Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

3-2006

Speech Categorization in Context: Joint Effects of Nonspeech and Speech Precursors

Lori L. Holt
Carnegie Mellon University, lholt@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/psychology

Published In
.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Speech categorization in context: Joint effects of nonspeech and speech precursors

Lori L. Holt[a)]
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 20 March 2005; revised 10 March 2006; accepted 14 March 2006)

The extent to which context influences speech categorization can inform theories of pre-lexical speech perception. Across three conditions, listeners categorized speech targets preceded by speech context syllables. These syllables were presented as the sole context or paired with nonspeech tone contexts previously shown to affect speech categorization. Listeners' context-dependent categorization across these conditions provides evidence that speech and nonspeech context stimuli jointly influence speech processing. Specifically, when the spectral characteristics of speech and nonspeech context stimuli are mismatched such that they are expected to produce opposing effects on speech categorization the influence of nonspeech contexts may undermine, or even reverse, the expected effect of adjacent speech context. Likewise, when spectrally matched, the cross-class contexts may collaborate to increase effects of context. Similar effects are observed even when natural speech syllables, matched in source to the speech categorization targets, serve as the speech contexts. Results are well-predicted by spectral characteristics of the context stimuli. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195119]

PACS number(s): 43.71.An, 43.71.−k, 43.66.Ba, 43.66.−x [ALF] Pages: 4016–4026

I. INTRODUCTION

Context plays a critical role in speech categorization. Acoustically identical speech stimuli may be perceived as members of different phonetic categories as a function of the surrounding acoustic context. Mann (1980), for example, has shown that listeners' categorization of a series of speech stimuli ranging perceptually from /ga/ to /da/ is shifted toward more "ga" responses when these target syllables are preceded by /al/. The same stimuli are more often categorized as "da" when /ar/ precedes them. Such context-dependent phonetic categorization is a consistent finding in speech perception (e.g., Lindblom and Studdert-Kennedy, 1967; Mann and Repp, 1981; see Repp, 1982 for review).

Consideration of how to account for context-dependent speech perception highlights larger theoretical issues of how best to characterize the basic representational currency and processing characteristics of speech perception. Relevant to this interest, an avian species (Japanese quail, *Coturnix coturnix japonica*) has been shown to exhibit context-dependent responses to speech (Lotto *et al.*, 1997). Birds operantly trained to peck a lighted key in response to a /ga/ stimulus peck more robustly in later tests when test syllables are preceded by /al/. Correspondingly, birds trained to peck to /da/ peck most vigorously to test stimuli when the are preceded by /ar/. Thus, birds exhibit shifts in pecking behavior contingent on preceding context analogous to context-dependent human speech categorization. The birds had no previous experience with speech, so their behavior cannot be explained on the basis of learned covariation of acoustic attributes across contexts or on the basis of existing phonetic categories. It is also unlikely that quail have access to specialized speech processes or knowledge of the human vocal tract. The parallels between quail and human behavior suggest a possible role for general auditory processing, not specific to speech or dependent upon extensive experience with the speech signal, in context-dependent speech perception.

In accord with the hypothesis that general, rather than speech-specific, processes play a role in context-dependent speech perception there is evidence that nonspeech acoustic contexts affect speech categorization by human listeners. Following the findings of Mann (1980), Lotto and Kluender (1998) synthesized two sine-wave tones, one with a higher frequency corresponding to the third formant (F3) offset frequency of /al/ and the other with a lower frequency corresponding to the /ar/ F3 offset frequency. When these nonspeech stimuli preceded a /ga/ to /da/ target stimulus series like that studied by Mann (1980), speech categorization was influenced by the precursor tones. Listeners more often categorized the syllables as "ga" when they were preceded by the higher-frequency sine-wave tone modeling /al/. The same stimuli were more often categorized as "da" when the tone modeling /ar/ preceded them. Thus, nonspeech stimuli mimicking very limited spectral characteristics of speech contexts also influence speech categorization.

