Preceding phonetic context affects perception of nonspeech[a] (L)

Joseph D. W. Stephens[b] and Lori L. Holt
Psychology Department and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 10 December 2002; revised 19 September 2003; accepted 30 September 2003)

A discrimination paradigm was used to detect the influence of phonetic context on speech (experiment 1a) and nonspeech (experiment 1b) stimuli. Results of experiment 1a were consistent with the previously observed phonetic context effect of liquid consonants (/l/ and /r/) on subsequent stop consonant (/g/ and /d/) perception. Experiment 1b demonstrated a context effect of liquid consonants on subsequent nonspeech sounds that were spectrally similar to the stop consonants. The results are consistent with findings that implicate spectral contrast in phonetic context effects.

[DOI: 10.1121/1.1627837]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Pc [PA] Pages: 3036–3039

I. INTRODUCTION

Perception of speech is highly dependent on surrounding phonetic context. For example, Mann (1980) found that ambiguous consonant–vowel (CV) syllables varying perceptually between /ga/ and /da/ were identified as /ga/ more often when preceded by /al/ than when preceded by /ar/. With this effect, and other phonetic context effects (e.g., Mann and Repp, 1980, 1981), speech identification is shifted in a direction opposite that of the acoustic assimilation caused by coarticulation in speech production, apparently ''compensating for coarticulation.'' The close correspondence between speech production and perception has led theorists to posit gestural origins for phonetic context effects, such that they arise either from listeners' implicit representations of articulatory gestures (Mann, 1980) or from direct perceptual recovery of articulatory gestures (Fowler *et al.*, 1990). Recent evidence, however, suggests that phonetic context effects may arise from perceptual interactions among the spectral characteristics of adjacent sounds. Lotto and Kluender (1998) found that nonspeech sounds lacking gestural information were sufficient to shift identification of subsequent speech: ambiguous CVs between /ga/ and /da/ were identified more often as /ga/ when preceded by a tone at the third formant (F3) offset frequency of /al/. Lotto *et al.* (1997) observed phonetic context effects in Japanese quail: quail trained to peck to /ga/ pecked more when stimuli were preceded by /al/ than when preceded by /ar/. These findings were interpreted as arising from spectrally contrastive perceptual mechanisms at a precategorical level.

If phonetic context effects arise from general perceptual interactions among spectral characteristics, then both speech and nonspeech sounds should elicit context effects. Nonspeech sounds have been shown to affect perception of subsequent speech (Holt, 1999; Holt *et al.*, 2000; Lotto and Kluender, 1998; Lotto *et al.*, 2003), but to date no effect of speech on subsequent nonspeech has been reported. Moreover, gesture-based theories of speech perception do not predict that such an effect will occur because phonetic context effects arise from information specifying articulatory gestures rather than auditory characteristics (e.g., Fowler *et al.*, 2000). The current study assessed the influence of preceding phonetic context on the perception of nonspeech sounds. Because the nonspeech stimuli were unfamiliar sounds for which participants had no labels, the current experiments used a discrimination paradigm in which labeling was unnecessary (modified from Mann and Liberman, 1983).

II. METHOD

A. Participants

Fifteen and 17 undergraduates at Carnegie Mellon University participated in experiments 1a and 1b, respectively. All reported normal hearing, were native English speakers, and received course credit for participation.

B. Stimuli

Two ten-member series of target stimuli were created for use in experiments 1a and 1b. For experiment 1a, target stimuli were CV syllables ranging perceptually from /ga/ to /da/. Target syllables consisted of 80-ms linear formant transitions followed by a 170-ms steady-state vowel. For experiment 1b, target stimuli were nonspeech sounds consisting only of the 80-ms F2 and F3 transitions from the CV syllables used in experiment 1a. Two syllables, /al/ and /ar/, were synthesized for use as precursors in both experiments. The precursors were 250 ms in duration and consisted of a 100-ms steady-state vowel followed by 150-ms linear formant transitions. Formant frequencies for all stimuli were identical to those used by Lotto and Kluender (1998). Stimuli were synthesized with 12-bit resolution and sampled at 10 kHz, using the cascade branch of the Klatt (1980) synthesizer for speech stimuli and the parallel branch of the synthesizer for nonspeech stimuli.

