

Contrast Effects Do Not Underlie Effects of Preceding Liquids on Stop-Consonant Identification by Humans

Carol A. Fowler and Julie M. Brown
Haskins Laboratories and University of Connecticut

Virginia A. Mann
University of California, Irvine

These experiments explored the claim by A. Lotto and K. Kluender (1998) that frequency contrast explains listeners' compensations for coarticulation in the case of liquid consonants coarticulating with following stops. Evidence of frequency contrast in experiments that tested for it directly was not found, but Lotto and Kluender's finding that high- and low-frequency precursor tones can produce contrastive effects on stop-consonant judgments were replicated. The effect depends on the amplitude relation of the tones to the third formant (F3) of the stops. This implies that the tones mask F3 information in the stop consonants. It is unknown whether liquids and following stops in natural speech are in an appropriate intensity relation for masking of the stop. A final experiment, exploiting the McGurk effect, showed compensation for coarticulation by listeners when neither frequency contrast nor masking can be the source of the compensations.

Talkers coarticulate, that is, they produce vocal tract gestures for consonants and vowels in overlapping time frames. Gestures are linguistically significant actions of the vocal tract. For example, the coordinated action of the lips and the jaw that closes the lips for /b/, /p/, or /m/ is a gesture. Talkers begin a gesture or gestures for a segment while those of another segment are ongoing (anticipatory coarticulation), and they complete production of a segment's gesture or gestures after those for another segment have begun (carryover or perseveratory coarticulation).

Considerable research has been interpreted as showing that listeners are remarkably attuned to coarticulation in speech. Indeed, they appear to "parse" acoustic speech signals along gestural lines. One index of gestural parsing is that, given acoustic information for a segment, say, y, that, due to coarticulation occurs in the domain of predominantly acoustic information for an earlier (or later) segment, x, listeners use that information to support perception of y (e.g., Fowler, 1984; Fowler & Brown, 2000; Fowler & Smith, 1986; Marslen-Wilson & Warren, 1994; Martin & Bunnell, 1981; Whalen, 1982, 1984). For example, they use acoustic information for /u/ in the frication noise for a preceding /s/ or /ʃ/ as information for /u/ (Whalen, 1984). When acoustic information for /u/ is spliced after frication that was originally produced

in the context of /a/, listeners' identification of /u/ is slowed by the misinformation in the frication. A second index of parsing along gestural lines is provided by discrimination experiments (Fowler, 1981; Fowler & Smith, 1986) showing that listeners do not hear segments as context sensitive that, due to coarticulation, are specified by context-sensitive acoustic signals. Asked to judge the relative similarity of pairs of consonant-vowel (CV) syllables, they judge acoustically different syllables, each in its proper coarticulatory context, as more similar than acoustically identical syllables, one of which is presented in a different phonetic context than that in which it was originally produced. These discrimination judgments, particularly coupled with identification judgments (Fowler & Smith, 1986) showing that listeners use the context-sensitive information nonetheless, suggest that listeners ascribe the context sensitivity to the relevant contextual segments.

A third index of listeners' attunements to coarticulated speech, and the one on which we focus, is sometimes called *compensation for coarticulation*. This refers to findings that listeners' identifications of consonants or vowels can be different in different coarticulatory contexts, as if listeners are "compensating" for probable acoustic consequences of coarticulation by contextual segments (Mann, 1980, 1986; Mann & Repp, 1980; Mann & Soli, 1991). The finding on which we focus was first reported by Mann (1980) who found that members of a continuum of syllables ranging from /da/ (with a high falling third formant [F3]) to /ga/ (with a low rising F3) were identified differently following precursor /a/ and /ar/ (ar) syllables. In particular, listeners identified ambiguous syllables more often as /ga/ following /a/ than following /ar/. Mann's (1980) account was that carryover coarticulation between /l/ and a following stop that has a more back place of articulation (/g/ in her stimuli) pulls the stop's place of articulation forward. Carryover from /r/, a more back consonant than /l/, does not; instead, it pulls the place of articulation of /d/ back. Listeners' attunement to coarticulated speech enables them to compensate for these coarticulatory effects, that is, to count more fronted conso-

Carol A. Fowler and Julie M. Brown, Haskins Laboratories, New Haven, Connecticut and Department of Psychology, University of Connecticut; Virginia A. Mann, Department of Psychology, University of California, Irvine.

This research was supported by National Institute of Child Health and Human Development (NICHD) Grant HD-01994 to Haskins Laboratories. A brief report of some of these results appeared in the proceedings of the 14th Congress of Phonetic Sciences, August 1–7, 1999, San Francisco, California.

Correspondence concerning this article should be addressed to Carol A. Fowler, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511. Electronic mail may be sent to fowler@haskins.yale.edu.

nants as /g/ in the context of /a/ than of /r/. In subsequent research, Mann (1986) found remarkably similar compensation among Japanese listeners who were unable to label /l/ and /r/ accurately or consistently. Fowler, Best, and McRoberts (1990) found qualitatively similar compensation in infants.

Recent evidence has suggested, however, that, in at least some instances of compensation for coarticulation, compensation is due to something other than acoustic or phonetic information about coarticulation. In several experiments, Elman and McClelland (1988) found compensation for effects of /s/ or /ʃ/ on identifications of following consonants ranging from /d/ to /g/ or /t/ to /k/ (with more /d/ or /t/ judgments following the more back /ʃ/, a well-known compensatory effect, e.g., Mann & Repp, 1980). However, in the research by Elman and McClelland, the identity of /s/ or /ʃ/ was determined by the lexical identity of the word in which it was embedded (e.g., "foolish" versus "Christmas"); the acoustic signals for /s/ and /ʃ/ were identical and ambiguous between signals characteristic of /s/ and /ʃ/. Compensation, in these experiments, appears to arise from information at a level of description of the speech stimuli above that on which coarticulation occurs (i.e., lexical rather than phonetic or acoustic). Recently, Pitt and McQueen (1998) have shown that the interpretation that the source of compensation is lexical is probably wrong. They determined that the compensation found by Elman and McClelland (1988) is not due to lexical information, but to prelexical knowledge of transition probabilities. Specifically, /s/ occurs more commonly after the final vowel in "Christmas" than does /ʃ/. /ʃ/ is more common after the final vowel in "foolish" than is /s/. This reinterpretation of the findings in terms of transition probabilities, however, does not change the fact that compensation for coarticulation can occur based on information other than acoustic or phonetic evidence of coarticulation.

The gestural parsing account of perception of coarticulated speech has been challenged from below the phonetic level as well. Recently, Lotto and Kluender (1998; Lotto, Kluender, & Holt, 1997) concluded that Mann's (1980) findings were not due to listeners' attunements to gestural overlap in speech but rather were due to frequency contrast.

Contrast effects may occur quite commonly. A frequency contrast effect is one in which, in the context of a high-frequency acoustic signal, a signal of intermediate frequency is judged lower in pitch than it is judged in the context of a low-frequency signal; this effect was reported by Cathcart and Dawson (1928–1929). However, contrast effects are much more general than that. For example, an object that is intermediate in weight among a set of objects is judged lighter by participants who have just hefted a heavier weight than by participants who have just hefted a lighter weight (Johnson, 1944). One account of these effects is offered by Warren (1985). It is that the effect reflects a kind of adaptive attunement of perceivers to the nature of stimuli with which they are interacting, such that they displace perceptual criteria (e.g., a boundary along the weight dimension that partitions weights into the categories heavy or light) in the direction of recently encountered stimuli. This renders perceivers more sensitive to differences in the vicinity of recently encountered stimuli. Accordingly, having experienced a very heavy weight, perceivers shift the heavy–light boundary toward the heavy end of the scale, and they judge some objects light that in other contexts they judged to be heavy. Lotto and Kluender (1998) allude to this account of contrast, but

they (see also Lotto et al., 1997) offer a somewhat different account of contrast that we will consider in the following discussion and in the General Discussion.

Results reported by Mann (1980) might be due to frequency contrast. In her stimuli, /l/ had a final F3 value that exceeded the starting F3 of the most /da/-like member of the /da/-/ga/ continuum (and so exceeded the starting F3 of every member of the continuum); /r/ had a final F3 value that was lower than the starting F3 value of the most /ga/-like member of the continuum (and so fell below the starting F3 value of every continuum member). These relations between the ending F3 of /l/ or /r/ and the starting F3 of the continuum members may have induced a contrast effect such that /a/ lowered the effective F3s of following consonants and /ar/ raised it. This would lead to a finding of more /g/ responses after /a/ than /ar/. Note that an account in terms of contrast renders Mann's (1986) findings with Japanese listeners and Fowler et al.'s (1990) findings with infants less surprising and remarkable than they are in the context of a gestural-parsing account.