Nonspeech-elicited context effects on speech categorization appear to be a general phenomenon. Holt (1999; Holt and Lotto, 2002) reports that sine-wave tones or single formants situated at the second formant (F2) frequency of /i/ versus /u/ shift categorization of syllables ranging perceptually from /ba/ to /da/ in the same manner as the vowels they model. Likewise, flanking nonspeech frequency-modulated glides that follow the F2 formant trajectories of /bVb/ and /dVd/ syllables influence categorization of the intermediate vowel (Holt *et al.*, 2000). A number of other studies demonstrate interactions of nonspeech context and speech percep-

[a)]Electronic mail: lholt@andrew.cmu.edu

 © 2006 Acoustical Society of America

tion (Fowler *et al.*, 2000; Kluender *et al.*, 2003; Watkins and Makin, 1994, 1996a, 1996b) and the effects appear to be reciprocal. Stephens and Holt (2003) report that preceding /al/ and /ar/ syllables modulate perception of following nonspeech stimuli. Follow-up studies have demonstrated that listeners are unable to relate the sine-wave tone precursors typical of these studies to the phonetic categories the tones model (Lotto, 2004); context-dependent speech categorization is elicited even with nonspeech precursors that are truly perceived as nonspeech events.

There is evidence that even temporally nonadjacent nonspeech precursors can influence speech categorization. Holt (2005) created "acoustic histories" composed of 21 sine-wave tones sampling a distribution defined in the acoustic frequency dimension. The acoustic histories terminated in a neutral-frequency tone that was shown to have no effect on speech categorization. In this way, the context immediately adjacent to the speech target in time was constant across conditions. The mean frequency of the acoustic histories differentiated conditions, with distribution means approximating the tone frequencies of Lotto and Kluender (1998). Despite their temporal nonadjacency with speech targets, the nonspeech acoustic histories had a significant effect on categorization of members of a following /ga/ to /da/ speech series. In line with previous findings, the higher-frequency acoustic histories resulted in more "ga" responses whereas the lower-frequency acoustic histories led to more "da" responses. These effects were observed even when as much as 1.3 s of silence or 13 repetitions of the neutral tone separated the acoustic histories and the speech targets in time.

In each of the cases for which effects of nonspeech contexts on speech categorization have been observed, the nonspeech contexts model limited spectral characteristics of the speech contexts. As simple pure tones or glides, they do not possess structured information about articulatory gestures. Moreover, even the somewhat richer nature of the acoustic history tone contexts of Holt (2005) are far removed from the stimuli that may be perceived as speech in sine-wave speech studies (e.g., Remez *et al.* 1994). The commonality shared between the tones composing the acoustic histories and sine-wave speech is limited to the fact that both make use of sinusoids. The tonal sine-wave speech stimuli are composed of three or four concurrent time-varying sinusoids, each mimicking the center frequency and amplitude of a natural vocal resonance measured from a real utterance. Thus, the sine-wave replicas that may give rise to speech percepts possess an overall acoustic structure that much more closely mirrors the speech spectrum it models. By contrast, the single sine-waves of, for example, Lotto and Kluender (1998) or the sequences of sine waves of Holt (2005) are far more removed from the precise time-varying characteristics of speech. The tones composing the acoustic histories of Holt (2005) are single sinusoids of equal amplitude, separated in time (not continuous), and randomized on a trial-by-trial basis. The nonspeech contexts provide neither acoustic structure consistent with articulation nor acoustic information sufficient to support phonetic labeling (see Lotto, 2004). What they do share with the speech contexts they model is a very limited resemblance to the *spectral* information that differentiates, for example, the /al/ from /ar/ contexts that have been shown to influence speech categorization (Mann, 1980).

