Markedness Asymmetries in Place Perception in Consonant Clusters

John Kingston and **Takahito Shinya**

Linguistics Department, University of Massachusetts, Amherst
jkingston@linguist.umass.edu, tshinya@linguist.umass.edu

ABSTRACT

We replicated and extended Repp's (1983) study of context effects on consonant place perception in VC1C2V strings. A 7-step [b-d-g] continuum was synthesized in which F2 and F3 followed mirror-image transitions in VC1 and C2V. VC1 was spliced together with C2V with silent gaps lasting 50, 100, 150, or 200 ms separating them. Single C responses were most frequent at the shortest closure duration, but didn't disappear even at the longest, and were only slightly less frequent when VC transitions differed acoustically from CV transitions. Single C responses were more often the unmarked coronal than the marked labial or dorsal places. Coronal responses to C1 were more frequent before a coronal C2, but more frequent to C2 after a non-coronal C1. Labial responses to both C1 and C2 were more frequent next to a non-labial C. Dorsal responses to C1 were also more frequent before a non-dorsal C2, but differed little to C2 with C1's place.

1 INTRODUCTION

Repp [1] reports that listeners frequently identify C1 in VC1C2V stimuli as identical to C2, particularly when the closure separating the VC1 formant transitions from the C2V transitions is short. On the other hand, they frequently identify C2 as different from C1, especially when the closure duration is long. Repp studied only labial and coronal stops, and he doesn't separate his results by the stop's perceived place. This study adds dorsal stops and breaks the responses down by place, to see if the unmarked coronal place behaves differently than the marked labial and dorsal places.

2 METHODS

2.1 PARTICIPANTS

Responses were analyzed from nineteen volunteers recruited from undergraduate linguistics courses at the University of Massachusetts, Amherst. None reported hearing or speaking pathologies. Data collected from nine participants weren't analyzed because they either didn't speak American English natively (3), their responses were incomplete or not determined by stimulus properties (6).

2.2 STIMULI

VCCV stimuli were constructed by splicing together nearly exact mirror image VC and CV syllables synthesized using the Sensimetrics implementation of the KLSYN88 terminal analogue synthesizer. The V was [a]. Figure 1 displays the [b], [g], and [d] endpoint frequencies for of the first three formants in the VC syllable; their frequencies followed mirror image trajectories in the CV syllables. F2 and F3 frequencies were 1220 and 2600 Hz, respectively, during their 210 ms steady states. To create a place of articulation continuum in the Cs, the frequencies of these formants changed over the remaining 60 ms of the syllables. Their offset and onset frequencies varied in seven equal log steps from [b] (F2 1100 Hz, F3 2150 Hz) to [g] (F2 1780 Hz, F3 2150 Hz) to [d] (F2 1780 Hz, F3 2800 Hz). F2 offset and onset frequencies alone vary from [b] to [g] and F3 onset and offset frequencies alone vary from [g] to [d]. F1, F4, and F5 followed the same trajectories in all stimuli: F1 was 700 Hz throughout its 250 ms steady state, and fell to 300 Hz over 20 ms at the vowel edges; F4 was constant at 3500 Hz and F5 constant at 4500 Hz through out all the stimuli.
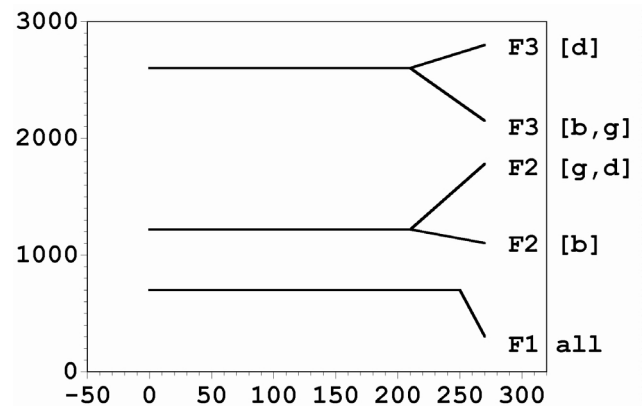

Figure 1. Trajectories of F1-F3 in the VC syllables, showing the F2 and F3 offset values for the best [b], [d], and [g] stimuli.

