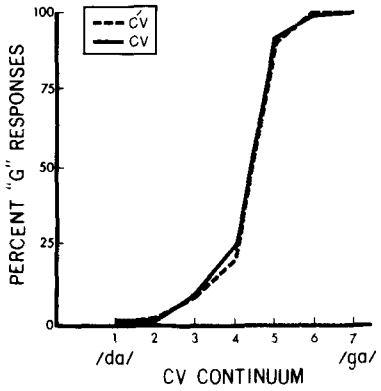

Figure 1. Percentage of "g" responses given to isolated CV stimuli from the synthetic [da]-[ga] continuum.

which he or she was seated in a soundproof room and listened to stimuli over TDH-39 earphones. The CV tape was presented first, followed by a short break. There was next a short practice sequence of hybrid disyllables that contained only the endpoint stimuli from the two CV continua; it was followed by two presentations of the VC-CV test tape. Thus, each stimulus was presented 12 times (ignoring token differences in the natural-speech portions).

In responding to the CV tape, the subjects were asked to identify each stop as "d" or "g." For the hybrid disyllables, they were asked to identify both the liquid as "l" or "r" and the following stop as "d" or "g."

RESULTS

In light of the novelty of the procedure for combining natural and synthetic syllables into test utterances, I was pleased to learn that the subjects' reactions to the hybrid disyllables were quite favorable. In fact, several listeners spontaneously praised the disyllables' resemblance to natural speech. None of the subjects had any difficulty hearing both liquids and stops; moreover, all of them were completely accurate in labeling the liquid consonants.

Consider first the pattern of responses to the isolated CV stimuli. Figure 1 plots the percentage of "g" responses given to each stimulus as a function of F_3 onset frequency. It can be seen that those stimuli at the first position, which contained a third-formant onset frequency appropriate for [da], received no "g" responses, while those at the seventh position, which contained a third-formant onset frequency appropriate for [ga], received 100% "g" responses. Between these two endpoints, the function follows the ogive pattern characteristic of identification functions obtained with stop-consonant continua. Note that the function obtained with stimuli whose duration, pitch contour, and amplitude contour were appropriate for a CV in stressed position (dashed line) is no different from that obtained with stimuli whose structure was appropriate for a CV in unstressed position (solid line).

Let us now turn to the main concern of this study, which was the question of whether labeling of stimuli along the [da]-[ga] continua would be altered by the presence of a preceding liquid. In the introduction, two possible effects were outlined, one concerning an effect of liquid category, the other concerning an effect due to the liquids' having been produced before [d] or [g]. The effect of liquid category membership was hypothesized to be that a preceding [l] would, in general, lead to more "g" responses than a preceding [r]. The relevant results are graphed in Figure 2, where it can be seen that the hypothesis was confirmed. There is a clear difference between the effects of preceding [l] (solid line) and preceding [r] (dashed line): Stops preceded by [l] were much more likely to be assigned a velar place of articulation. This effect was highly significant [$F(1,9) = 52.16$, $p < .0005$] and is primarily due to [l], for while there was no significant difference between the percentage of "g" responses given to CV stimuli preceded by [r] and that for CV stimuli presented in isolation, labeling of stimuli preceded by [l] significantly differed from the baseline [$F(1,9) = 50.1$, $p < .0005$]. A comparison of

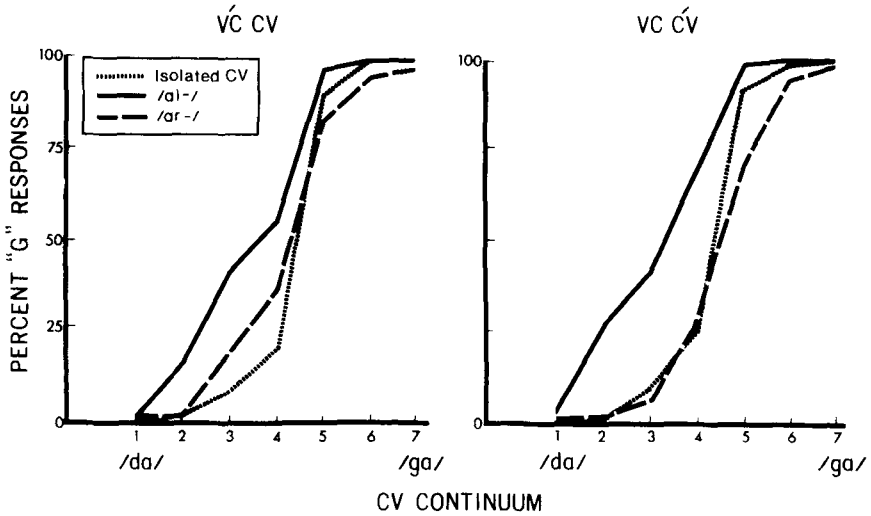

Figure 2. Percentage of "g" responses given to CV stimuli as a function of the category of the preceding liquid.