To test the contrast hypothesis, Lotto and Kluender (1998) used precursor tones rather than precursor syllables. The tones were sine waves that either tracked the center frequency of /l/'s or /r/'s F3 (Experiment 2) or were level tones at the final F3 value of /l/ or /r/ (Experiment 3). Listeners reported more /ga/s following the higher frequency tone than following the lower frequency tone, supporting the contrast interpretation. To test further the idea that apparent compensation for coarticulation is due to an auditory process such as contrast rather than to attunement to gestural overlap, Lotto et al. (1997) tested quail on stimuli like those of Mann (1980). Quail were first trained on endpoint /da/ and /ga/ syllables that were presented in the context of each of three precursor syllables, /a/, /al/, and /ar/, for three of the four quail but presented in isolation for the remaining quail. (Whether or not training included the precursor syllables did not affect the outcome; accordingly, data from the fourth quail were pooled with those of the other quail in its condition.) Two quail were trained to peck to /da/ syllables and to withhold responding to /ga/; the remaining two quail responded to /ga/, not /da/. After training, the quail were tested on all continuum members in the context of the three precursors, and the investigators measured the frequencies of their pecks to the continuum members depending on the identity of the precursor syllable. The quail that were trained to peck to /ga/ did so more frequently with precursor /al/ than with /ar/; those that were trained to peck to /da/ did so less frequently with precursor /al/ than with /ar/. These findings, too, are consistent with an account in terms of contrast. It is assumed that quail are not attuned to acoustic consequences of gestural overlap in human speech.

Coarticulation quite generally causes acoustic assimilation. That is, coarticulation by one gesture in the domain of another has acoustic consequences that are similar to the acoustic signal in the coarticulating segment's own domain. If effects of context segments on perception of their neighbors are generally contrastive, they will qualitatively compensate for the coarticulatory effects. Accordingly, the account of compensation offered by Lotto and Kluender and by Lotto et al. may have very broad applicability. It need not be restricted to compensation for coarticulatory effects of /l/ and /r/ on /da/ and /ga/.

Despite the strength of the evidence offered by Lotto and colleagues, we chose to test the contrast account further. We had

several reasons, both empirical and theoretical, to doubt that contrast provides an accurate account of compensation for coarticulation. In addition, in our view, the account stems from a mistaken perspective on the nature of the perceptual system that subserves speech perception.

One empirical reason to doubt the contrast account of compensations for coarticulation derives from the first index of listeners' attunement to gestural overlap that we listed earlier. Contrast should eliminate or at least substantially reduce the availability to listeners of coarticulatory information for phonetic identity. (For example, in the perception of an utterance of /arda/, a frequency contrast effect exerted by /ar/ should eliminate or reduce perceptibility of the /r/ coloring of /d/ by raising the effective F3 onset of /d/.) But many studies show that listeners use coarticulatory information for phonetic identity (e.g., Fowler, 1984; Fowler & Brown, 2000; Fowler & Smith, 1986; Marslen-Wilson & Warren, 1994; Martin & Bunnell, 1981; Whalen, 1982, 1984). It is true, as the contrast account predicts, that listeners do not hear a segment specified by a context-sensitive acoustic signal as context sensitive (our second index of gestural parsing listed earlier; e.g., Fowler, 1981; Fowler & Smith, 1986). But listeners' use of the coarticulatory information as information for the coarticulating segment even when their discrimination judgments reveal a context-free percept shows that the information has not been lost as it would be according to a contrast account. Rather, it has been properly ascribed to the coarticulating segment.

A second empirical reason to doubt the contrast account is provided by a finding of Mann and Liberman (1983). They presented the disyllables of Mann (1980) to listeners' left ear, except the critical F3 transition for members of the /da/-/ga/ continuum, which they presented to the right ear. Under these conditions, listeners appear to hear the F3 transition in two ways simultaneously, exhibiting *duplex perception*. They hear it both as part of the CV syllable, namely the part that determines whether the syllable is /da/ or /ga/, and they hear it as a pitch glide (a "chirp"). Listeners participated in two sessions in which they took an AXB discrimination test on the same stimuli. (In an AXB test, three stimuli [here, three duplex disyllables] are presented successively. Listeners judge whether the middle stimulus [X] is more like the first [A] or the third [B].) In the tests, the CVs of A and B were three steps apart along the /da/-/ga/ continuum, and X was identical either to A or to B. In one session, listeners attended to the speech and judged whether the disyllable X sounded more like A or B. In the other session, they attended to the chirps and decided whether the middle chirp X sounded more like the chirp of A or that of B. In the speech task, listeners exhibited a shift in the place along the /da/-/ga/ continuum where discrimination performance peaked depending on whether the precursor syllable was /a/ or /r/. Discrimination peaks were closer to the /da/ end of the continuum when the precursor was /a/ than when it was /r/. Because discrimination peaks tend to occur at category boundaries, this is consistent with listeners hearing more /ga/s following /a/ than /r/. Discrimination functions based on responses to the chirps were alike, however, revealing no differential effect of the precursor vowel-consonant (VC) syllables on discriminations. If contrast effects were the source of the discrimination peak shift in the speech test, a peak shift should have been apparent in the chirp discrimination functions as well, because the acoustic stimuli were identical in the discrimination tests.

A third reason to doubt the contrast account is that the version of it offered by Lotto et al., general as it is, is not sufficiently general to explain compensation for coarticulation. To explain the ubiquity of contrast effects, Lotto et al. (1997) suggest that they are adaptive:

Due to the variables of inertia and mass, physical systems tend to be assimilative. The configuration of a system at time t is significantly constrained by its configuration at time $t - 1$ Perceptual systems have developed in an environment governed by particular physical laws and it is probable that perceptual processes respect these laws. (p. 1139)

The reason why this particular account of contrast as compensation for inertia is insufficiently general is that it explains only compensation for carryover coarticulation. However, coarticulation is bidirectional, and listeners compensate bidirectionally. It happens that the particular instance of coarticulation examined by Mann (1980) was an instance of carryover coarticulation, which some researchers have ascribed to inertia of articulators. (See Daniloff and Hammarberg, 1973, for a review and for a statement of some deficiencies of the account of carryover effects as due to inertia.) However, anticipatory coarticulation occurs as well, and compensating for it requires a direction of compensation that is opposite to that proposed by Lotto et al. Listeners do compensate for anticipatory coarticulation. For example, Mann and Repp (1980) and Mann and Soli (1991) found compensation for anticipatory lip rounding in listeners' identifications of preceding fricatives.

A final empirical reason to doubt the contrast account is that it has not been established that frequency contrast occurs with stimuli like those used by Lotto and Kluender (1998). They cited two studies (Cathcart & Dawson, 1928-1929; Christman, 1954) as providing evidence for contrastive effects of frequency on pitch judgments. The studies do report contrast effects, but with stimuli that are nothing like the syllables or tones used by Lotto and Kluender (1998). For example, Christman (1954) presented what he called *satiating tones* to the left ear for 1 or 2 minutes followed by a standard tone (600 Hz) to the left ear and a variable tone to the right ear. He determined the frequency of the tone presented to the right ear that was perceived to match the standard in the satiated ear, and found contrastlike consequences of satiation. Cathcart and Dawson do not tell the reader how long the stimuli were that they used to induce contrast; but for the experiments that provided the most consistent evidence of contrast, the tones were produced by a dulcitone, which is described as a piano with tuning forks in place of strings. The sounds produced by the dulcitone were unlikely to have been as short as the 250-ms tones used by Lotto and Kluender.

Aside from these reasons stemming from research findings, there are also two theoretical reasons why we doubt the contrast account of Lotto and Kluender (1998) and of Lotto et al. (1997). One is that, as a general account of compensation for coarticulation (as Lotto and Kluender's, 1998, title implies: "General Contrast Effects in Speech Perception . . ."), it has an implication that is implausible. The implication is that, although acoustic signals provide information about coarticulation, and although listeners compensate for coarticulation (as if they used the information as such), listeners do not use information about coarticulation as such. Consider these observations. Compensation for coarticulation sometimes occurs due to information in transition probabili-

ties about segment identity (Pitt & McQueen, 1998). This is information at a level of description of a speech event higher than the gestural-phonetic level at which coarticulation occurs and has acoustic consequences. According to a general contrast account, otherwise compensation occurs due to auditory contrast effects. Auditory contrast arises at a level of description of a speech event below that on which coarticulation occurs. This implies that compensation never occurs due to information about coarticulation itself in the acoustic signal or in a sequence of phones; that is, it never occurs at the very level of description of speech where information about coarticulation is available. We find this implausible.