The directionality of the context-dependence is likewise predictable from this spectral information. Across the observations of context-dependent speech categorization for speech and nonspeech contexts, the pattern of context-dependent categorization is spectrally contrastive (Holt, 2005; Lotto *et al.*, 1997; Lotto and Kluender, 1998); precursors with acoustic energy in higher frequency regions (whether speech or nonspeech, e.g., /al/ or nonspeech sounds modeling the spectrum of /al/) shift categorization toward the speech category characterized by lower-frequency acoustic energy (i.e., /ga/) whereas lower-frequency precursors (/ar/ or nonspeech sounds modeling /ar/) shift categorization toward the higher-frequency alternative (i.e., /da/). The auditory perceptual system appears to be operating in a manner that serves to emphasize spectral change in the acoustic signal. Contrastive mechanisms are a fundamental characteristic of perceptual processing across modalities. General mechanisms of auditory processing that produce spectral contrast may give rise to the results observed for speech and nonspeech contexts in human listeners with varying levels and types of language expertise (Mann, 1986; Fowler *et al.*, 1990) and in quail subjects (Lotto *et al.*, 1997). Neural adaptation and inhibition are simple examples of neural mechanisms that exaggerate contrast in the auditory system (Smith, 1979; Sutter *et al.*, 1999), but others exist at higher levels of auditory processing (see e.g., Delgutte, 1996; Ulanovsky *et al.*, 2003; 2004) that produce contrast without a loss in sensitivity (Holt and Lotto, 2002). The observation of nonspeech context effects on speech categorization when context and target are presented to opposite ears (Holt and Lotto, 2002; Lotto *et al.*, 2003) and findings demonstrating effects of nonadjacent nonspeech context on speech categorization (Holt, 2005) indicate that the mechanisms are not solely sensory.[1] Moreover, there is evidence that mechanisms producing spectral contrast may operate over multiple time scales (Holt, 2005; Ulanovsky *et al.*, 2003, 2004).

By this general perceptual account, speech- and nonspeech-elicited context effects emerge from common processes that are part of general auditory processing. These mechanisms are broadly described as spectrally contrastive in that they emphasize spectral change in the acoustic signal, independent of its classification as speech or nonspeech or whether the signal carries information about speech articulation. So far, observed effects have been limited to the influence of speech *or* nonspeech contexts on speech categorization (or, conversely, the effects of speech contexts on nonspeech perception, Stephens and Holt, 2003). However, an account that relies upon spectral contrast makes strong directional predictions about context-dependent speech categorization in circumstances in which both speech and nonspeech contexts are present. Specifically, this account predicts that when both speech and nonspeech are present as context, their effects on speech categorization will be dictated by their spectral characteristics such that they may either cooperate or conflict in their direction of influence on speech categorization as a function of how they are paired. If the speech and nonspeech contexts are matched in the distri-

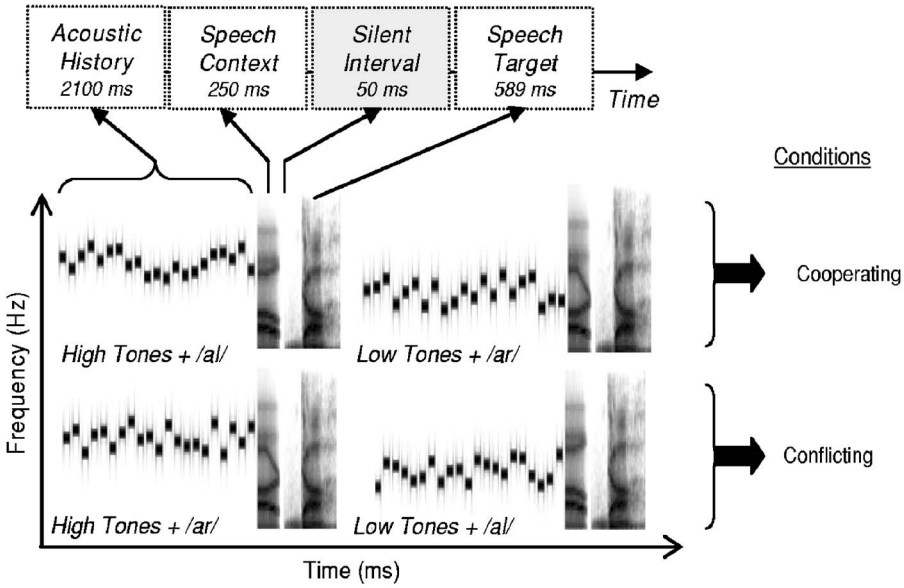

FIG. 1. At the top, an illustration displays the elements of each stimulus. Representative spectrograms (on time × frequency axes) below show example stimuli from Experiment 1 conditions. Stimuli from the Cooperating condition (top row) are composed of spectrally matched speech and nonspeech contexts that have been shown previously to shift speech categorization in the same direction. Examples of Conflicting condition stimuli for which spectrally mismatched nonspeech and speech precursors have opposing effects on speech categorization are shown in the bottom row.

bution of spectral energy that they possess such that they are expected to shift speech categorization in the same direction, then nonspeech may collaborate with speech to produce greater effects of context than observed for speech contexts alone. Conversely, when nonspeech and speech contexts possess spectra that push speech categorization in opposing directions, nonspeech contexts should be expected to lessen the influence of speech contexts on speech categorization. As a means of empirically examining the hypotheses arising from this account, the present experiments examine speech categorization when *both* speech and nonspeech signals serve as acoustic context, specifically investigating the degree to which they may jointly influence speech categorization.