Only the F0 contours differed between the VC and CV syllables: F0 fell linearly from 140 to 130 Hz across the VC syllables, and from 125 to 115 Hz across the CV syllables. As a result, when the VC and CV syllables were spliced together, F0 fell continuously throughout the resulting VCCV stimuli. The VC and CV stimuli were spliced together with either 50, 100, 150, or 200 ms of silence between them, yielding 196 stimuli (4 silence or closure durations x 7 VC syllables x 7 CV syllables).

2.3 PROCEDURES

The stimuli were presented to listeners one at a time over TDH-49 headphones at self-selected comfortable listening levels. They identified the first C as "b", "d", or "g" by

pressing labeled buttons. A rapid tone triplet was then played to prompt a response to the second C. Listeners pressed one of the same labeled buttons again if they heard a second C, or they pressed a button labeled "0" if they heard only one C. The listeners had up 1500 ms in each interval to identify the two consonants. The second response was followed by 1000 ms of silence before the next stimulus was presented.

Listeners were familiarized with the stimuli with an initial block of trials in which they heard all combinations of the best exemplars of the three places of articulation and all closure durations once (3 VC x 3 CV x 4 closure durations = 36 training stimuli). The listeners were then presented with four blocks of trials in each of which they heard all 196 stimuli once.

3 RESULTS

Responses from the 19 listeners were pooled because no single listener responded enough times to any single stimulus to analyze his or her responses separately. Results are therefore based on a maximum of 76 responses per stimulus, but because listeners occasionally failed to respond to a stimulus before the maximum time to respond had elapsed, the analysis is based on an average of 72 responses/stimulus.

3.1 SINGLE CONSONANT RESPONSES

Figure 2 displays the mean proportion of single consonant responses to all stimuli (light bars) and to stimuli in which F2 and F3 offset frequencies in VC were not identical to their onset frequencies in CV (dark bars), as a function of closure duration.
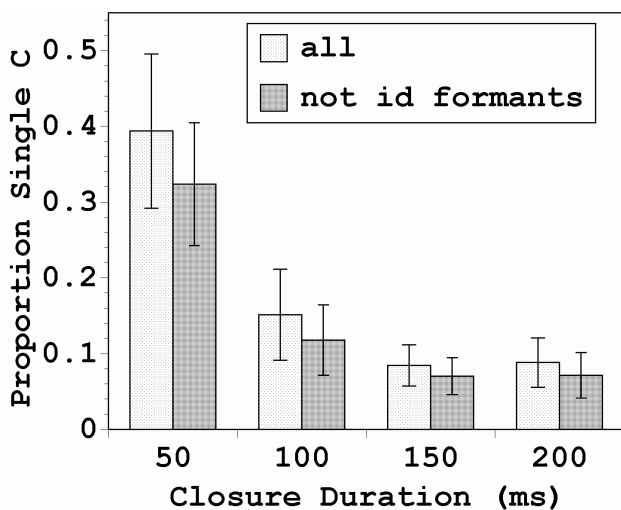
Figure 2. Mean proportion of single C responses (± 95% confidence intervals) for each closure duration to all stimuli (light bars) and to stimuli in which VC and CV formant transitions weren't acoustically identical (dark bars).

Unsurprisingly, single consonant responses drop in frequency as closure duration increases, and are somewhat fewer when VC and CV formant frequencies aren't identical. This figure shows, too, that there is an irreducible number of single C responses, of roughly 0.07 of responses, regardless

of the difference between VC and CV formant transitions or their separation.