The final theoretical ground on which we doubt the contrast account is the perspective on the nature of perceptual systems that it implies. It implies that, because there are general, lawful kinds of properties and events in the world (for Lotto et al., things have mass and inertia), there will have evolved very general adaptive resources or mechanisms for perceiving events having those properties. Contrast effects enhance accurate perception of real-world events (see the General Discussion for an elaboration of this idea). However, other points of view are possible, and we consider them more plausible. One, which we ascribe to motor theorists (e.g., Liberman & Mattingly, 1985), is that speech signals have many special properties, including consequences of coarticulation. Therefore, special mechanisms have evolved to deal with those special properties. Another view, is that perceivers perceive by attuning to a specifying structure in air for hearing, light for seeing and so on, which serves as information for its causes (e.g., Fowler, 1996). In speech, linguistically significant gestures of the vocal tract cause acoustic signals, and distinctive gestures cause distinctive acoustic signals, which, therefore, can specify their causes. Both of these latter views disagree with Lotto and Kluender's proposal that speech perception relies on very general processing resources. Rather, perception is attuned to the special properties of speech signals. In contrasting Lotto and Kluender's (1998) account of compensation with our own, we contrast their theoretical perspective with these others as well.

The aim of our first experiment was to test for frequency contrast directly, a test that failed to provide evidence of contrast. Our second experiment both replicated the findings of Lotto and Kluender showing effects of a precursor tone on /da-/ga/ identifications and provided some evidence that masking underlies the effect. Our third experiment shows that compensation for coarticulation occurs under bimodal conditions in which neither contrast nor masking can explain its occurrence.

Experiments 1a and 1b

We designed Experiment 1 to test for frequency contrast, by determining whether a precursor tone would affect the pitch of a following tone. Following Lotto and Kluender (1998), we used 250-ms precursors. Our test tones were 250-ms long in Experiment 1a to mirror the durations of /da-/ga/ syllables in the research of Lotto and Kluender and 50-ms long in Experiment 1b to mirror the durations of the /da-/ga/ formant transitions, which were the parts of the syllables that were supposed to be affected by contrast.

Method

Participants. Participants were 31 undergraduates who participated in the research for course credit. They were native speakers of English who reported having normal hearing. Sixteen participants took part in Experiment 1a and 15 in Experiment 1b.

Stimulus materials. Precursor tones were sine waves 250 ms in duration having a 5-ms amplitude ramp at onset and offset. They were synthesized using the program SWS (sine wave synthesis) at Haskins Laboratories (New Haven, CT). Following Lotto and Kluender (1998), we used two precursor tones, one at 1700 Hz and one at 2800 Hz. (These frequencies were chosen to match endpoint F3 frequencies of syllables /a/ and /ar/.) The precursor tones were followed after a 50-ms silence by the test tones. These were 10 steady-state test tones with frequencies that ranged in 100 Hz steps from 1800 Hz to 2700 Hz. In Experiment 1a, the test tones were 250 ms in duration; in Experiment 1b, they were 50 ms in duration. Like the precursor tones, they were ramped up in amplitude over the first 5 ms and were ramped down over the last 5 ms. The test tones were matched in amplitude to the precursor tones.

These stimuli were used to compose listening tests, which were recorded on audiotape. There were three tests. The first was a 20-item randomization of the endpoint (1800 Hz, 2700 Hz) test tones with 3.5 s between items to give participants time to make their identifications. The second was a 100-item randomization of 10 instances of each of the 10 test tones. The third was a 200-item randomization of 10 instances of each of the test tones preceded either by the 1700-Hz precursor or 2800-Hz precursor. Inter-stimulus intervals in these two tests were 3.5 s, except after trials corresponding to the end of a column on the answer sheet; these intervals were 6.5 s.

Procedure. Participants were tested in groups of 1 to 3 in a quiet room. They listened to the stimuli over headphones. We first gave them experience with the test tone endpoints (1800 Hz, 2700 Hz) by playing them in alternation three times. We instructed listeners to identify the lower pitched tone as *L* and the higher pitched tone as *H*. To verify that they understood the assignment of letters to tones, we had them listen to the 20-item randomization of the two endpoints and identify each tone as *H* or *L*. Next participants took part in two tests designed after those of Lotto and Kluender (1998). In the first, they heard all 10 test tones, which were each presented 10 times in random order. We then told the participants that they would be listening to tones of a variety of pitches including the tones that they had just learned to identify and other tones that were intermediate in pitch between the endpoint tones. Their task was to identify a tone as *H* if its pitch was more like that of the *H* than the *L* endpoint and as *L* if its pitch was more like that of the *L* than the *H* endpoint. They were required to choose either *H* or *L* on each trial. In the third and final test in which test tones were presented after the precursors, we told listeners that their task was the same except that each tone to be classified was preceded by a precursor tone.

Results

In their research, Lotto and Kluender (1998) used the results of the 100-item test to eliminate data of participants whose responses to the endpoints fell below 90%. In this and the remaining experiments that we report, we did not eliminate data based on this or any other criterion. In Experiment 2a, which provides a replication of Lotto and Kluender's (1998) Experiment 3, we analyzed the data both including and excluding participants' data that did not meet the 90% criterion. The results were very similar and were statistically the same in the two analyses. We prefer not to exclude data on this basis, because the percentages of excluded participants can be high (20% and 23.5% in the two experiments of Lotto and Kluender, 1998, in which they provide the numbers), and this can render the remaining sample unrepresentative of their population.

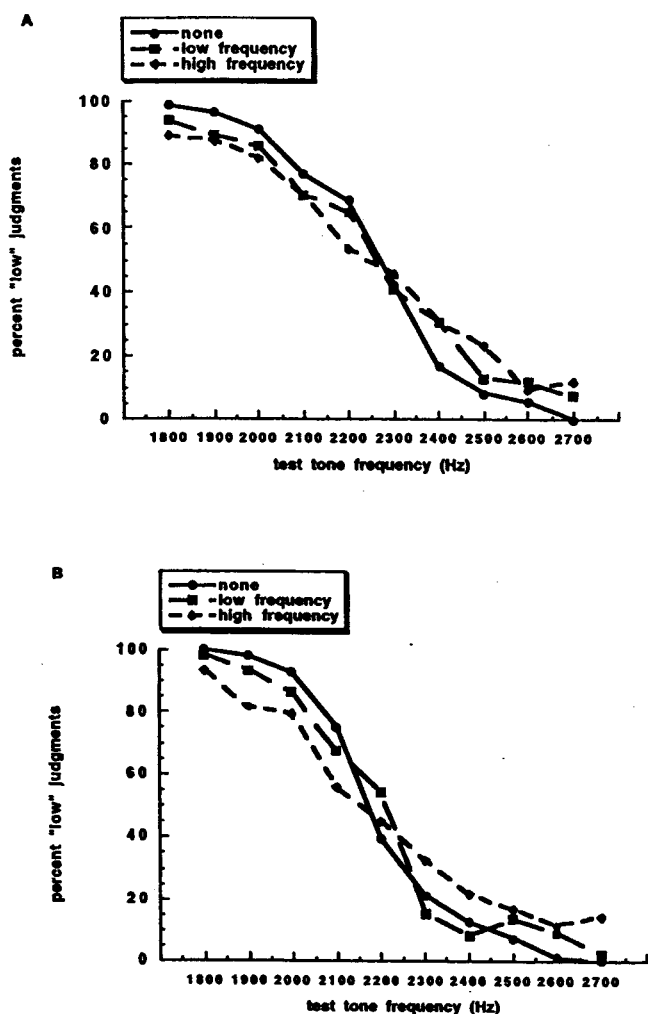


Figure 1. Percentage of "low" judgments to the (A) 250-ms and (B) 50-ms tones heard in isolation (no precursor) or following a low- or high-frequency precursor tone. Data from Experiments 1a and 1b.

Every participant scored 100% correct on the verification test that included only the endpoint tones. Figures 1A and 1B present the averaged findings from the remaining two tests on the 250-ms and 50-ms tones, respectively. Figure 1 shows the percent of *L* responses assigned to tones along the 10-item continuum. Figure 1 plots the data separately for the three precursor conditions (no precursor, 1700 Hz precursor, 2800 Hz precursor). A contrast effect would be manifest if there were more *H* responses in the low-frequency precursor condition than in the high-frequency precursor condition. There is no evidence for contrast in the outcome.