II. EXPERIMENT 1

The aim of this study thus is to assess the relative influence of speech and jointly presented nonspeech contexts on speech categorization. Experiment 1 examines speech categorization of a /ga/ to /da/ syllable series across three contexts: (1) preceding /al/ and /ar/ syllables; (2) the same speech syllables paired with spectrally matched nonspeech acoustic histories (as described by Holt, 2005) that shift speech categorization in the same direction (e.g., High Mean acoustic histories paired with /al/); (3) the same speech syllables paired with spectrally mismatched nonspeech acoustic histories that shift speech categorization in opposing directions (e.g., Low Mean acoustic histories paired with /al/). Whereas the speech contexts remain consistent across conditions, the nonspeech contexts vary. Thus, if speech and nonspeech contexts fail to jointly influence speech categorization there will be no significant differences in speech categorization across conditions and, as in previous studies, speech targets preceded by /al/ will be more often categorized as "ga" than the same targets preceded by /ar/. If, however, the two sources of acoustic context mutually influence speech categorization as predicted by a general perceptual/cognitive account of context effects in speech perception then the observed context effects will vary across conditions and the relative influence of each context source on speech categorization can be assessed.

A. Methods

1. Participants

Ten adult monolingual English listeners recruited from the Carnegie Mellon University community participated in return for a small payment or course credit. All participants reported normal hearing.

2. Stimuli

Stimulus design is schematized in Fig. 1. For each stimulus an acoustic history composed of 21 sine-wave tones preceded a speech syllable context stimulus, a 50-ms silent interval, and a speech target drawn from a stimulus series varying perceptually from /ga/ to /da/.

a. Speech. Speech target stimuli were identical to those described previously (Holt, 2005; Wade and Holt, 2005). Natural tokens of /ga/ and /da/ spoken in isolation were digitally recorded from an adult male monolingual English speaker (CSL, Kay Elemetrics; 20-kHz sample rate, 16-bit resolution). From a number of natural productions, one /ga/ and one /da/ token were selected that were nearly identical in spectral and temporal properties except for the onset frequencies of F2 and F3. LPC analysis was performed on each of the tokens and a nine-step sequence of filters was created (Analysis-Synthesis Laboratory, Kay Elemetrics) such that the onset frequencies of F2 and F3 varied approximately linearly between /g/ and /d/ endpoints. These filters were excited by the LPC residual of the original /ga/ production to create an acoustic series spanning the natural /ga/ and /da/ end points in approximately equal steps. Each stimulus was 589 ms in duration. The series was judged by the experimenter to comprise a gradual shift between natural-sounding /ga/ and /da/ tokens and this impression was confirmed by regular shifts in phonetic categorization across the series by participants in the Holt (2005) and Wade and Holt (2005)

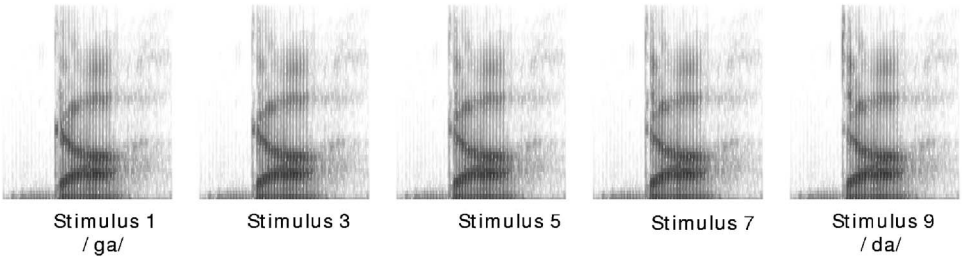

FIG. 2. Spectrograms of the odd-numbered stimuli along the nine-step /ga/ to /da/ series that served as categorization targets in Experiments 1 and 2.

studies. These speech series members served as categorization targets for each experimental condition. Spectrograms of odd-number series stimuli are shown in Fig. 2.

