COMPENSATION FOR COARTICULATION IN AUDIOVISUAL SPEECH PERCEPTION

Carol A. Fowler*, Julie M. Brown*, and Virginia A. Mann+

*University of Connecticut, +University of California, Irvine, USA

ABSTRACT

Recently the claim that listeners compensate for coarticulation by closely tracking information about the talker’s coarticulatory behavior has been challenged by findings of compensation-like responses to speech on the part of Japanese quail and by findings of compensation-like responses of humans to tonal precursors of syllables. These findings led to a conclusion that apparent compensations for coarticulation reflect a low-level frequency contrast effect. We disconfirm that account in a first experiment and then confirm that precursor tones, if sufficiently intense, can mask following syllables giving rise to compensation-like response patterns. However, we rule out that account and any other account in terms of auditory contrast effects as underlying compensation for coarticulation generally by showing compensation for coarticulatory effects of precursor syllables when information for the coarticulatory effects is optical.

1. INTRODUCTION

Listeners appear to be remarkably sensitive to what talkers do to produce coarticulated speech. One index of their sensitivity is their compensation for coarticulation. A well-known example of compensation was provided by Mann [7] She presented members of a /da/-/ga/ continuum in the context of a preceding /al/ or /ar/ syllable. Listeners identified more continuum members as /ga/ in the context of /al/ than of /ar/. In natural speech, the alveolar constriction of /l/ may exert a carryover coarticulatory effect on a following /g/, pulling its place of articulation forward. /l/ may pull the place of articulation of /d/ back. Listeners behave as if they are taking that into account perceptually, allowing more front members of the /da/-/ga/ continuum to count as /ga/ in the context of /l/ than of /r/. If this account of listeners’ behavior is accurate, then listeners are adept at using phonetic information for coarticulatory influences of one phonetic segment on another as information regulating their perceptual identifications.

The literature offers two other ways in which apparent compensations for coarticulation may occur. Neither of these implies that listeners are particularly sensitive to talkers’ coarticulatory behaviors. Elman and McClelland [3] and Pitt and McQueen [10] have shown that compensation-like effects of fricatives on a following /d/-/g/ continuum can be due, not to phonetic or acoustic evidence of coarticulation, but rather, to listeners’ knowledge of their language. In the research by Elman and McClelland [3] listeners compensated for effects of /s/ and /S/ on following /d/-/g/ continua when the information for the identity of the fricative was lexical not acoustic (e.g., the final consonants of Christmas and foolish having been made acoustically identical. Recently, Pitt and McQueen [10] have shown that, in fact, the effective information in the research of Elman and McClelland was not lexical, but knowledge of the transition probabilities between successive phonemes. (In English, /s/ is more probable after the final vowel in Christmas; /S/ is more probable after the final vowel of foolish.

Recently, Lotto and Kluender [5] (see as well, [6] have challenged the claim that compensation for coarticulation ever needs to be ascribed to any remarkable sensitivity on the part of listeners to talkers’ coarticulatory behavior. They ascribed the apparent compensatory behavior in Mann’s [7] research, and by implication in any research findings of compensation, to frequency contrast. Mann’s /da/-/ga/ continuum members differed in F3; the /da/ endpoint had a falling F3 and the /ga/ endpoint a rising F3. The /al/ precursor syllable had a higher terminal F3 than the onset F3 of the /da/ endpoint; the /ar/ precursor had a lower terminal F3 than the onset F3 of the /ga/ endpoint. A frequency contrast effect of /al/’s terminal F3 would make following onset F3s perceptually lower than they are; a contrastive effect of /ar/ would make following onset F3s perceptually higher. This would foster more /ga/ identifications after /al/ than /ar/. Lotto and Kluender tested the frequency contrast account by substituting tones for the precursor syllables of Mann [7]. A high frequency precursor tone was set to the terminal center frequency of F3 of /al/; a low frequency precursor tone was set to the terminal center frequency of /ar/’s F3. The high and low frequency precursor tones affected /da/ and /ga/ identifications in the same way that /al/ and /ar/ had, respectively, in the research by Mann [7]. Additionally, Lotto, et al [6] showed that Japanese quail exhibit qualitatively the same effects of /al/ and /ar/ on /da/-/ga/ identifications as humans. Show Japanese quail are not likely to be highly sensitive to coarticulation by humans; a contrast account of their behavior is plausible.