Figure 3 displays the percentage of single consonant responses in which listeners responded with each place of articulation, as a function of closure duration. Responses to stimuli in which the VC formant frequencies aren't identical to the CV formant frequencies are displayed on the left, and those in which they are identical are displayed on the right. Single "d" responses are more frequent at all closure durations than either single "b" or "g" responses, and they are also more frequent at longer than shorter closure durations. As closure duration increases, listeners respond with single "g" less often and more often with single "d" and to lesser extent single "b". All these effects are more pronounced when VC formant frequencies are identical to CV formant frequencies (right) than when they're not (left).
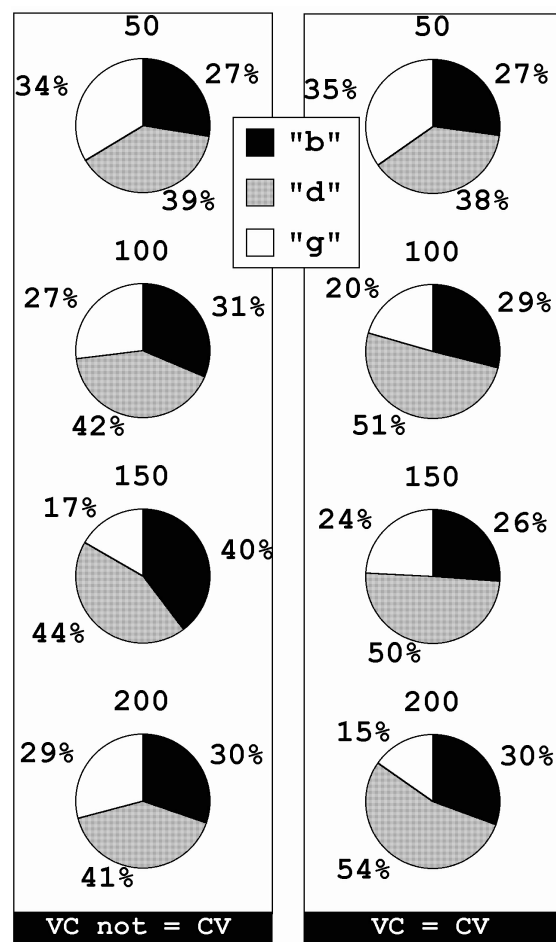
Figure 3. Single "b", "d", and "g" response percentages as a function of closure duration for stimuli in which VC formant frequencies are identical to CV formant frequencies (left) vs those in which they are not (right).

3.2 CONTRAST VS ASSIMILATION

Following Repp (1983), one C *assimilates* to the other when listeners identify that C's place of articulation more often as the same as that of the other consonant but *contrasts* when

they instead identify with a different place of articulation. Expressions such as that in (1) quantify this difference:

$$[P(\text{"}b\text{"}) \,/\, a\_.ba] - [P(\text{"}b\text{"}) \,/\, a\_.\{d,g\}a] \quad (1)$$

"b" in C1 assimilates when the value in (1) is greater than 0, and contrasts when it's less that 0. Scores were calculated for "b" responses before the best [b] vs. all other stimuli. (Calculations before the stimulus most often labeled "b" yielded very similar results, and aren't reported here.) Contrast-assimilation scores were calculated including the single C responses among the assimilated responses (lower case symbols and dashed lines in Figures 4 and 5) and leaving them out altogether (upper case symbols and solid lines in Figures 4 and 5). Analogous calculations were carried out for the other two places of articulation and for Cs in CV as well as VC. Figures 4 and 5 display the effects of C2's place on the percept of C1's place and of C1's place on the percept of C2's place, respectively.
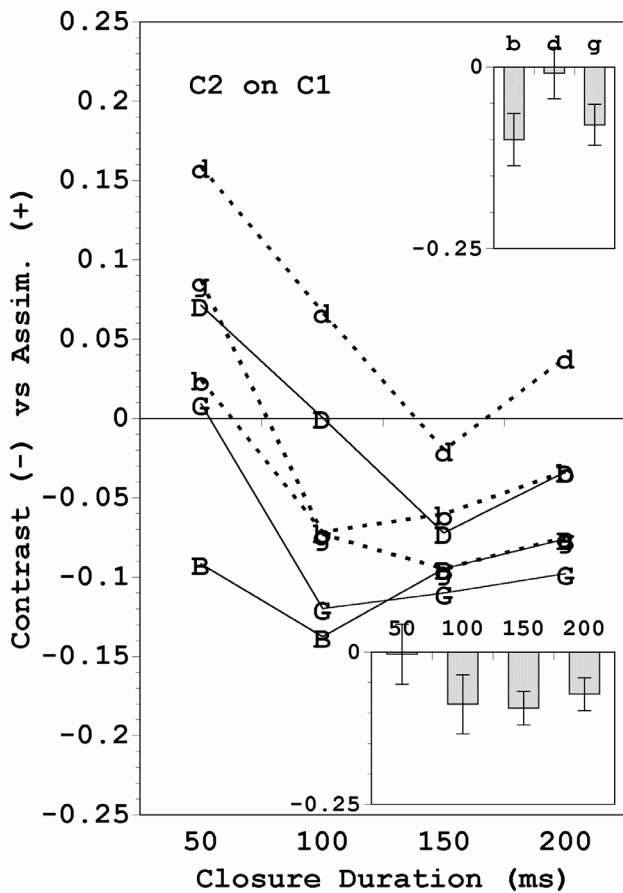

Figure 4. Mean assimilation to vs contrast with C2 for "b", "d", and "g" responses to C1 for all responses (lower case, dashed lines) and excluding single C responses (upper case, solid lines). Insets: Place and Duration means (95% CIs) excluding single C responses.