In addition, there were two speech context stimuli. These 250-ms syllables corresponded perceptually to /al/ and /ar/ and were composed of a 100-ms steady-state vowel followed by a 150-ms linear formant transition. Stimuli were synthesized using the cascade branch of the Klatt (1980) synthesizer. These stimuli were identical to those shown in earlier reports to produce spectrally contrastive context effects on perception of speech (Lotto and Kluender, 1998) and nonspeech (Stephens and Holt, 2003). Lotto and Kluender (1998) provide full details of stimulus synthesis.

b. Nonspeech. Acoustic histories were created as described by Holt (2005). Each acoustic history was composed of 21 70-ms sine-wave tones (30-ms silent intervals) with unique frequencies. Distributions' mean frequencies (1800 and 2800 Hz) were chosen based on the findings of Lotto and Kluender (1998), who demonstrated that single 1824 versus 2720 Hz tones produce a spectrally contrastive context effect on speech categorization targets varying perceptually from /ga/ to /da/. "Low Mean" acoustic histories were composed of 1300–2300 Hz tones ($M=1800$ Hz, 50-Hz steps). "High Mean" acoustic histories possessed tones sampling 2300–3300 Hz ($M=2800$ Hz, 50-Hz steps).

To minimize effects elicited by any particular tone ordering, acoustic histories were created by randomizing the order of the 21 tones on a trial-by-trial basis. Each trial was unique; acoustic histories within a condition were distinctive in surface acoustic characteristics, but were statistically consistent with other stimuli drawn from the distribution defining the nonspeech context. Thus, any influence of acoustic histories on speech categorization is indicative of listeners' sensitivity to the long-term spectral distribution of the acoustic history and not merely to the simple acoustic characteristics of any particular segment (for further discussion see Holt, 2005).

Tones comprising the acoustic histories were synthesized with 16-bit resolution and sampled at 10 kHz using MATLAB (Mathworks, Inc.). Linear onset/offset amplitude ramps of 5 ms were applied to all tones. Target speech stimuli were digitally down-sampled from their recording rate of 20–10 kHz and both tones and speech tokens were digitally matched to the rms energy of the /da/ end point of the target speech series.

As discussed in Sec. I, very broad interpretation of the kind of acoustic energy that may carry articulatory information may cause concern that the High and Low mean acoustic histories could serve as information about articulatory events and perhaps lead listeners to identify the nonspeech acoustic histories phonetically. To allay this concern, 10 monolingual English participants who reported normal hearing were tested in a pilot stimulus test. These participants did not serve as listeners in any of the reported experiments and had not participated in experiments of this sort before. These listeners identified the High and Low mean acoustic histories as "al" or "ar" in the context of the following speech syllable pairs described above. If the limited spectral information that the acoustic histories model from the /al/ and /ar/ contexts serves as information about articulatory events, we should expect High mean acoustic histories to elicit more "al" responses and Low mean acoustic histories to elicit more "ar" responses. This was not the case. Listeners' phonetic labeling of the High versus Low mean acoustic histories as "al" was not greater for the High mean acoustic histories ($M_{\text{High}}=51.1$, $SE=0.52$) than Low mean acoustic histories ($M_{\text{Low}}=51.0$, $SE=1.19$; $t<1$ in a paired-samples t-test).

c. Stimulus construction. Two sets of stimuli were constructed from these elements. To create the hybrid nonspeech/speech contexts preceding the speech targets, each of the nine /ga/ to /da/ target stimuli was appended to the /al/ and /ar/ speech contexts with a 50-ms silent interval separating the syllables. Each of the resulting 18 disyllables was appended to two nonspeech contexts, one an acoustic history defined by the High Mean distribution and the other an acoustic history with a Low Mean. This pairing of disyllables with acoustic histories was repeated 10 times, with a different acoustic history for each repetition. This resulted in 360 unique stimuli, exhaustively pairing /al/ and /ar/ speech contexts with High and Low mean nonspeech contexts and the nine target speech series stimuli across 10 repetitions. A second set of stimuli with only speech contexts preceding the speech targets also was created; /al/ and /ar/ stimuli were appended to each of the speech target series members with a 50-ms interstimulus silent interval for a total of 18 stimuli. These stimuli were presented 10 times each during the experiment.