[a]Portions of this work were presented in ''Effect of preceding speech on nonspeech sound perception,'' at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA, June 2002.
[b]Electronic mail: jds2@andrew.cmu.edu

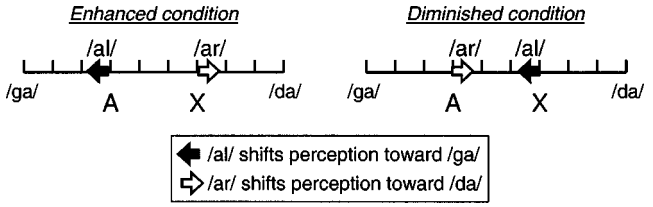

FIG. 1. Design of experiments. On each trial, participants heard two target stimuli and decided whether they were the same or different. Discrimination conditions exploited a known context effect to increase or decrease perceptual distance between target (''enhanced'' and ''diminished'' conditions). In experiment 1b, target syllables from /ga/ to /da/ were replaced by nonspeech sounds that modeled their spectral characteristics.

Each of the target stimuli was matched in root mean square (rms) amplitude and combined with each of the two precursors to yield 40 precursor-target combinations. A silent interval of 50 ms was inserted between precursors and targets, resulting in an overall stimulus duration of 550 ms in experiment 1a and 380 ms in experiment 1b. Stimuli were converted from digital to analog and presented by TDT System II hardware (Tucker-Davis Technologies) over Sennheiser HD-265 linear headphones at 65–70 dB SPL.

C. Procedure

Experiments 1a and 1b each consisted of two phases. In the first phase, participants performed AX discrimination of target stimuli presented in phonetic context. On each trial, two precursor-target combinations were presented with an interstimulus interval of approximately 750 ms. Participants were told to attend to the target stimuli and indicate whether the two targets were different using buttons labeled ''SAME'' and ''DIFFERENT.'' Participants were told that target stimuli would always sound similar and that they should respond ''SAME'' only if they thought the targets were *exactly* the same. On each trial, one target stimulus was preceded by /al/ and the other was preceded by /ar/. Target stimuli either were identical (catch trials) or differed by three steps along the ten-step series (discrimination trials).

The effect of context was tested by comparing two conditions defined by the arrangement of precursors and targets in each trial. In the ''enhanced'' condition, target stimuli with lower F3 onset were preceded by /al/ and target stimuli with higher F3 onset were preceded by /ar/. In the ''diminished'' condition, the opposite arrangement was used. Based on the effect of preceding liquids on stop consonant identification (Mann, 1980), the discrimination of target pairs was expected to be more accurate in the enhanced condition than in the diminished condition. The experimental design is illustrated in Fig. 1. The within-trial order of precursor-target pairs was counterbalanced to yield 28 unique discrimination trials and 20 unique catch trials. All 48 trials were presented in a single block and there were eight repetitions of the trial block, for a total of 384 trials. Order of presentation was random within each trial block. Participants were given a short break half-way through the task.

The second phase of each experiment was an identification task in which participants heard target stimuli one at a time, in isolation, and indicated whether each stimulus sounded like /ga/ or /da/ by pressing buttons labeled ''GA'' and ''DA'' (in experiment 1b participants were asked, ''For each non-speech sound, *if it were a syllable*, would it be /ga/ or /da/?''). All ten targets were presented in each of ten randomly ordered blocks, for a total of 100 trials.