We ran analyses of variance (ANOVAs) on the data plotted in Figure 1; factors were continuum member (1–10) and precursor. The results were the same in the two analyses. The effect of continuum member was highly significant in Experiment 1a, $F(9, 135) = 174.83$, $p < .0001$, and in Experiment 1b, $F(9, 126) = 249.19$, $p < .0001$, reflecting the monotonic decreases in *L* responding for tones of higher frequency. The effect of precursor was nonsignificant (both F s < 1), but the interaction was significant in Experiment 1a; $F(18, 270) = 3.75$, $p < .0001$; and

Experiment 1b; $F(18, 252) = 4.81$, $p < .0001$. Figure 1 suggests that the interactions in Experiments 1a and 1b occurred because performance was more consistent at the extremes of the continuum in the no-precursor condition than in the others and that the higher frequency precursor tone was associated with the least consistent performance of all.

Consistent with the nonsignificance of the precursor effects in the ANOVA, *t* tests comparing pairs of precursor conditions on percent *L* judgments were uniformly nonsignificant in both experiments.

Discussion

We failed to obtain any indication of a contrast effect and also failed, therefore, to confirm that the frequency contrast effects that Cathcart and Dawson (1928–1929) and Christman (1954) found with quite different precursor tones and methods generalize to precursor tones like those used by Lotto and Kluender (1998). Presumably, then, the effects of the precursors that Lotto and Kluender found were not due to frequency contrast.

Lotto and Kluender (1998) briefly allude to a different kind of contrast effect that they call *spectral contrast*. A. Lotto (personal communication, May 8, 1998) has augmented the account of spectral contrast offered in that paper. Spectral contrast occurs when presentation of a tone reduces the effective amplitude of that tone's frequency, and perhaps nearby frequencies, in a subsequently presented acoustic stimulus. If spectral contrast underlies the findings of Lotto and Kluender, then perhaps we did not see any contrastive effect in Experiment 1 because none of our test tones, which were sine waves, had energy at the frequency of either precursor tone where the effect should have been largest.

In Experiment 2, we will offer evidence, albeit not definitive, that some such account may be accurate. We point out that what Lotto has described (see also Lotto and Kluender, 1998, p. 616) is likely a masking effect with contrast as a consequence. Moore (1988) reviews evidence that acoustic masks tend to reduce sensitivity to frequencies including and surrounding their own; the range of frequencies affected increases with the amplitude of the mask. This may be another reason why Experiments 1a and 1b failed to give evidence of contrast. Perhaps the precursor tones were insufficiently intense in relation to the intensity of the test tones to produce an effect. Experiment 2 was designed to confirm that we could find the effect of precursor tones on /da/-/ga/ judgments that Lotto and Kluender (1998) reported. It also provided an opportunity for a preliminary test of the masking account.¹

¹ We did not originally design Experiment 2 to test a masking account of Lotto and Kluender's (1998) findings. We designed it to replicate their Experiment 3. Using our own stimuli, we failed on a number of attempts. That led us to ask A. Lotto to send us his stimuli, which he kindly did. Acoustic comparison of our stimuli with those of Lotto and Kluender uncovered a mistake in our synthesis of the /da/-/ga/ stimuli. We had made F3 unnaturally relatively intense. That suggested the possibility of a test for masking.

Experiment 2a provides a replication of Lotto and Kluender's Experiment 3, using their stimuli.² Experiment 2b provides a replication using our own stimuli that differ in a critical way from those of Lotto and Kluender (1998). The amplitudes of the F3s of our stimuli are unnaturally high relative to the amplitudes of the lower formants.

Experiments 2a and 2b

In Experiments 2a and 2b, following Experiment 3 of Lotto and Kluender (1998), we obtained *D* and *G* judgments from listeners who heard members of a continuum of syllables presented in isolation or following a high- or low-frequency precursor tone.

Method

Participants. Thirty-one participants took part in the experiments; 18 in Experiment 2a and 13 in Experiment 2b. They were native speakers of English who reported having normal hearing. They received course credit for their participation.

Stimulus materials. The stimuli supplied by Lotto and used in Experiment 2a are described in the Method section of Lotto and Kluender's Experiment 1. We paraphrase that description here. Stimuli were synthesized using the Klatt synthesizer (Klatt, 1980). Endpoint /da/ and /ga/ stimuli were synthesized to copy a natural-speech production of each syllable by a male talker. The onset frequency of F3 varied in ten 100-Hz steps from 1800 Hz to 2700 Hz. The transition shifted the onset frequency linearly to the steady-state vowel F3 of 2450 Hz. The frequency of the first formant (F1) rose from 300 Hz to 750 Hz; the frequency of the second formant (F2) decreased from 1650 to 1200 Hz. The transitions of these formants were 80 ms in duration. Total syllable duration was 250 ms. The fundamental frequency of the syllables was 110 Hz falling to 95 Hz over the last 50 ms. Precursor tones were like those in our Experiments 1a and 1b except that they were matched in root mean square (RMS) amplitude to that of the synthetic CV syllables (Lotto & Kluender, 1998, p. 613).

The stimuli we used in Experiment 2b were modeled after those of Lotto and Kluender (1998), but they differed in two ways. First, we did not use the Klatt synthesizer, but rather we used a parallel synthesizer (and the software program SYN) at Haskins Laboratories (New Haven, CT). Acoustic comparison of the CVs produced by the two synthesizers revealed one notable difference. The frequency of the fourth formant (F4) was closer to that of F3 in the syllables produced by the Haskins synthesizer than in those produced by the Klatt synthesizer. This may increase the effective amplitude of the syllables in the frequency vicinity of F3 and provide some protection against masking. Second, we set the amplitude of F3 of the syllables to a value 20 dB higher than that of F1 rather than lower than those of F1 and F2 as occurs in natural productions of the syllables. This also should have the effect of protecting F3 against masking effects of the precursor tones.

Our precursor tones were those of Experiments 1a and 1b matched in RMS amplitude to the CV stimuli.

We devised two listening tests with the stimuli from each experiment. They were recorded on audiotape. In one test, the 10 continuum members were presented in isolation 10 times each in random order. In the other, the 10 continuum members were presented 10 times each in the context of the high- and low-precursor tones (200 trials in all). The interstimulus intervals in both tests were as in Experiment 1.

Procedure. Participants were tested individually in a quiet testing room. They listened over headphones. We first presented the /da/ and /ga/ endpoints in alternation to familiarize participants with the clearest /da/ and /ga/. Next, listeners took the 100-item test in which each continuum member was presented in isolation 10 times. We instructed them to write *D* or *G* on their answer sheets to identify each stimulus. We warned them

that some stimuli were ambiguous but that they were required to write down either *D* or *G* on each trial, guessing if necessary. Finally, participants took the 200-item tests in which continuum members were presented in the context of the high- and low-precursor tones. Again, we instructed participants to identify each syllable by writing *D* or *G* on their answer sheets.

Results

Figure 2A shows the findings (percent *G* judgments) with the stimuli used by Lotto and Kluender (1998). Our outcome was very similar to theirs and showed clear evidence of a contrastive effect of the precursor tones on the identification of /da/ and /ga/. That is, more *G* responses were reported in the context of the high-frequency than the low-frequency precursor tone. The responses in the no-precursor condition do not fall between those in the two precursor conditions, but rather they generally fall below both precursor conditions. This may or may not signify that only the high precursor was effective. Because listeners identified the isolated syllables in a preceding test that involved precursors, response criteria may have shifted from the first test to the second.

In an ANOVA with precursor condition and continuum member as factors, both main effects and the interaction were significant: precursor, $F(2, 34) = 11.15, p = .0002$; continuum, $F(9, 153) = 195.04, p < .0001$; Precursor \times Continuum, $F(18, 306) = 5.08, p < .0001$. As for the effect of precursor, more *G* responses occurred in the high-precursor than in the low-precursor condition (64.6% vs. 54.2%), $t(17) = 5.97, p < .0001$. The difference in the percentage of *G* responses in the no-precursor versus the high-precursor condition was significant, $t(17) = 4.15, p = .0007$; the difference between the no- and low-precursor conditions was marginal, $t(17) = 1.92, p = .07$. However, there were fewer *G* responses in the no-precursor condition (43.5%) than in the low-precursor condition, an unpredicted effect that may, however, reflect the fact that the no-precursor trials were blocked with respect to precursor trials. The effect of continuum was significant because the percentage of *G* responses decreased almost monotonically as F3 increased in starting frequency. The interaction reflected the fact, clear in Figure 2A, that the curves representing the low- and high-precursor conditions were separate except at the continuum endpoints where performance approached ceiling or floor.