3. Design and procedure

The pairing of speech and nonspeech contexts in stimulus creation yielded the two experimental conditions illustrated in Fig. 1. Stimuli making up the Conflicting condition possessed acoustic histories and speech context syllables that have been shown to have opposing effects on speech categorization (Holt, 2005; Lotto and Kluender, 1998; Mann, 1980). The Cooperating condition was made up of stimuli possessing speech and nonspeech precursor contexts that

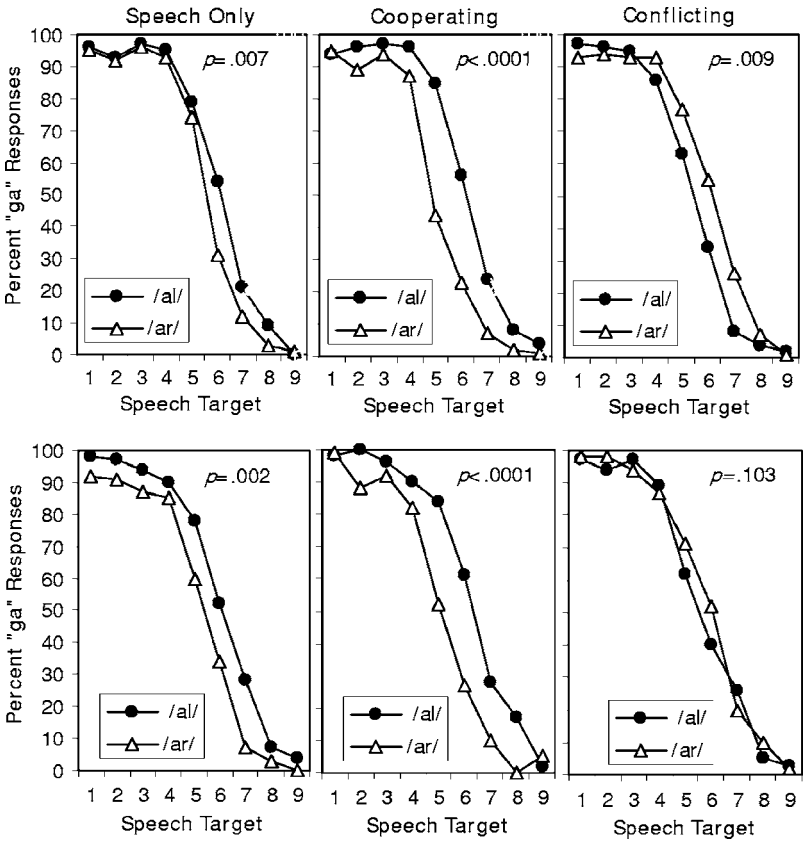

FIG. 3. Mean "ga" responses to speech series stimuli for Experiment 1 (top panel) and Experiment 2 (bottom panel). The "Speech Only" panels present categorization data for /al/ and /ar/ contexts. The other two panels illustrate categorization when the same stimuli are preceded by High and Low Mean acoustic histories and the /al/ or /ar/ precursors. In the "Cooperating" condition, speech and nonspeech precursors are expected to shift categorization in the same direction (High+/*al*/, Low+/*ar*/). In the "Conflicting" condition, acoustic histories and speech precursors exert opposite effects on speech categorization (Low+/*al*/, High+/*ar*/).

shift speech categorization in the same direction. Note that these pairings can also be described in terms of the spectral characteristics of the component context stimuli because spectral characteristics well-predict the directionality of context effects on speech categorization (e.g., Holt, 2005; Lotto and Kluender, 1998). For example, High Mean acoustic histories were matched with /al/ (also possessing greater high-frequency acoustic energy) in the spectrally matched Cooperating condition and with /ar/ (with greater low-frequency energy) in the spectrally mismatched Conflicting condition.