For all three place responses in Figure 4, the dashed lines lie above the solid ones, which shows that single C responses are assimilative. Even excluding the single C responses, "d" responses to C1 are more assimilative or less contrastive at all closure durations than "b" or "g" responses, and "b"

responses are more contrastive than "g" responses at the shortest closure duration. That is, when C2 is the best [d] stimulus, listeners are more likely to identify C1 as "d" than they are to identify it as "b" or "g" before the best [b] and [g] stimulus, respectively. At the shortest closure duration, they are also less likely to identify a stimulus as "b" before the best [b] than to identify it as "g" before the best [g]. Finally, "d" and "g" responses to C1 shift from being assimilative or neutral at the shortest closure duration to contrastive at longer closure durations (> 50 ms), but "b" responses are contrastive even at the shortest closure duration.
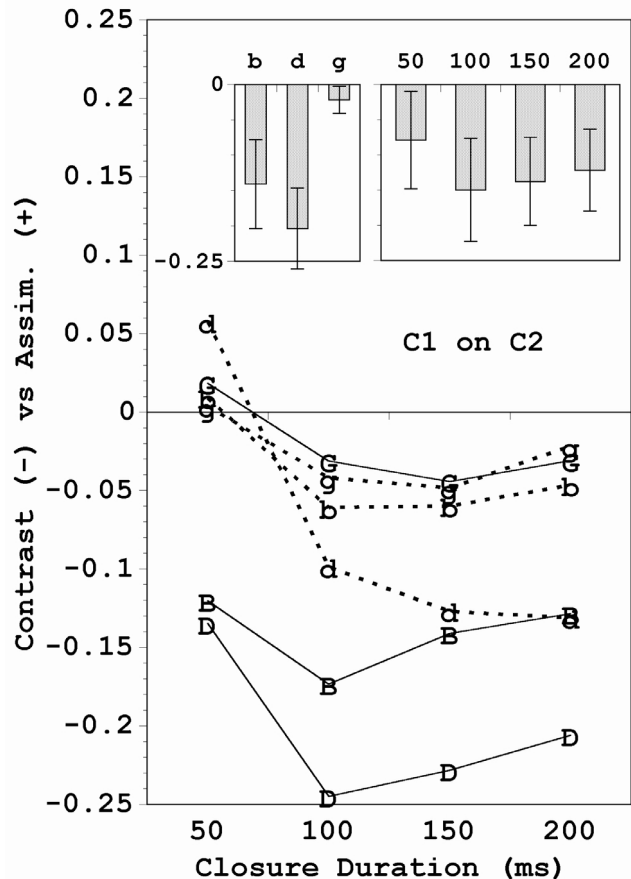

Figure 5. Mean assimilation to vs contrast with C1 for "b", "d", and "g" responses to C2 for all responses (lower case, dashed lines) and excluding single C responses (upper case, solid lines). Insets: Place and Duration means (95% CIs) excluding single C responses.

The dashed lines lie reliably above the solid ones for the "b" and "d" but not "g" responses in Figure 5, which shows that single C responses are only assimilative in C2 for the labial and coronal places. Once these responses are excluded, the "b" and especially the "d" scores are strongly contrastive. Listeners are much more likely to hear one of these two places after any C1 than the best [b] or [d] stimuli, respectively. "g" responses are neither assimilative nor contrastive at the shortest closure duration and at best weakly contrastive at longer durations.

Repeated measures ANOVAs were run on assimilation-contrast scores calculated as in (1): including vs excluding

single C responses for regressive (C2 on C1) vs progressive (C1 on C2) effects of the context stimulus. The independent variables were the Place of articulation of the response ("b" vs "d" vs "g") and the closure Duration (50 vs 100 vs 150 vs 200 ms). These tests are equivalent to an analysis in which items are the random variable.