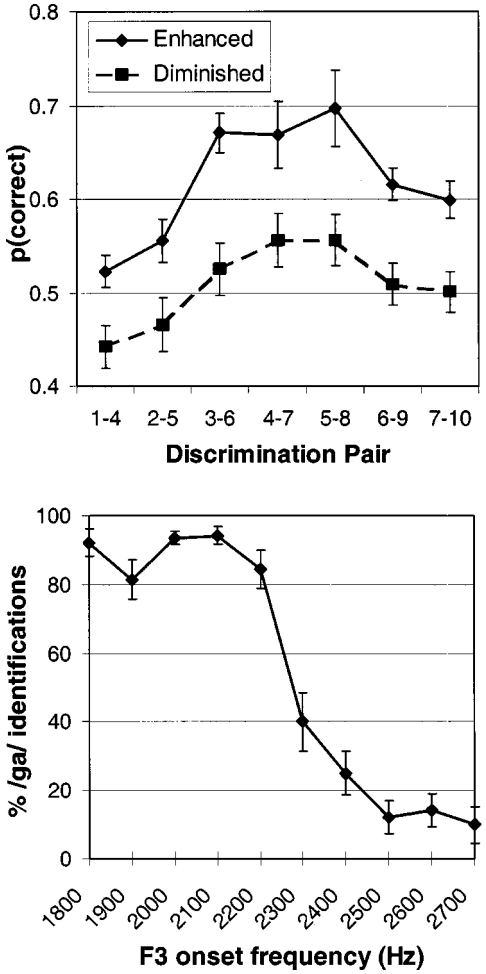

FIG. 2. Results of experiment 1a (speech targets). Mean discrimination (top panel) was better in the enhanced condition than in the diminished condition. Mean identification responses (bottom panel) revealed a typical category boundary along the /ga/-/da/ series. Error bars reflect standard error of the mean.

III. RESULTS

A context effect of preceding liquid on target perception was observed for both speech and nonspeech targets. Data from the identification tasks provided evidence that the target stimuli in experiment 1b were not perceived as speech.

Averaged results of experiments 1a and 1b are shown in Figs. 2 and 3, respectively. Discrimination performance was evaluated by calculating an unbiased measure of proportion correct for each discrimination pair, in each condition, for each participant.[1] Proportion correct was calculated according to Eq. (5.6) of Macmillan and Creelman (1991). Identification responses were evaluated by computing the percentage of trials in which each target stimulus was identified as ''GA.''

For the discrimination task in experiment 1a, an analysis of variance (ANOVA) on the proportion=correct data revealed main effects of enhanced versus diminished condition, $F(1,14)=33.8$, $p<0.001$, and discrimination across tar-

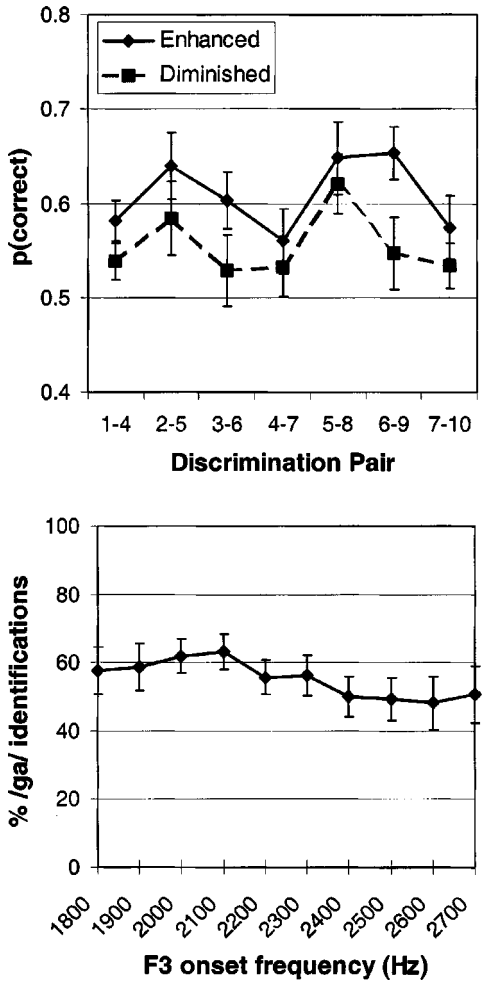

FIG. 3. Results of experiment 1b (nonspeech targets). Mean discrimination (top panel) was better in the enhanced condition than in the diminished condition. Mean identification responses (bottom panel) indicated that participants were unable to correctly assign speech labels to nonspeech stimuli. Error bars reflect standard error of the mean.

get stimulus pair, $F(6,84)=8.05$, $p<0.001$. Target stimulus pairs were more accurately discriminated in the ''enhanced'' condition than in the ''diminished'' condition, indicating that the perceptual distance between target pairs was increased or decreased depending on the arrangement of precursors. Discrimination performance varied significantly across discrimination pairs, with better performance in the middle of the series than at the ends. The identification data from one participant in experiment 1a were discarded due to computer error. Identification data for the remaining participants exhibited a categorical pattern typical of stop-consonant series. An ANOVA on the identification data revealed a significant effect of target stimulus, $F(9,117)=68.5$, $p<0.001$.