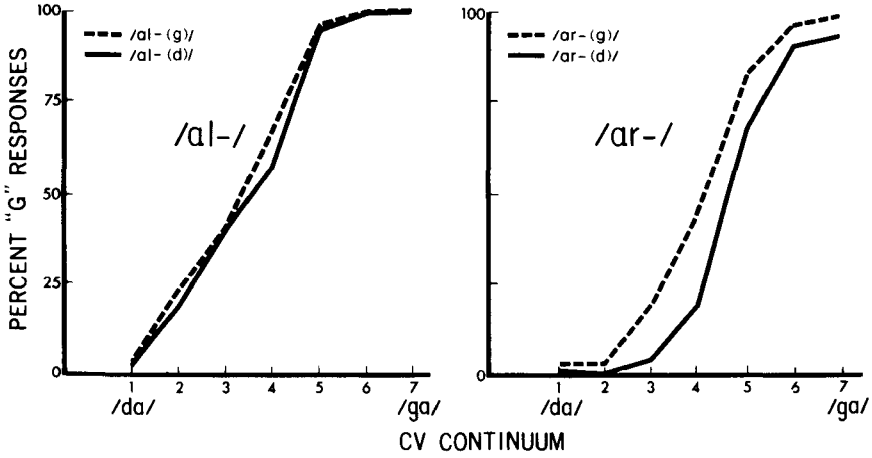

Figure 3. Percentage of "g" responses given to CV stimuli as a function of whether the preceding liquid had originally been produced before [d] or [g].

the left and right panels of Figure 2 also reveals that the difference between the effects of [l] and [r] on stop perception was somewhat greater when the liquid-final syllable did not receive primary stress [$F(1,9) = 8.13$, $p < .025$]. However, this paradoxical effect of stress did not appear to hold for all individual tokens of [al] and [ar], since it fell short of significance in a min F′ analysis (Clark, 1973). The effect of liquid context, on the other hand, remained significant [min $F'(1,5) = 18.4$, $p < .01$].

The second question asked in the introduction was whether tokens of [al] and [ar] that had been produced before [ga] would lead to more "g" responses than those produced before [da], all other things being equal. In that case, the relevant results are graphed in Figure 3, where the left panel shows the percentage of "g" responses to synthetic CV stimuli preceded by [al] and the right shows the corresponding percentages for stimuli preceded by [ar]. In each panel, it can be seen that liquids that had been produced before [ga] (dashed line) led to more "g" responses than those produced before [da] (solid line). It is also evident that the effect is considerably stronger for [ar] than for [al]. An analysis of variance computed on the percentage of "g" responses reveals a significant effect of original stop ([g] vs. [d]) [$F(1,9) = 35.63$, $p < .0005$] and an interaction between this effect and liquid category [$F(1,9) = 13.32$, $p < .005$]. Neither of these effects was influenced by the stress pattern of the disyllables, and both are upheld by the results of a min F′ analysis with tokens treated as a random variable. [For the effect of original context, min $F'(1,11) = 28.0$, $p < .0005$; for the interaction between this effect and liquid type, min $F'(1,7) = 6.74$, $p < .05$.]

DISCUSSION

Through a technique of combining natural and synthetic syllables into hybrid disyllables, the present experiment revealed that certain attributes of a preceding liquid can influence the perceived place of stop occlusion. Two influences are evident in the pattern of stop labeling functions obtained when naturally produced tokens of [al] and [ar] preceded stimuli along a [da]-[ga] continuum. First, there was an influence of liquid category: Many more "g" percepts occurred when synthetic CV stimuli were preceded by [l] than when preceded by [r]. Second, there was an effect due to liquids' having been produced before [d] or [g]: Many more "g" percepts occurred when the preceding liquid had been originally produced before [g] than when it had been produced before [d], this effect being much stronger for [r] than for [l].