Figure 2B shows the results with our continuum in which F3 was raised in intensity and F4 lay close to F3. There is no evidence of a contrastive effect of precursor tone in Experiment 2b. Rather, the three curves fall one on top of the other. In overall means, there were 53.9% *G* responses in the context of the high-precursor tones and 55.2% in the context of the low precursor.

In an ANOVA, the main effect of continuum, $F(9, 108) = 163.90, p < .0001$, and the interaction, $F(18, 216) = 2.76, p = .0003$, were significant. The main effect reflects the generally decreasing percentages of *G* responses as F3 increased in fre-

² In Lotto and Kluender's Experiments 1 and 2, they used a 10-item /da/-/ga/ continuum; in Experiment 3, they used a 7-item continuum modeled after that of Mann (1980). A. Lotto sent us both continua. In pretesting, we found more consistent classification of the endpoints of the 10-item continuum. Accordingly, we used that for our replication of Lotto and Kluender's Experiment 3.

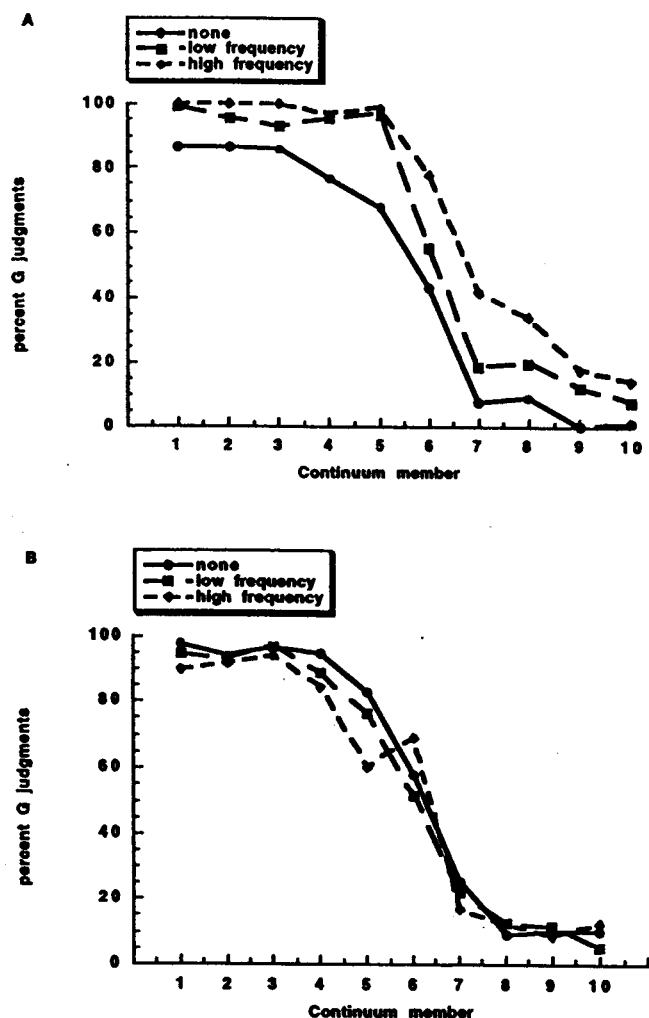


Figure 2. Percentage of *G* judgments given to synthetic consonant-vowel (CV) syllables heard in isolation or following a high- or low-frequency tone. Figure 2A presents data from Experiment 2a in which the synthetic syllables were those of Lotto and Kluender, 1998. Figure 2B presents data from Experiment 2b in which synthetic /da/-/ga/ syllables had relatively intense third formants (F3s).

quency. The interaction may largely reflect the unevenness of responses to the sixth and seventh continuum members in the high-precursor condition. We did not explore the source of the significant interaction further, because it does not relate to the predictions of an account of precursor effects in terms of contrast or masking. Results of *t* tests comparing the percentage of *G* responses between pairs of precursor conditions were uniformly nonsignificant.

Discussion

In Experiment 2a, we replicated the findings of Lotto and Kluender (1998) that high- and low-frequency precursor tones can affect listeners' identification of syllables as /ga/ or /da/ in a way similar to effects of preceding /a/ or /ar/ syllables. These effects were contrastive in direction in that a high-frequency precursor

tone increased identifications of syllables as *G*. Indeed, we found in listening to the same ambiguous syllable sequentially in the context of the high- and low-precursor tones that the contrastive effect is clearly audible.

The results of Experiment 2b in relation to those of Experiment 2a provide some support for an account of the effect in terms of masking of F3 information in the /da/ and /ga/ stimuli by the precursor tones. Raising the amplitude of F3 did not affect the identifiability of /da/ and /ga/. Responses to continuum endpoints were highly systematic. However, raising the amplitude does appear to have protected the syllables from any effect of the precursor tones. We infer that the effect that Lotto and Kluender identify as a contrast effect is contrast due to masking, and the amplitude of the precursor tone in relation to relevant acoustic structure in the following syllable is critical in determining whether there is or is not an effect of the context tone.

Given the apparent importance of relative amplitude to the occurrence of this effect, as we have noted, it is relevant to ask whether, in natural speech, the amplitude of F3 in /a/ and /ar/ syllables is sufficient to mask F3 of a following /da/ or /ga/ syllable. To our knowledge, these amplitude relations have not been reported. Accordingly, it is unknown whether natural speech provides conditions in which this masking effect might arise.³

Rather than address this question next, we chose to test the adequacy of the masking account in a different way. We asked whether listeners would compensate for coarticulatory effects of /l/ and /t/ on /da/-/ga/ continuum members when the information distinguishing /l/ from /t/ was optical rather than acoustic. In this way, we explored whether compensation for coarticulation occurs under conditions in which neither auditory contrast nor masking accounts can predict an effect.

Experiments 3a and 3b

We used our own continuum of /da/ and /ga/ syllables, rather than that of Lotto and Kluender. Despite their inappropriate F3

³ However, there is evidence that masking does not underlie compensation for coarticulatory effects of /a/ and /ar/ on *D* and *G* judgments that researchers have found. First, Mann (1980) manipulated stress in natural speech productions of the four disyllables and did not find a larger precursor syllable effect of stressed than unstressed /a/ and /ar/ on, respectively, unstressed and stressed /da/-/ga/ syllables. The initial syllables of the naturally produced disyllables served as precursors to members of either of two synthetic /da/-/ga/ continua. Stressed /a/ and /ar/ preceded members of a synthetic /da/-/ga/ continuum synthesized to mimic unstressed naturally produced /da/ and /ga/ syllables in duration, amplitude, and fundamental frequency contour. Unstressed /a/ and /ar/ preceded synthetic /da/ and /ga/ syllables that were synthesized to mimic naturally produced stressed /da/ and /ga/ syllables. Stress did not affect the magnitude of the precursor effect statistically. Numerically, the effect of unstressed /a/ and /ar/ was greater than the effect of the stressed precursors (Mann, 1980, Figure 2.) This provides a little evidence that the relative amplitudes of the VC and CV syllables are unimportant over some range for the occurrence of compensation for coarticulation. Second, Mann and Liberman (1983) found compensation for coarticulation when the precursor /a/ and /ar/ syllables were presented to the left ear, whereas the critical F3 transition for the /da/-/ga/ continuum members was presented to the right ear. This dichotic presentation should reduce or eliminate masking, but it did not eliminate compensation for coarticulation.

amplitude, the syllables sound as natural as synthetic syllables, and listeners were systematic in identifying them in Experiment 2B. Because the stimuli were deviant, we ensured that participants would compensate for coarticulation in their identifications of members of this continuum by performing an initial experiment (Experiment 3a) in which continuum members were preceded by acoustic /a/ and /a/ syllables and then performing a second experiment (Experiment 3b) in which information distinguishing /a/ from /a/ was optical.

In Experiment 3b, we synthesized a syllable that we judged to be ambiguous between /a/ and /a/. We dubbed that signal onto a videotape of a male speaker hyperarticulating /a/ or /a/. That is, we exploited the McGurk effect (e.g., McGurk & MacDonald, 1976) in which, under some conditions of dubbing, the sight of one syllable being produced is dubbed onto the acoustic signal for another syllable, and listeners report hearing a syllable that may be the same as the visible syllable or that may be an integration of information from the two modalities. For example, a video /da/ dubbed onto acoustic /ba/ typically leads to /da/ judgments of the acoustic syllable; video /da/ dubbed onto acoustic /ma/ leads to /na/ judgments.