Seated in individual sound-attenuated booths, listeners categorized the speech target of each stimulus by pressing electronic buttons labeled "ga" and "da." Listeners completed two blocks in a single session; the order of the blocks was counterbalanced. In one block, the hybrid nonspeech plus speech contexts preceded the speech targets. In this block, stimulus presentation was mixed across the Conflicting and Cooperating conditions. In the other (Speech Only) block, participants heard only /al/ or /ar/ preceding the speech targets. Thus, each listener responded to stimuli from all three conditions.

Acoustic presentation was under the control of Tucker Davis Technologies System II hardware; stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A).

B. Results

Results were analyzed in terms of average percent "ga" responses across stimulus repetitions and are plotted in the top row of Fig. 3. The nonoverlapping categorization curves illustrated in each of the top panels of Fig. 3 are indicative of an influence of context for each condition (see also the mar-

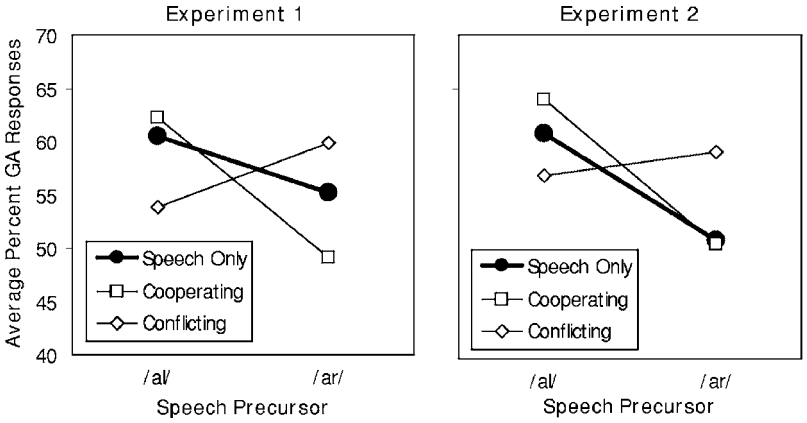

FIG. 4. Marginal means across condition and experiment.

TABLE I. Category boundaries were estimated for each participant's response to each condition of the experiment. The mean probit boundary across participants is presented in terms of the stimulus step across the nine-step /ga/ to /da/ categorization target series. The results parallel those of the ANOVA analyses across the speech stimulus series reported in the text.

Experiment	Condition	Precursor	Mean probit boundary	Standard error	t-test
1	Speech Only	/al/	7.0	0.21	$t(9)=3.13$, $p=0.01$
		/ar/	6.46	0.27	
	Cooperating	High+/al/	7.16	0.23	$t(9)=3.71$, $p=0.005$
		Low+/ar/	5.96	0.29	
	Conflicting	Low+/al/	6.12	0.21	$t(9)=3.76$, $p=0.005$
		High+/ar/	6.82	0.25	
2	Speech Only	/al/	6.79	0.24	$t(9)=3.59$, $p=0.01$
		/ar/	5.97	0.36	
	Cooperating	High+/al/	7.21	0.23	$t(9)=5.94$, $p<0.0001$
		Low+/ar/	5.98	0.24	
	Conflicting	Low+/al/	6.70	0.22	$t(9)=0.3$, $p=0.77$
		High+/ar/	6.64	0.19	

ginal means plotted in Fig. 4). Critically, although the immediately preceding speech context was constant across conditions, the observed context effects were not identical. Repeated-measures analysis of variance results are described in the following. Probit boundary analysis (Finney, 1971) of participants' category boundaries across conditions reveals the same pattern of results. The results of these analyses are provided in Table I.

1. Speech Only condition

The average percent "ga" responses across participants were submitted to a 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA. This analysis revealed a significant effect of Context, $F(1,9)=12.12$, $p=0.007$, $\eta_p^2=0.574$. Consistent with earlier findings (Lotto and Kluender, 1998; Mann, 1980), listeners categorized speech targets preceded by /al/ as "ga" significantly more often ($M=60.44$, $SE=2.86$, here and henceforth, means refer to "ga" responses averaged across target speech stimuli and participants) than the same targets preceded by /ar/ ($M=55.22$, $SE=2.57$). These data confirm that, on their own, the speech context precursors have a significant effect on categorization of neighboring speech targets. Probit boundary values are presented in Table I.

