

GENERAL AUDITORY PROCESSES MAY ACCOUNT FOR THE EFFECT OF PRECEDING LIQUID ON PERCEPTION OF PLACE OF ARTICULATION.

Andrew J. Lotto and Keith R. Kluender

Dept. of Psychology, University of Wisconsin-Madison, USA

ABSTRACT

The perception of syllable-initial /d/ and /g/ can be affected by the composition of the preceding syllable [1]. It has been suggested that this result demonstrates the existence of a mechanism which compensates for coarticulation. In a series of perceptual experiments, including the use of an avian species, this effect is shown to be due to general auditory processes which may be described as frequency contrast.

I. INTRODUCTION

Immediate context plays an enormous role in perception of acoustic information specifying phonemes. Many context effects in speech perception have the appearance of serving as compensation for effects of coarticulation. For example, perception of syllable-initial /d/ and /g/ can be affected by the composition of preceding acoustic information such that, for a series of synthesized consonant-vowel syllables (CVs) varying in onset characteristics of the third formant (F3) and varying perceptually from /da/ to /ga/, subjects are more likely to perceive /da/ when preceded by the syllable /ar/, and to perceive /ga/ when preceded by /al/ [1]. This effect has been found for speakers of Japanese who cannot distinguish between /l/ and /r/ [2] and for prelinguistic infants [3]. The received interpretation of these findings has been that listeners are somehow sensitive to articulatory implementation. The following experiments assess the degree to which these perceptual effects are specific to qualities of human articulatory sources.

II. EXPERIMENT 1

Others have concluded that the perceptual effect described above results from a mechanism specialized to compensate for vocal tract constraints through the use of tacit knowledge of articulatory dynamics [1], or that the effect is due to the recovery of vocal tract actions [3]. The first experiment was performed to examine the extent to which perceptual effects of coarticulation are dependent upon maintaining a unitary source. Presumably a mechanism serving to accommodate coarticulation would be specific to the speech stream of a single speaker. After all, the perception of a single talker's speech often must take place in the presence of other acoustic sources including other talkers.

Stimuli.

A 10-step series of /da-ga/ syllables varying in F3 onset frequency was synthesized with endpoint stimuli based on the natural productions of a male talker. For these CVs, the onset frequency of F3 varied from 1800 to 2700 Hz in 100 Hz steps and changed linearly over 80 msec to a steady state value of 2450 Hz. All other parameters remained unchanged between members of the series. The first formant rose from 300 to 750 Hz and F2 decreased from 1650 to 1200 Hz over 80 msec. Fundamental frequency was 110 Hz from onset until decreasing to 95 Hz over the last 50 msec. Total stimulus duration of these synthesized syllables was 250 ms. These stimuli were preceded by two separate natural speech versions of preceding /ar/ and /al/. One version was produced by a relatively tall adult male talker (190 cm in height) after

whom the /da-ga/ series was modelled with an average f_0 of 110 Hz. The other was produced by a relatively short female talker (157 cm) with an average f_0 of 210 Hz. Formant frequency values averaged about 12% higher for the female VCs.

Procedure.

Twelve English-speaking subjects participated in a two-choice forced identification task. All stimuli were passed through a 16-bit D/A at a sampling rate of 10 kHz, low-pass filtered at 4.8 kHz, amplified, and randomly presented under the control of a microcomputer over headphones at 75 dB SPL. Listeners responded after each disyllable by pressing either of two buttons labelled 'd' and 'g'.

Results.

Identification functions depicting results are displayed in Figure 1. In addition to replicating the original finding for disyllabic stimuli for which both syllables are modeled after the same talker [1], we found that the effect extends to disyllabic stimuli for which there is a clear mismatch between sources for the first (♀) and second (♂) syllable. It appears that the context effect of preceding /al/ and /ar/ is not critically sensitive to the entire stimulus complex being produced by a single talker or even by a modestly similar vocal tract.