The McGurk effect is most effective when syllables with labial and nonlabial consonants are cross-dubbed; /l/ and /r/ are not commonly studied in McGurk experiments because cross dubbing them is unlikely to give rise to strong McGurk effects. Even though /r/ has a lip-rounding constriction, it is not a labial consonant; neither is /l/. However, /l/ and /r/ are in different viseme classes (Walden, Prosek, Montgomery, Scherr, & Jones, 1977). If optical information for /l/ and /r/ give rise to any McGurk effect at all, it may be strongest if the acoustic signal presented with the video clips is ambiguous between /l/ and /r/. That is how we designed our stimuli.

Method

Participants. Twenty-six participants were run in Experiments 3a and 3b. Thirteen participated in each experiment for course credit. They were native speakers of English who reported normal hearing.

Stimulus materials. The continuum members were those of Experiment 2b. In addition, we synthesized /a/ and /a/ syllables for Experiment 3a using the description of their synthesis by Lotto and Kluender (1998) as a guide. The fundamental frequency of the syllables was fixed at 110 Hz. The steady-state values of F1, F2, and F3 for /a/ were 750 Hz, 1200 Hz, and 2450 Hz, respectively, and were constant over the first 100 ms of each syllable. The ending values of the formants in the /a/ syllable were 564 Hz, 956 Hz, and 2700 Hz; for /a/, they were 549 Hz, 1517 Hz, and 1600 Hz. These ending values were achieved by linear interpolation over 150 ms. We matched these syllables to members of the /da/-/ga/ continuum in RMS amplitude.

We made the ambiguous syllable of Experiment 3b by modifying the ending frequency of the formants for /a/ and /a/ until we found a set of values that we judged to be maximally ambiguous between /a/ and /a/. These values were 556 Hz, 1300 Hz, and 2150 Hz for F1, F2, and F3, respectively. In addition, to match the duration of the hyperarticulated syllables of our videotaped speaker, we increased the syllable in duration so that the whole syllable duration was 450 ms with a 200-ms steady-state component. After 50 ms of silence, the ambiguous syllable was followed by each of the continuum members.

For the video component of Experiment 3b, we videotaped a male speaker⁴ producing tokens of /alda/, /arda/, /alga/, and /arga/. We asked the speaker to hyperarticulate, emphasizing the rounding for /r/ and the tongue-tip constriction for /l/.

It was not feasible for us to use the same video clip of our talker producing /da/ or /ga/ in the context of both /a/ and /a/ by dubbing the video /da/ or /ga/ from one disyllabic utterance onto the other. This is because any differences in the positioning of the speaker's head between the two disyllabic utterances would make the video image jump unnaturally at the syllable break of the dubbed disyllable. Accordingly, the two video clips that we used were undubbed and therefore had nonidentical final syllables. We asked 15 students to look at 10 tokens each of each of the four disyllables in a video-only condition, with trial types randomized. They identified both consonants in a forced-choice test. For the video clips /alda/ and /arda/, we found that observers identified the second consonant as *D* equally often overall in the /a/ (67% of responses) and /a/ (65%) contexts. The corresponding percentages of *G* responses in the /alga/ and /arga/ conditions were 50% and 25%, respectively. To ensure that the video information about the second consonant of each disyllable did not bias observers to report *G* more often in the context of /a/, we used the /alda/ and /arda/ video clips.

The listening test of Experiment 3a was run like those of Experiments 1 and 2. That is, there were two tests, one that presented the 10 continuum members 10 times each in random order and a second 200-item test that presented the continuum members 10 times each with each of the two precursor syllables /a/ and /a/. In both tests, there were 3.5 s between trials, except in those trials that corresponded with the ends of columns on the answer sheet; these intervals were 6.5 s.

We ran the audiovisual component of Experiment 3b using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), which dubbed each acoustic disyllable with each video clip by outputting each speech file and video clip simultaneously. (Using Adobe Premier [Adobe Systems, Inc., San Jose, CA], we added silence to the onset of the audio files so that the audio files were timed appropriately to output concurrently with the video clips.) In this test, there were 200 trials in which each continuum member was preceded by the ambiguous syllable dubbed onto a video clip of our speaker mouthing /a/ or /a/.

Procedure. In Experiment 3a, participants listened to the tests over headphones. They were tested individually. They were first given experience with the continuum endpoints. Then, they took the 100-item test and wrote *D* or *G* on an answer sheet to identify the consonant that they heard. Finally, they took the 200-item test and again wrote *D* or *G* to signal the consonant that they heard on each trial.

The first phase of Experiment 3b was like that of Experiment 3a. In the first test, stimuli were presented acoustically. Listeners were given experience with the /da/ and /ga/ continuum endpoints, then they took the 100-item test by listening over headphones and writing their identification responses. After that, they sat at a computer facing the monitor. They were given experience with the four audiovisual endpoint stimuli (i.e., video /alda/ or /arda/ dubbed onto the ambiguous syllable followed after 50 ms by endpoint /da/ or endpoint /ga/). Each video display was approximately 4.5 in. in width and 4 in. in height (320 × 240 pixels) on the computer monitor. These stimuli were identified for the listener by the experimenter.

Next, participants took the 200-item audiovisual test. On each trial, participants saw and heard the speaker producing one of the disyllables. After that, a black screen with boxes representing the response choices (/alda/, /arda/, /alga/, /arga/) appeared, and participants clicked the computer's mouse in the box representing the disyllable that they heard. We instructed participants to guess if they were not sure of the identification of either consonant.

Results

Figures 3A and 3B present the results of Experiments 3a and 3b, respectively. In Experiment 3a, we obtained the effect of precursor

⁴ We thank Justin Bates for serving as our speaker.

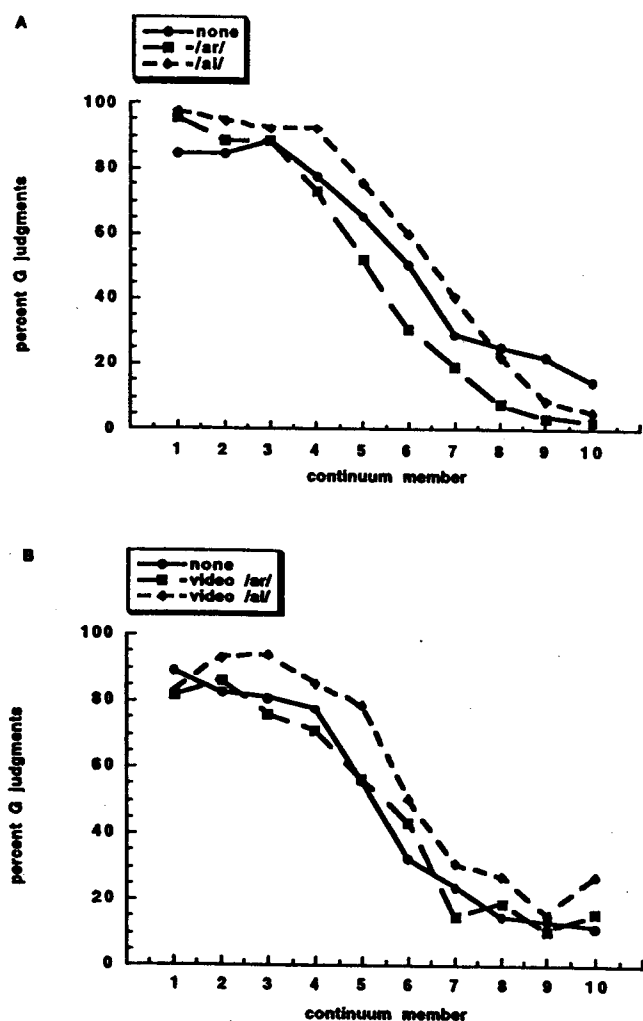


Figure 3. Percentage of G judgments given to the synthetic syllables of Experiment 2b heard in the context of acoustic /a/ or /ar/ or in isolation (A) or heard in the context of audiovisual /a/ or /ar/ or in isolation (B).

syllable first reported by Mann (1980). That is, participants gave more G responses in syllables following /a/ than /ar/. The overall percentages of G responses to CVs after /a/, /ar/, and in isolation were 59.0%, 46.2%, and 54.3% respectively. In an ANOVA, the effect of precursor syllable was significant, $F(2, 24) = 6.30$, $p = .006$. Paired t tests showed that there were significantly more G responses following /a/ than /ar/, $t(12) = 3.57$, $p = .0004$, and more G responses in the no-precursor condition than after /ar/, $t(12) = 2.70$, $p = .02$. In addition, there was a highly significant effect of continuum member, $F(9, 108) = 112.11$, $p < .0001$, and a significant interaction, $F(18, 216) = 2.37$, $p = .0019$. The interaction most likely reflects the crossing of the responses to the isolated syllable condition over the other two curves at the endpoints of the continuum. Because this condition was the first one that listeners took, the cleaner performance at the endpoints in the /a/ and /ar/ conditions is likely a practice effect. A second contributor to the interaction may be the separation of the /a/ and /ar/ curves except at continuum endpoints.