The finding that [l], in general, led to more "g" percepts than [r] is remarkably similar to a finding observed in studies of the influence of preceding fricatives on stop perception (Mann & Repp, in press): [l], which has a more forward place of articulation than [r], leads to relatively more velar stop responses, just as does [s], which has a more forward place of articulation than [ʃ]. The fact that [s] leads to more velar responses than [ʃ] has been attributed to the fact that subjects are, in some sense, aware that stops that follow [s] can receive a relatively more forward place of articulation than those that follow [ʃ]. Perhaps the contrasting effects of [l] and [r] could be similarly explained. Certainly, this contrast cannot be explained reasonably in terms of the relative frequencies of various liquid-stop clusters in the English language, especially since the effect operates across a syllable boundary. On the other hand, the design of the present experiment does not eliminate the possibility that the results are due to some auditory interaction involving VC offset and CV onset spectra. For example, the contrasting effects of [l] and [r] could conceivably be the consequence of some form of auditory contrast between the concentration of energy in the F_3 region at the end of the preceding VC and

that in the F_3 region in the beginning of the following CV. Perhaps the relatively higher F_3 offset frequency in [l] led to the perception of a lower F_3 onset frequency in the following CV syllable and thus to more [g] percepts. Nevertheless, the conjecture outlined in the introduction also remains plausible—namely, that stops that follow [l] were more often perceived as "g," because stops that follow [l] tend to be produced with a relatively more forward place of articulation than those that follow [ar].

In an attempt to gain some support for this contention, we turn to spectrographic measurements of the natural CV syllables from which the test materials were constructed. (See the appendix for a discussion of the method employed.) Average formant transitions for these syllables are shown in Figure 4 with values for [da] and [ga] represented separately. Comparison of the transitions for stops preceded by [l] (dashed line) with those for stops preceded by [r] (solid line) reveals that stops that followed [l] had greater separation between the onset values of F_2 and F_3. Since velar stops typically show a greater convergence of the onset values for these two formants than for alveolar ones, this finding accords with the view that stops that follow [l] can receive a relatively more forward place of occlusion. Certainly, the extent to which such fronting is typical of all speakers remains an open question. For the moment, however, it is sufficient to note that the present perceptual context effect was obtained with the voice of a speaker who tended to front stops after [l].

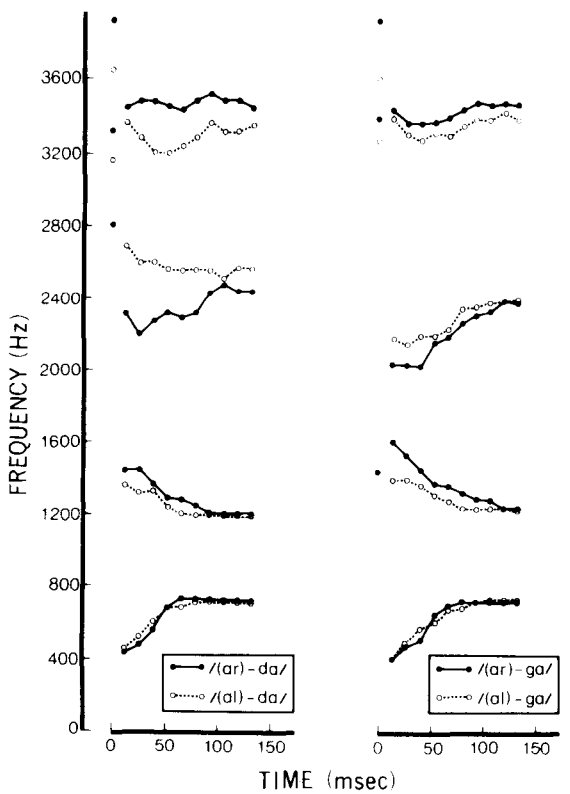

Figure 4. Average formant values for the first 145 msec of natural tokens of [da] and [ga], plotted separately for tokens produced after [al] and [ar].