III. EXPERIMENT 2

Because results from the first experiment indicated that precise matching of articulatory/acoustic characteristics for the initial VC and following CV was not essential for the effect, Experiment 2 tested whether simple schematized nonspeech versions of F3 information alone could affect labeling of following CVs.

Stimuli.

Two sine-wave glides were synthesized. Each was matched in frequency characteristics to the F3

transitions of the /al/ or /ar/ syllables of Experiment 1 and were matched in amplitude to the energy within a critical band of the center frequency of F3. The same series of synthetic /da-ga/ syllables used in the Experiment 1 followed the sine-wave glides.

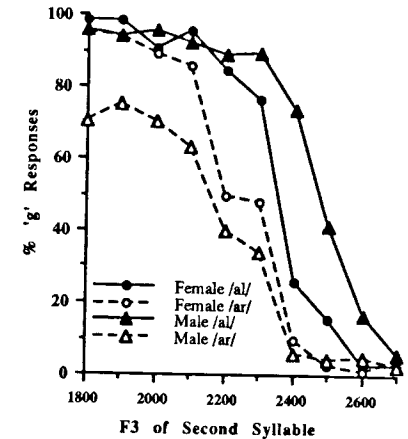


Figure 1. Identification functions from Experiment 1

Procedure.

The procedure was identical to that of Experiment 1. Thirteen subjects identified the consonant as 'd' or 'g'.

Results.

Identification functions depicting results are displayed in Figure 2. Although the context effect of sine-wave glides was not as great as the effect for full spectrum speech /al/ and /ar/, there is a significant effect of glide type as reflected in responses to the synthesized /da-ga/ series. Even a preceding stimulus that consists of only a sine-wave caricature of a portion of rich full-spectrum speech (F3) is adequate to give rise to the context effect on perception of the following CV as /da/ or /ga/.

IV. EXPERIMENT 3

Conceivably, the sine-wave glides in the second experiment may have resembled speech sufficiently to activate

some special speech mechanism. To address this possibility, simpler constant-frequency tones were used as precursor stimuli for this experiment.

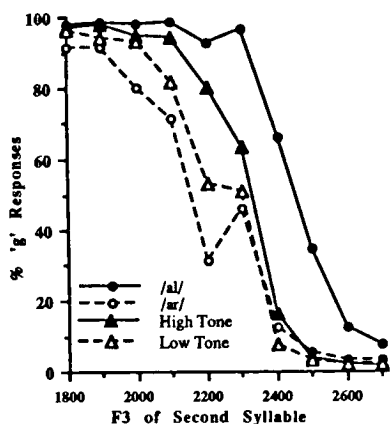


Figure 2. Identification functions for Experiment 2

Stimuli.

Two constant-frequency sine waves with frequencies equal to the offset frequency of F3 for the /al/ and /ar/ speech stimuli were used as precursor stimuli. Sine-wave amplitude was set to the RMS amplitude of the full-syllable male /al/ and /ar/ from Experiment 1. In contrast to the nonspeech stimuli of Experiment 2, these stimuli were not matched to the speech stimuli either in frequency change over time or in amplitude as a function of frequency. The same series of synthetic CV syllables used in Experiments 1 and 2 and varying perceptually from /da/ to /ga/ followed the sine-wave tones.

Procedure.

As in the preceding experiments, subjects (16) responded by pressing 'd' or 'g'.

Results.

Identification functions depicting the results are displayed in Figure 3. The difference in frequency of preceding

tones resulted in a shift in identification boundaries despite the fact that the only characteristic tones shared with /al/ and /ar/ was that they contained substantial energy in the region of F3. Although the context effect of constant-frequency sine waves was not as great as the effect for full spectrum speech /al/ and /ar/ in this experiment, the effect of these nonspeech precursors is of comparable magnitude to Mann's [1] for natural speech stimuli.