In Experiment 3b, accuracy identifying /a/ and /ar/ was quite good, averaging 93.4% for /a/ and 84.0% for /ar/. We scored responses to the continuum members contingent on the participants' accurate identification of the precursor syllables.⁵ This performance, shown in Figure 3B, was somewhat noisier than that of participants in Experiment 3a, but otherwise quite similar. Overall percentages of G responses in the /a/, /ar/, and isolated syllable conditions were 58.4%, 47.3%, and 48.1%, respectively. In the ANOVA, the main effect of precursor syllable was significant, $F(2, 24) = 8.78$, $p = .0014$, with all differences significant, in paired t tests, except between /ar/ and the isolated syllable condition, /a/ versus /ar/, $t(12) = 3.87$, $p = .0002$; and /a/ versus isolated syllable, $t(12) = 3.22$, $p = .007$. In addition, the main effect of continuum member was significant, $F(9, 108) = 86.77$, $p < .0001$; the interaction did not reach significance.

The overall precursor effect (that is, percentage G responses in the /a/ versus /ar/ contexts) was 12.8% in Experiment 3a and 11.1% in Experiment 3b. In an ANOVA with experiment as a factor, the main effect of experiment and its interaction with precursor were both nonsignificant (both F s < 1).

We performed a final analysis to determine whether responses in Experiment 3b might have been due to some kind of response bias. That is, listeners may have given more G responses to syllables in the McGurk /a/ than /ar/ context, because they somehow knew that ambiguous syllables are more likely to be /ga/ following /l/ than /r/. To make this determination, we looked at responses on the trials that had been excluded from the original analysis of Experiment 3b. That is, we looked at G responses on trials on which listeners had misidentified the precursor consonant. If listeners were aware that ambiguous syllables are more likely to be /ga/ after /a/ than /ar/, then we should have seen more G responses after syllables incorrectly identified as /a/ than after syllables incorrectly identified as /ar/. In fact, the percentages of G responses were statistically the same, $t(12) = 1.37$, $p = .20$, after /a/ incorrectly identified as /ar/ (62.5%) and after /ar/ incorrectly identified as /a/ (51.4%). Most likely, these responses tended to occur on trials on which, for some reason (e.g., inattention), the video information was ineffective. Accordingly, the effect of the precursor was whatever the effect of our acoustically ambiguous syllable alone was on perception of /da/ and /ga/. However, it is interesting that the numerical trend is consistent with the consonant signaled by the video clip, not with the consonant that observers identified. This trend is similar to Mann's (1986) more reliable finding that Japanese listeners who are unable to identify /h/ and /l/ systematically nonetheless compensate for coarticulation.

Discussion

In Experiment 3, we first replicated the finding of Mann (1980) that syllables ambiguous between /da/ and /ga/ are more frequently

⁵ Given Mann's (1986) findings that Japanese listeners show differential compensation for coarticulation even when they are at chance at distinguishing /h/ from /l/, this contingent scoring of D and G responses should be unnecessary. Accordingly, we also scored D and G responses on all trials regardless of accuracy in identifying L and R. It is not surprising, given the high performance on L and R identifications, that the results were qualitatively and statistically the same as in the contingent analysis.

identified as /ga/ following /a/ than following /ar/. We ran the replication to ensure that responses to our continuum syllables, with their too-intense F3s, which had resisted masking in Experiment 2, would show the same evidence of compensation for coarticulation that responses to other versions of the stimuli have shown. They did, and that allowed us to go on to ask whether these same syllables would show effects of precursor syllables whose identities as /a/ and /ar/ were determined by the visible articulatory gestures of a speaker rather than by the acoustic speech signal. Such a finding would rule out masking and any other auditory contrastive effect as the origin of compensation for coarticulation in these stimuli. That was our finding.

A proponent of contrast accounts might suggest, even so, that our results are due to contrast. There may have been a contrastive effect of the visible /a/ and /ar/ on the visible place of articulation of the following CV. However, we can rule this out as being responsible for the patterning of data that we obtained. Because /ar/ is lip rounded, /ar/, which has the more back major constriction, was visibly labial. The alveolar constriction behind the lips and teeth for /a/ was visible during its production. Any contrast effect, therefore, would have worked backward to explain the data. More *G* responses occurred in the context of the more front, but visibly more back, consonant than in the context of the more back, but visibly more front, consonant.

In the introduction, we summarized recent findings that compensation for coarticulation sometimes occurs due to information that is superordinate to the phonetic level of description of speech utterances. Pitt and McQueen (1998) have shown compensation due to knowledge of transition probabilities between phonetic segments. Perhaps this is what underlies our findings if contrast does not. That is, perhaps more *G* judgments follow /a/ than /ar/ because /g/ is more likely after /l/ than is /d/ and because /d/ is more likely after /r/ than is /g/.

To address the question more directly, we searched a computer lexicon of approximately 24,000 words. We searched for all occurrences of /l/ (and separately /r/) in the phonetic transcription of words that occurred in prefinal position in the word. Next, we searched those subsets of the lexicon for occurrences of following /g/ or /d/ either in the same syllable or across a syllable boundary. We found 5,833 occurrences of /l/ and 7,504 of /r/. The conditional probabilities of /g/ given /l/ and of /d/ given /l/ were .0026 and .0273, respectively. The probabilities of /g/ given /r/ and of /d/ given /r/ were .0015 and .0219, respectively. That is, /g/ is much less likely than /d/ in both contexts. If we subtract the probabilities in each pair (/d/ minus /g/), we get .0247 in the context of /l/ and .0204 in the context of /r/. These differences indicate the degree to which /g/ is dispreferred in the contexts of /l/ and /r/. Even though the probabilities are very low, and even though /g/ is less frequent than /d/ in both contexts, is /g/ less infrequent after /l/ than /r/? The answer is no; /g/ is relatively more infrequent than /d/ after /l/ than after /r/. Knowledge of transition probabilities does not underlie the findings of Experiment 3b.

We conclude that, in Experiment 3b, compensation for coarticulation did not occur due to any kind of auditory or optical contrast effect, and it did not occur due to superordinate knowledge of the relative frequencies of different phoneme sequences in the language. It occurred when perceivers used information about coarticulation at the same level of description of a speech utterance at which speakers coarticulate. As Lotto et al. (1997, p. 1134) point

out, there is "remarkable symmetry between perception and production" of speech.

General Discussion

Experiment 1 disconfirmed a prediction that we derived from a frequency contrast account of compensation for coarticulation. We found no effect of the frequency of a precursor tone on pitch judgments of a following tone. We assume that we did not get the same effects as Cathcart and Dawson (1928–1929) and Christman (1954) due to the substantial differences in stimulus materials and methods across the studies. Our stimulus materials were selected to mimic the durations and amplitude relations of the /a/, /ar/, /da/, and /ga/ syllables invoked by Lotto and Kluender (1998) in their frequency contrast account. Our methods were selected to mimic those of Lotto and Kluender.

Experiment 2 replicated Lotto and Kluender's finding of contrastive effects of precursor tones on /da/-/ga/ judgments. It extended those findings by showing that the relative amplitudes of the precursor tones and the relevant part of the stop-vowel syllables (F3) are likely to be important in determining whether the contrastive effect occurs. This suggests that the contrast effect is caused by masking. It remains to be determined whether the relative amplitudes of relevant components of /a/ and /ar/ versus /da/ and /ga/ that are produced in natural speech are the same as those required for masking to occur outside the laboratory. As we point out in Footnote 3, there is already reason to doubt that masking underlies the compensation for coarticulation observed by Mann (1980), Mann and Liberman (1983), and our Experiment 3a.

Experiment 3 was designed to eliminate the possibility of auditory masking by using the McGurk effect to determine the identity of the precursor syllables. We found no reduction in the effect of /a/ and /ar/ on /da/-/ga/ judgments when the information distinguishing them was optical rather than acoustic. In addition, we ruled out accounts of the findings in terms of optical contrast effects or English listener's knowledge of transition probabilities between segments in words. In Experiment 3, we infer that compensation for coarticulation was due to perceivers' use of phonetic gestural information in the audiovisual stimuli as phonetic gestural information.