Table 2
Average Formant Offset Frequencies in Naturally Produced VC Syllables

	F_1		F_2		F_3		F_4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AR-(DA)	400	67	1473	17	1680	143	2727	89
AR-(GA)	407	30	1306	49	1786	106	3453	218
AL-(DA)	447	141	927	40	2773	41	3553	119
AL-(GA)	420	49	1020	79	2649	39	3573	200

Thus, a plausible explanation of the effect of liquid category membership is that it reflects perceptual compensation for left-to-right, or perseverative, coarticulation in the production of liquid-stop sequences. The effect due to [al] and [ar] having been produced before [d] or [g] likewise may derive from a coarticulatory influence—but for one that is right to left, or anticipatory, in nature. This second effect is also different from the first in that it is a direct consequence of coarticulation-induced variation rather than a compensation for that variation. Thus, it is analogous to the finding (Repp & Mann, in press) that when synthetic stimuli from a [da]-[ga] continuum are preceded by fricative noises excised from naturally produced fricative-stop sequences, they tend to be perceived as the stop that originally followed the fricative. For fricatives, however, it has further been shown that the acoustic consequence of coarticulation with a following stop is an observable change in noise spectrum. The implication, then, is that when [al] or [ar] preceded velar or alveolar stops, they may have contained cues to the following stop, because stop production systematically influenced some aspect of their acoustic structure. The fact that such systematic influences were indeed present can be seen in Table 2, where, for tokens of [al] and [ar], the average offset spectra are given as a function of whether they preceded [da] or [ga]. (The method used in obtaining these measurements is described in the appendix.) For both [al] and [ar], offset spectrum was considerably influenced by the place of the following stop. Indeed, the following stop had a relatively greater influence on [ar], which is consistent with the perceptual results obtained with these stimuli. The fact that listeners are able to make correct use of such influences as cues to stop perception attests to the view that speech perception must somehow operate with tacit reference to the dynamics of speech production and its acoustic consequences. How else can we explain the fact that such a multiplicity of cues seems capable of influencing stop-consonant perception? The commonality between those cues is neither their acoustic structure

nor their location in time, but rather that they reflect one and the same "articulatory act" (Repp et al., 1978).

In summary, the high degree of consistency between the present perceptual findings and the dynamics of speech production is reminiscent of that seen in several previous studies of contextual influences on stop-consonant perception. Clearly, the conclusion to be drawn from this consistency is that the observed influences of liquid context reflect listeners' sensitivity to the coarticulatory influences involved in the production of liquid-stop sequences. There are two aspects of this sensitivity that are particularly relevant to our understanding of the type of mechanisms that must be accomplishing human speech perception: First, that perception takes into account coarticulatory influences in both directions, that is, from left to right and right to left; second, that it can operate across a well-defined syllable boundary. These results, which cannot easily be explained by models of speech perception that place extensive reliance on either phoneme- or syllable-sized templates, accord with the view that speech perception is an active process guided as if by some tacit knowledge of articulatory dynamics.

REFERENCE NOTE

1. Gay, T. *Cinefluorographic and electromyographic studies of articulatory organization* (Status Report on Speech Research, SR-50, 77-92). New Haven, Conn: Haskins Laboratories, 1977.

REFERENCES

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.

Kuhn, G. M. Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance. *Journal of the Acoustical Society of America*, 1977, **65**, 774-788.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. The role of selected variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 1952, **65**, 497-516.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, **28**, 213-228.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, **39**, 151-168.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 621-637.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, in press.

APPENDIX

In measuring the formant frequencies of the naturally produced syllables, I relied on spectral cross sections generated by a Federal Scientific UA-6A spectrum analyzer and displayed as point plots on a Hewlett-Packard 1300 oscilloscope, together with a computer-generated spectrogram and waveform display. All spectral information was smoothed and preemphasized. The cross sections were derived from 25.6-msec windows in 12.8-msec steps. The precise location of the first window could not be controlled; thus, the first section of each syllable usually included some of the silence preceding the utterance, and spectral peaks usually were not evident until the second section. The location of peaks for the first four formants was estimated visually, the maximum resolution being 40 Hz.

Two portions of each disyllable were of particular interest: the offset of the VC syllable and the transitions in the CV syllable. For each portion, I computed formant values that were subsequently averaged across the three tokens of each disyllable in each of the two stress patterns. Spurious peaks that were not common to all six tokens were omitted. Table 2 gives the average formant values for the last cross section of the VC syllable that contained peaks for each of the first four formants. Figure 4 shows the formant values for the initial 12 sections of the CV syllable, starting with the first section with measurable spectral energy.

(Received for publication June 6, 1980;
accepted revision received August 9, 1980.)