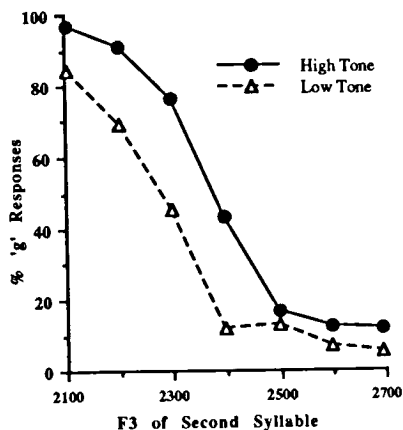


Figure 3. Identification functions for Experiment 3.

V. EXPERIMENT 4

Results from the first three experiments indicate that the original effect of preceding /al/ and /ar/ on perception of following /da/ or /ga/ generalized quite broadly. These findings suggest that the effect could potentially be general to auditory perception. If this is so, similar results should arise when nonhuman animals serve as subjects. In the last experiment, Japanese Quail (*Coturnix japonica*) were used to test the generality of this effect.

Procedure.

Two Japanese Quail were trained (by operant procedures) to peck a lighted key when presented with either the syllable /da/ or /ga/ and to refrain from

pecking it when presented with the alternative syllable (/ga/ or /da/, respectively). After the birds learned to reliably peck differentially to /da/ or /ga/, they were presented with test syllable pairs consisting of the synthesized /al/ or /ar/ followed by one of the ambiguous intermediary members of the /da/-/ga/ series.

Results.

Histograms representing the results of this study are shown in Figure 4. Avian responses evidenced an effect of the preceding syllable such that 'labeling' shifted to more /ga/ responses following /al/. The most parsimonious explanation for this result is that the effect is due to general auditory processes of frequency contrast.

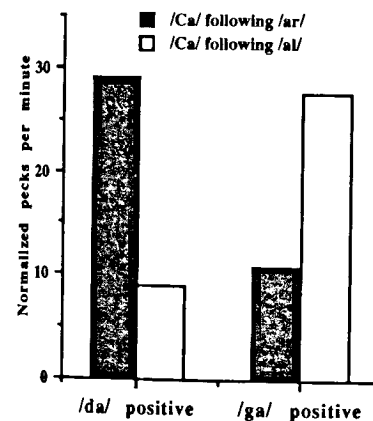


Figure 4. Histograms of Normalized Peck Rates from Experiment 4.

DISCUSSION

Results from these four experiments suggest that this context effect is not specific to speech in any conspicuous way. At least part of the original effect is likely due to some general property of the auditory system.

One potentially could describe this hypothesized general auditory effect as frequency contrast. The data from these experiments are consistent with this description. Following a high tone or

glide, the identification boundary shifts toward higher F3 onsets (more are perceived as low, "ga") relative to when the preceding tone is lower.

At this time, the term is not meant to suffice as an explanation or to designate a specific process. The point is simply that the effect reported first in Mann [1] and replicated here may be due, at least in part, to general auditory processes. If the effect is of a general nature, then the earlier findings with Japanese listeners [2] and with infants [3] should not be unexpected.

In general, contrast may be useful for many of the acoustic transformations arising from coarticulation. It appears that all coarticulation in speech results in frequency assimilation. Consequently, contrastive perceptual effects between contiguous speech sounds should generally be adaptive.

V. REFERENCES

- [1] Mann, V. A. (1980), "Influence of preceding liquid on stop consonant perception", *Perception and Psychophysics*, vol. 28, pp. 211-235.
- [2] Mann, V. A. (1986), "Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English 'l' and 'r'", *Cognition*, vol. 24, pp. 169-196.
- [3] Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990), "Young infants' perception of liquid coarticulatory influences on following stop consonants", *Perception & Psychophysics*, vol. 48, pp. 559-570.

VI. ACKNOWLEDGMENTS

This work was supported by NIDCD Grant DC-00719 and NSF Young Investigator Award DBS-9258482 to the second author.