In view of these findings of Experiment 3, it is appropriate to ask why quail show response patterns that are qualitatively like those in Figure 3 when they receive acoustic disyllables like those that are used in Mann's (1980) study and in our own. Are they compensating for coarticulation? Although one of us (Fowler, 1996) has argued that nonhuman animals may well perceive vocal tract actions from acoustic speech signals, it strains even her credulity to suppose that their perception is sensitive enough to support compensation for coarticulation. For the present, we suppose that quail showed the response patterns that they did because masking occurred. That is, we suppose that the intensity relations between the initial and final syllables of the stimuli used by Lotto et al. (1997) were such that masking occurred. Additional research is required to pin down what those intensity relations have to be in order for masking to occur (particularly, in humans) and to determine what the intensity relations tend to be in natural speech.

As we see it, the disagreement in theoretical viewpoint that underlies the expectations that compensations for coarticulation are due to general auditory contrast effects or that they are due to

perceiving what talkers do is a disagreement about the nature of the perceptual system subserving speech perception. For Lotto and Kluender, perceptual systems are general-purpose devices with built-in resources for handling invariant properties of real-world events; for motor theorists, the speech perception system itself is a special purpose device; and for a theory of direct realism, perceptual systems attune themselves to stimulation so that they are capable of being a variety of special purpose devices (cf. Runeson, 1977). Our results are consistent with either of the latter viewpoints, but not with the first.

This does not mean that we deny that contrast effects occur or that masking with contrastive consequences occurs. Moreover, we are willing to accept that contrast effects may be adaptive in the way that, for example, Warren (1985) describes. We do suggest, however, that contrast effects occur under considerably more limited circumstances than Kluender and his colleagues have supposed. As for masking, it seems to us unlikely to reflect a general-purpose resource for dealing with invariant properties of real-world events and more likely to index limitations on the rate at which discrete stimuli can be processed. In any case, we are willing to leave the investigation of these matters to more expert researchers. Our claim here is only that neither general contrast effects nor masking with contrast as a consequence underlies the tendency of human listeners to compensate for coarticulation. What underlies this tendency, instead, is listeners' use of the structure in acoustic speech signals as information for its gestural causes.

As we noted in the introduction, Warren (1985) suggested that contrast effects occur when perceivers attune to the particulars of the environment in which they are working. An initial boundary between a judgment that a weight is heavy (or a tone is low pitched) or light (high pitched) will shift in the direction of weights (tones) presented at one extreme of the possible set of weights (tones) or the other. This has the effect of sensitizing perceivers to differences in the vicinity of recently encountered stimuli.

Although Lotto and Kluender (1998) allude to this account of contrast without rejecting it, it was, a priori, unlikely to apply to compensation for coarticulation. Warren (1985) points out that these criterion shifts (as he calls them, rather than contrast effects) are highly context sensitive. He gives two illustrative examples of this context sensitivity. In Johnson's (1944) study of contrast effects in weight judgments, if participants were asked to move some books between a pair of trials in which they were to judge the heaviness of two stimulus objects, the weight of the books had no impact on weight judgments. Warren's second set of examples comes from the literature on selective adaptation in speech perception, which he ascribes to criterion shifts. Extensive research in this domain has shown that, if listeners are repeatedly exposed to a syllable, for example, /ta/, they subsequently identify ambiguous syllables along a /da/-/ta/ continuum as /da/ that they identified as /ta/ before adaptation (e.g., Eimas & Corbit, 1973). In this literature, the adaptation effect is found to be highly context sensitive. In research by Cooper (1979), for example, /pae/ was less effective than /bae/ as an adaptor of members of a /bae/ to /dae/ continuum. Effects of VCs on CVs were absent in research by Ades (1974). Accordingly, /al/ and /ar/ would be expected to have no effect on /da/ and /ga/ identifications. Precursor tones should likewise be ineffective.

Lotto et al. (1997) suggest a somewhat different way of thinking about contrast. They suggest that contrast can enhance sensitivity to change. As they point out, things in the world (including in the vocal tract) have mass and exhibit inertia. Accordingly, the state of a system at time t is markedly constrained by the state at time $t - 1$. This leads, for example, to carryover coarticulation effects with assimilatory acoustic consequences. Contrast eliminates or reduces the assimilatory effects, thereby enhancing the change from one phonetic segment to another. Although contrast effects would have that result, we are confident that they do not underlie compensations for coarticulation. As we pointed out in the introduction, this account in terms of contrast does not explain compensation for coarticulation in speech perception. Compensation is bidirectional, occurring in the anticipatory coarticulatory direction as well as the carryover direction; but in the account of Lotto et al. (1997) and Lotto and Kluender (1998), contrast is unidirectional. In addition, coarticulatory effects are not eliminated for perceivers, who use them to identify coarticulating segments. Although we have verified that masking effects with contrastive consequences can occur with speech synthesized so that the masking stimulus is sufficiently intense to serve as a masker, the literature provides no evidence suggesting that the required intensity relations are met in natural productions of disyllables in which liquids precede stops. Finally, contrast cannot explain the compensation for coarticulation that occurred in Experiment 3b.

Contrast effects have been invoked frequently by Kluender and colleagues (Diehl & Kluender, 1989; Diehl & Walsh, 1989; Diehl, Walsh, & Kluender, 1991; Kluender, Diehl, & Wright, 1988) to explain characteristics of speech perception. One of us (Fowler, 1991, 1992) has pointed out empirical and theoretical deficiencies of this view as it has been applied to perceivers' use of durational information in speech perception. Our present research reveals comparable deficiencies in accounts of frequency contrast and spectral contrast. In our view, it is time to acknowledge that general contrast effects do not support perception of speech.

References*

- Ades, A. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 16, 61-66.
- Adobe Premier [Computer software]. (1998). San Jose, CA: Adobe Systems, Inc.
- Cathcart, E. P., & Dawson, S. (1928-1929). Persistence (2). *British Journal of Psychology*, 19, 343-356.
- Christman, R. (1954). Shifts in pitch as a function of prolonged stimulation with pure tones. *American Journal of Psychology*, 67, 484-491.
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods and Instrumentation*, 25, 257-271.
- Cooper, W. E. (1979). *Speech perception and production*. Norwood, NJ: Ablex Publishing.
- Daniiloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, 1, 239-248.
- Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- Diehl, R., & Walsh, M. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, 85, 2154-2164.
- Diehl, R., Walsh, M., & Kluender, K. (1991). On the interpretability of

- speech/nonspeech comparisons. *Journal of the Acoustical Society of America*, 89, 2905-2909.
- Eimas, P. D., & Corbit, J. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- Elman, J., & McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143-165.
- Fowler, C. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, 46, 127-139.
- Fowler, C. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, 36, 359-368.
- Fowler, C. A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America*, 89, 2910-2915.
- Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There is no contrast effect here. *Journal of Phonetics*, 20, 143-165.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730-1741.
- Fowler, C. A., Best, C., & McRoberts, G. (1990). Young infants' perception of liquid coarticulatory effects on following stop consonants. *Perception & Psychophysics*, 48, 559-570.
- Fowler, C. A., & Brown, J. B. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, 62, 21-32.
- Fowler, C., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-136). Hillsdale, NJ: Erlbaum.
- Johnson, D. M. (1944). Generalization of a scale of values by the averaging of practice effects. *Journal of Experimental Psychology*, 34, 425-436.
- Klatt, D. (1980). Software for a cascade/parallel format synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.
- Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153-169.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory revised. *Cognition*, 21, 1-36.
- Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602-619.
- Lotto, A., Kluender, K., & Holt, L. (1997). Perceptual compensation for coarticulation by Japanese quail (*coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102, 1134-1140.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407-412.
- Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition*, 24, 169-196.
- Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, 14, 211-235.
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [s]-[ʃ] distinction. *Perception & Psychophysics*, 28, 213-228.
- Mann, V. A., & Soli, S. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics*, 49, 399-411.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes and features. *Psychological Review*, 101, 653-675.
- Martin, J., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, 69, 559-567.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Moore, B. (1988). *An introduction to the psychology of hearing*. London: Academic Press.
- Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.
- Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, 18, 172-179.
- Walden, B., Prosek, R., Montgomery, A., Scherr, C., & Jones, C. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*, 20, 130-145.
- Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, 92, 574-584.
- Whalen, D. H. (1982). *Perceptual effects of phonetic mismatches*. Unpublished doctoral dissertation, Yale University.
- Whalen, D. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, 35, 49-64.

Received February 23, 1999

Revision received September 4, 1999

Accepted September 4, 1999 ■