In the analysis of the regressive effect of C2 on C1 (Figure 4), Place was significant in the analyses including and excluding single C responses [including: $F(2,12) = 13.934$, $p < .001$; excluding: $F(2,12) = 5.331$, $p < .025$], and Duration was, too [including: $F(3,18) = 16.277$, $p < .001$; excluding: $F(3,18) = 7.112$, $p < .005$]. These variables did not interact significantly in the analysis including single C responses but they did in the one excluding them [including: $F(6,36) = 2.063$, $p < .10$; excluding: $F(6,36) = 2.941$, $p < .025$]. The latter interaction is probably significant because "d" scores differ less from "b" and "g" scores at longer closure durations (> 100 ms) and "b" responses are only more contrastive than "g" responses at the shortest closure duration. A planned contrast in the analysis excluding single C responses showed that "d" scores were significantly more assimilative than the "b" and "g" scores [$F(1,6) = 17.169$, $p < .001$]. Similarly, a planned contrast showed that scores at the shortest closure duration were also significantly more assimilative than the scores at the longer durations [$F(1,6) = 8.943$, $p < .025$].

In the analyses of the progressive effect of C1 on C2 (Figure 5), Place wasn't significant in either analysis [including: $F(2,12) = 1.502$, $p > .25$; excluding: $F(2,12) = 2.277$, $p > .10$], but Duration was significant for both [including: $F(3,18) = 18.394$, $p < .001$; excluding: $F(3,18) = 5.262$, $p < .005$]. In the analysis including the single C responses, these two variables also interacted significantly, but they did interact in the analysis in which those responses were excluded [including: $F(6,36) = 5.476$, $p < .001$; excluding: $F(6,36) = 1.314$, $p > .25$]. This interaction is significant in the analysis including single C responses because "d" scores change much more, from assimilative to contrastive, as a function of closure duration than "b" or "g" scores. Planned contrasts in the analysis excluding single C responses showed that "g" scores were significantly less contrastive than "b" and "d" scores [$F(1,6) = 65.347$, $p < .001$]. Although there is also trend for scores to be less contrastive at the shortest than the longer closure durations, it's at best marginally significant [$F(1,6) = 5.347$, $p < .10$].

4 DISCUSSION

Listeners are most likely to hear just a single C when the closure duration is shortest, even when the formant transitions at the end of the VC syllable aren't identical (in mirror image) to those at the beginning of the CV syllable (Figure 2). In fact, they frequently hear just a single C at this duration, on 3-4 per 10 trials. The frequency of single C responses drops sharply at the next longer closure duration, to 1-2 per 10 trials, but listeners still respond with just a single C even at the longest closure duration, and they continue to do so for stimuli in which VC and CV transitions

differ acoustically. It would be unsurprising if listeners found it hard to detect acoustic differences between VC and CV transitions at the shortest closure duration. That they continue to respond with single Cs at longer closure durations even when VC and CV transitions differ shows that they instead occasionally integrate the place information from these two sources into a single place percept.

At all closure durations, single "d" responses outnumber single "b" or "g" responses (Figure 3). Single "b" responses also outnumber single "g" responses at all but the shortest closure duration. Both differences are obtained for stimuli in which the VC transitions differ from the CV transitions as well as those in which they are identical. when they hear just a single C, listeners are biased to respond with the unmarked coronal place, or with the labial rather than the dorsal place among the marked places.

The contrast-assimilation scores (Figures 4 and 5) show that "d" responses to C1 are significantly more assimilative than "b" or "g" responses, and that "b" and especially "d" responses to C2 are significantly more contrastive than "g" responses. The unmarked coronal place clearly varies the most as a function of its position in the VCCV string: listeners hear C1 as "d" when C2 is the best [d] stimulus, but hear C2 as "d" when C1 is anything but the best [d] stimulus. In both C1 and C2, listeners are more likely to hear "b" when the other C is anything but the best [b]. "g" is similar in C1 except at the shortest closure duration where "g" responses are as likely before the best [g] as any other C2. In C2, however, "g" responses are at most slightly more likely at any closure duration after any C1 than the best [g].

In other words, coronal percepts are weak in C1 unless supported by coronal in C2, but they are strong in C2. Labial percepts are strong in both C1 and C2, and dorsal percepts are reasonably strong in C1 but neither strong nor weak in C2. These strength differences reflect an interaction of the place of articulation's characteristic acoustics with position in syllable but not with differences in the salience of the place information in the syllable coda. By design, the quality of the place information was identical in codas and onsets. Coronals in codas therefore assimilate in place to following onsets but labials and dorsals do not because coronal's acoustics are perceived less reliably in codas.

REFERENCE

[1] B. H. Repp, "Bidirectional context effects in the perception of VC-CV sequences," *Perception & Psychophysics*, vol. 33, pp. 147-155, 1983.