Lotto, et al [6] suggest that contrast effects reflect the adaptation of perceptual systems to invariant properties of the world, namely that events exhibit inertia. In inertial systems the state of the system at time t is constrained by its state at t-1. Contrast effects remove the effects of the state at t-1 on the state at t. In production of /al/ga/, carryover coarticulation may be viewed as inertial; /l/ continues to be produced as /g/ is initiated. Contrast effects remove the /l/ coloring of /g/.

This account of contrast effects provided the starting point for our further investigation of sources of compensations for coarticulation. One flaw in the account is that it explains compensation for carryover coarticulation, but not compensation for anticipatory coarticulation, which is observed to occur (e.g., [8]). A second flaw is that it suggests perceptual elimination of coarticulatory effects. However, the effects are not eliminated; listeners use information in the domain of one segment that is due to coarticulation by another segment as information for the coarticulating segment [e.g., 3]. Perception does not eliminate coarticulatory effects; it properly ascribes them to the coarticulating segment. A final weakness is that the occurrence of frequency contrast effects with stimuli like those of Lotto and Kluender [5] has not been established. Whereas Lotto and Kluender cited two papers [1, 2] that do, indeed, report frequency contrast, the contrastive findings were obtained with stimuli nothing like those of Lotto and Kluender. For example, precursor tones in [2] lasted either one or two minutes (i.e., 60,000 or 120,000 ms), not the 250 ms of Lotto and Kluender’s tones.

page 639 ICPPhS99 San Francisco
2. EXPERIMENTS

2.1 Experiment 1

2.1.1 Introduction. In the first experiment, we looked for a frequency contrast effect using tonal precursors and target tones.

2.1.2 Method. We synthesized two sets of target tones. In one set, the tones were 250 ms long (the duration of /da/-/ga/ syllables in research by Lotto and Kluender [5] in the other set they were 50 ms long (the duration of /da/-/ga/ transitions). Tones were steady state sinewaves with frequencies between 1800 and 2700 Hz. They had 5 ms amplitude rise times and fall offs at the beginnings and ends. There were ten tones in all, with a 100 Hz separation between tones. High frequency (2800 Hz) and low frequency (1700 Hz) precursor tones were synthesized as well. They were 250 ms in duration. Target tones followed test tones after a 50 ms silent interval.

Sixteen listeners participated in the condition with 250 ms target tones; 15 participated in the test with 50 ms target tones. In both groups, listeners first learned to identify the tone-continuum endpoints as H (high) or L. They then took a 100 item test in which all continuum members were presented in isolation ten times each and identified each tone as H or L. Finally, they took a 200 item test in which 20 tokens of each continuum tone were presented preceded ten times by the high precursor tone and ten times by the low precursor. They identified the target tones as H or L.

2.1.3 Results. The results shown in Figures 1 and 2 revealed no evidence of a frequency contrast effect. Precursor tones reduced identification consistency at the continuum endpoints, but there was no tendency for the high precursor to increase L judgments or for the low precursor to increase H judgments. In analyses of variance performed separately on the data from the 250 ms target tones and the 50 ms tones, effects of the precursor were nonsignificant (both Fs < 1).

2.2 Experiment 2

2.2.1 Method. We used our precursor tones from Experiment 1, but matched them in RMS amplitude to target /da/-/ga/ stimuli, following the procedure of Lotto and Kluender [5] We used two sets of /da/-/ga/ stimuli. One set consisted of the syllables used by Lotto and Kluender. The other consisted of syllables that we synthesized to match the parameters of their syllables, but we increased the intensity of F3. Listeners (18 listened to the Lotto-Kluender syllables; 13 listened to the other set) first learned to identify the endpoints of the /da/-/ga/ continuum. Then they identified ten tokens of each of the continuum members. Finally they identified 20 tokens of each of the continuum members, half preceded by the high precursor tone and half preceded by the low precursor.

2.2.2 Results and Discussion. The results of the final test are shown in Figures 3 and 4. We obtained a clear contrastive effect of the precursor tones on the Lotto-Kluender stimuli (F(1, 17) =
35.65, p < .0001), but no evidence of an effect on the stimuli with enhanced F3 intensity (F<1). We infer that the effect of precursor tones is likely a masking effect due to the relative intensities of the precursor tones and the F3 transition for the /da/-/ga/ stimuli. It is relevant that it is presently unknown whether the relative intensities of F3 in /al/ and /ar/ in Mann’s (1980) research are such that masking can account for apparent compensations for coarticulation that she observed. In addition, it is presently unknown whether, in nature, the relative intensities of syllables for which listeners appear to compensate for coarticulation are such that a masking account is feasible.