The data from one participant in experiment 1b were excluded from analysis due to incorrect execution of the discrimination task (only five ''same'' responses in 384 trials). For the discrimination task in experiment 1b, an ANOVA on the proportion=correct data revealed a significant main effect of enhanced versus diminished condition, $F(1,15)=12.0$, $p<0.005$. Thus, phonetic context effectively influenced discriminability of nonspeech targets. A main effect of discrimination across target stimulus pairs did not reach significance, $F(6,90)=2.16$, $p=0.054$. An ANOVA on the identification data revealed no effect of target stimulus on identification responses, indicating that, on average, participants were unable to assign consistent category labels to the stimuli.

Inspection of identification data from individuals in experiment 1b revealed that some participants *were* able to assign labels to the stimuli. Such response patterns could indicate either that listeners heard the nonspeech stimuli as speech sounds, or that they simply managed to consistently assign arbitrary labels as a function of the F3 onset cue. Therefore, additional analyses were performed to determine whether the results of the discrimination task depended on participants' ability to assign speech labels the nonspeech stimuli. Participants were sorted into three groups: those whose labeling was consistent with the analogous speech categories ($N=7$); those whose labeling was categorical, but opposite to the analogous speech categories ($N=3$); and those whose labeling was not categorical ($N=6$). A participant's identification responses were considered ''categorical'' if the average of his or her ''GA'' responses to the first three members of the stimulus series and the last three members of the stimulus series differed by at least 20 percent. An ANOVA was performed on the discrimination data of experiment 1b, including labeler type as a between-subjects variable. There was no interaction of labeler type and condition, indicating that the effect of phonetic context on nonspeech targets was not related to participants' labels for the nonspeech stimuli. An additional ANOVA was performed on just the discrimination data of the six ''noncategorical'' listeners. An influence of speech precursors upon nonspeech discrimination was observed, $F(1,5)=7.64$, $p=0.04$.

Additional analyses were conducted to compare the results of experiments 1a and 1b. An ANOVA that included data from both experiments revealed a significant interaction of experiment and condition, $F(1,29)=5.54$, $p=0.025$, reflecting the larger effect of context for speech versus nonspeech stimuli. The interaction of experiment and target stimulus pair was also significant, $F(6,174)=3.89$, $p=0.001$, indicating that the pattern of discrimination performance across the target stimulus series differed for speech sounds compared to nonspeech sounds. An ANOVA comparing discrimination data from the seven ''speechlike'' labelers of experiment 1b to discrimination data from experiment 1a also revealed a significant interaction of experiment and target stimulus pair, $F(6,120)=2.84$, $p=0.013$. Thus, even the experiment 1b participants with speechlike identification exhibited a significantly different pattern of discrimination across targets from listeners in experiment 1a.

IV. DISCUSSION

The present results indicate that phonetic context affects the perception of nonspeech sounds, as predicted by a spectral contrast account of phonetic context effects (Lotto and Kluender, 1998). However, some alternative interpretations of the data should be considered.

Fowler *et al.* (2000) proposed that speech and nonspeech context effects originate from different mechanisms, with nonspeech influences upon phonetic perception arising

from masking of the target F3 by nonspeech precursor sounds. Citing Moore (1988), Fowler *et al.* noted 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.'' (p. 881) If masking causes nonspeech context effects, then these effects could be attributed to peripheral auditory mechanisms, which, in the account of Fowler *et al.*, are presumably unrelated to the perceptual mechanisms involved in phonetic context effects. However, due to the construction of stimuli used in experiment 1b, masking is unlikely to be responsible for the current results. The acoustically reduced, nonspeech target stimuli were matched in rms amplitude to the acoustically rich precursor syllables, so that the F3 frequencies of targets had greater energy than the F3 frequencies of precursors. Additionally, recent work has ruled out a peripheral masking account for effects of nonspeech precursors on speech targets. Lotto *et al.* (2003) found effects of preceding nonspeech on consonant perception when precursors and targets were presented to opposite ears, and when delays of up to 175 ms were inserted between precursors and targets. Identical experiments using only speech stimuli (Holt and Lotto, 2002) yielded very similar results, consistent with the hypothesis that common mechanisms underlie both effects.