2.3 Experiment 3

In the final experiment, we asked whether compensation for coarticulation might occur under conditions in which neither frequency contrast nor masking accounts of compensation should apply. In particular, we used an acoustic precursor syllable that was ambiguous between /al/ and /ar/ and disambiguated it by dubbing it onto a face mouthing /al/ or /ar/ [9]. In this way, information distinguishing /al/ from /ar/ is optical, not acoustic.

2.3.1 Method. We synthesized three precursor syllables, /al/ and /ar/ using synthesis parameters of Lotto and Kluender [5] and a syllable ambiguous between /al/ and /ar/. One condition of our experiment served as a replication of Mann’s [7] original demonstration, using our /da/-/ga/ continuum with its enhanced F3 intensity. This was to ensure that, despite not seeing evidence of masking of these continuum members in Experiment 2, we nonetheless would see compensation for coarticulation. In this condition, listeners (N=13) were given experience with continuum endpoints, then identified continuum members presented in isolation, and finally identified the continuum members presented in the context of the precursor /al/ and /ar/ syllables. Numbers of trials and other aspects of the procedure were as in Experiment 2. For the audiovisual condition, we videotaped a male speaker producing the disyllables /alda/, /arda/, /alga/ and /arga/. We asked the speaker to hyperarticulate, emphasizing the alveolar constriction of /al/ and the lip rounding of /ar/. We chose video clips of /alda/ and /arda/ for the experiment. (A video only test showed that observers identified the second consonant as /da/ equally often from the two video clips.) We dubbed our ambiguous “al/r” syllable followed after 50 ms by each of the ten continuum members onto each video clip. In this condition, in an audio alone test, listeners (N=13) first learned to identify the continuum endpoints and then they identified each continuum member presented in isolation. Finally, they viewed and listened to a 200 item audiovisual test in which each continuum member occurred ten times dubbed following the ambiguous “al/r” syllable, onto the /alda/ video clip and ten times dubbed onto the /arda/ video clip. In this test, participants identified both consonants on each trial.

2.3.2 Results. Results are presented in Figures 5 and 6. Figure 5 presents the findings from the replication of Mann’s [7] experiment. Despite not finding spectral contrast in Experiment 2, we did find compensation for coarticulation in this condition of Experiment 3. The effect of precursor syllable was highly significant (F(1,12) = 12.73, p = .004). In the audiovisual condition, we scored the proportion of /ga/ responses contingent on accuracy reporting the first consonant as /l/ or /r/. Accuracy identifying /l/ was 93.4%; accuracy on /r/ was 84.0%. Figure 4 shows that observers compensated for coarticulation even though the information supporting compensation was optical rather than acoustic. The effect of precursor syllable was highly significant (F(1,12) = 14.96, p = .002).
the findings of Experiment 3. We searched a lexicon of approximately 24,000 words to find out whether /g/ is more probable following /l/ than is /d/ and whether /d/ is more probably following /r/ than is /g/. In short, in both the /l/ and /r/ contexts, /d/ is considerably more probable than is /g/. It is more probable than /ga/ to a slightly greater degree following /l/ than /r/. Accordingly, our findings of compensation are not due to information about transition probabilities between liquids and following /d/ and /g/. We conclude that compensation for coarticulation occurred in Experiment 3 because listener/observers used information about coarticulation in the sequence of phonetic segments as information for coarticulation.

3. CONCLUSION

We ascribe human listeners’ performance in Lotto and Kluender’s [5] research, and Japanese quail’s performance in [6], to masking. It does not follow, however, that compensation for coarticulation by humans outside the laboratory is also due to spectral contrast (i.e., masking). It remains to be determined whether intensity relations between acoustic information for a coarticulating segment and relevant parts of an influenced segment in nature and in other research are such that masking would occur. In any case, it is unlikely that spectral contrast underlies the findings of compensation generally. As noted, Lotto and Kluender’s contrast account implies unidirectional compensation for coarticulation, that is compensation for carryover coarticulation only, but compensation occurs as well for anticipatory coarticulation (e.g., [8]). Moreover, the account in terms of spectral contrast implies that coarticulatory information is eliminated perceptually, but it is well known not to be. Listeners use coarticulatory information about a segment in the domain of another segment as information for the coarticulating segment [e.g., 11]. Compensation merely shows that the coarticulatory information is properly ascribed by listeners to the coarticulating segment not the influenced, segment

ACKNOWLEDGMENTS

The research was supported by grant HD-01994 to Haskins Laboratories.

NOTES

1. We thank Andrew Lotto for supplying the acoustic stimuli for one condition of Experiment 2.

REFERENCES