The interpretation of the current results also depends on the validity of the assumption that the perception of nonspeech sounds involves general auditory mechanisms. Due to the overlearned nature of speech perception, ''speech-specific'' perceptual mechanisms might generalize somewhat to acoustically similar nonspeech sounds. Thus, the context effect observed in experiment 1b might have resulted from speechlike processing of the nonspeech target stimuli. Although most listeners did not label the nonspeech stimuli according to analogous speech categories, the identification task could simply have been less sensitive to subtle effects of speech-specific processing than the discrimination task. However, the pattern of discrimination responses in experiment 1b also provides evidence that speech mechanisms were not involved in the perception of the nonspeech stimuli. Discrimination performance did not vary across the nonspeech stimulus series, whereas discrimination of speech stimuli was significantly better in the middle of the stimulus series than at the ends [consistent with the typical finding of a discrimination peak near a category boundary (Liberman *et al.*, 1957)]. Mann and Liberman (1983) interpreted the absence of a discrimination peak for F3 chirps in duplex perception as evidence that chirps were perceived in a nonspeech ''mode.'' It is therefore reasonable to conclude that the results of experiment 1b are not due to speech-specific processes. A replication of the current results in nonhuman listeners would further support for this conclusion, as it has for other context effects (Lotto *et al.*, 1997).

An additional aspect of interest in the current data is that speech precursors had a greater effect on discrimination of speech targets than nonspeech targets. This difference might result from participants' tendency to rely on category labels when discriminating speech sounds. Due to the presence of a category boundary along the speech target series, an increase or decrease in perceptual distance between speech targets should increase or decrease, respectively, the probability that those targets fall into different categories. Thus, category membership could exaggerate the perceptual context effect for speech stimuli compared to nonspeech stimuli.

In summary, the current findings are consistent with the hypothesis that phonetic context effects result from general perceptual processes sensitive to spectral characteristics. They are more difficult to reconcile with theories (Fowler *et al.*, 1990, 2000) that attribute phonetic context effects to implicit knowledge of articulatory dynamics or direct recovery of articulatory gestures from speech input.

ACKNOWLEDGMENTS

The authors thank Christi Adams, Monica Datta, and Camilla Kydland for help with data collection and Andrew Lotto, Randy Diehl, and James Sawusch for helpful comments. This work was supported by an NDSEG Fellowship to JDWS, by a research grant from NOHR to LLH, and by the Center for the Neural Basis of Cognition.

[1] *d′* was not computed due to the complexity of the ''roving'' same–different design, in which multiple discrimination pairs are presented within a single block (see Chap. 6 of Macmillan and Creelman, 1991). Proportion correct for an unbiased observer is an adequate measure of sensitivity for the purposes of the current study.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). ''Young infants' perception of liquid coarticulatory influences on following stop consonants,'' Percept. Psychophys. **48**, 559–570.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). ''Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans,'' J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Holt, L. L. (**1999**). ''Auditory constraints on speech perception: An examination of spectral contrast,'' unpublished doctoral dissertation, University of Wisconsin, Madison.

Holt, L. L., and Lotto, A. J. (**2002**). ''Behavioral examinations of the level of auditory processing of speech context effects,'' Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). ''The discrimination of speech sounds across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368.

Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). ''Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*),'' J. Acoust. Soc. Am. **102**, 1134–1140.

Lotto, A. J., Sullivan, S. C., and Holt, L. L. (**2003**). ''Central locus for non-speech context effects on phonetic identification,'' J. Acoust. Soc. Am. **113**, 53–56.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U.P., New York).

Mann, V. A. (**1980**). ''Influence of preceding liquid on stop-consonant perception,'' Percept. Psychophys. **28**, 407–412.

Mann, V. A., and Liberman, A. M., (**1983**). ''Some differences between phonetic and auditory modes of perception,'' Cognition **14**, 211–235.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [ʃ]-[s] distinction,'' Percept. Psychophys. **28**, 213–228.

Mann, V. A., and Repp, B. H. (**1981**). ''Influence of preceding fricative on stop consonant perception,'' J. Acoust. Soc. Am. **69**, 548–558.

Moore, B. (**1988**). *An Introduction to the Psychology of Hearing* (Academic, London).