2. Cooperating condition

A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA revealed that there was also a significant effect of Cooperating nonspeech/speech contexts on speech categorization, $F(1,9)=40.22$, $p<0.0001$, $\eta_p^2=0.817$. As would be expected from the influence that speech and nonspeech contexts elicit independently (Lotto and Kluender, 1998; Holt, 2005), the effect observed in the Cooperating condition was spectrally contrastive; categorization was shifted in the same direction as in the Speech Only condition. When listeners heard speech targets preceded by High Mean acoustic histories paired with /al/, they more often categorized the targets as "ga" ($M=62.22$, $SE=2.05$) than when the same targets were preceded by Low Mean acoustic histories paired with /ar/ ($M=49.11$, $SE=2.34$).

The primary aim of this study was to examine potential joint effects of speech and nonspeech acoustic contexts in influencing speech target categorization. A $2\times2\times9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA of the categorization patterns of the Speech Only condition versus those of the Cooperating condition indicates that when speech and nonspeech contexts are spectrally matched such that they are expected to influence speech categorization similarly, they collaborate to produce an even greater context effect on speech target categorization ($M_{\text{High+}/al/}=62.22$ vs $M_{\text{Low+}/ar/}=49.11$) than do the speech targets on their own ($M_{/al/}=60.44$ vs $M_{/ar/}=55.22$), as indicated by a significant Context by Condition interaction, $F(1,9)=6.42$, $p=0.03$, $\eta_p^2=0.416$.

3. Conflicting Condition

A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA of responses to Conflicting condition stimuli revealed that when the spectra of speech and nonspeech contexts predicted opposing effects on speech categorization, there was also a significant effect of context, $F(1,9)=25.97$, $p=0.001$, $\eta_p^2=0.743$. Note, however, the direction of this effect. Listeners more often categorized target syllables as "ga" when they were preceded the High Mean acoustic histories paired with /ar/ speech precursors (% "ga" responses: $M_{\text{High+}/ar/}=59.89$, $SE=2.41$ vs $M_{\text{Low+}/al/}=49.11$, $SE=2.34$). In this example, the /ar/ speech context independently predicts more "da" responses (Mann, 1980) whereas the High Mean nonspeech acoustic histories independently predict more "ga" responses (Holt, 2005). Listeners more often responded "ga," following the expected influence of the nonspeech context rather than that of the speech context that immediately preceded the speech targets. These results indicate that when the spectra of nonspeech and speech contexts are put in conflict, the influence of temporally nonad-

jacent nonspeech context may be robust enough even to undermine the expected influence of temporally adjacent speech contexts.

Of note, a $2\times2\times9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA comparing the Conflicting condition to the Speech Only condition revealed no main effect of Context, $F(1,9)=2.98$, $p=0.119$, $\eta_p^2=0.249$, but a significant Condition by Context interaction, $F(8,72)=83.17$, $p<0.0001$, $\eta_p^2=0.902$. This indicates that the context effect produced by the speech contexts plus conflicting nonspeech contexts was statistically equivalent in magnitude, although opposite in direction, to that produced by the speech contexts alone.

4. Comparison of Cooperating vs Conflicting conditions

The relative contributions of speech and nonspeech contexts can be assessed with a $2\times2\times9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA comparing the effects of nonspeech/speech hybrid contexts across Cooperating and Conflicting conditions. This analysis reveals an overall main effect of Context (context was coded in terms of the nonspeech segment of the precursor), $F(1,9)=37.207$, $p<0.0001$, $\eta_p^2=0.805$, such that listeners more often labeled speech targets as "ga" when nonspeech precursors were drawn from the High Mean acoustic history distribution ($M=61.06$, $SE=2.01$) than the Low Mean distribution ($M=51.50$, $SE=2.09$). The contribution of the speech contexts to target syllable categorization is reflected in this analysis by the significant Condition by Acoustic History interaction, $F(1,9)=9.69$, $p=0.01$, $\eta_p^2=0.518$. With /al/ precursors, targets were somewhat more likely to be categorized as "ga" ($M=58.056$, $SE=1.9$) whereas with /ar/ precursors the same stimuli were less likely to be categorized as "ga" ($M=54.50$, $SE=2.13$). Thus, across conditions there is evidence of the joint influence of speech and nonspeech contexts. Moreover, the directionality of the observed effects is well-predicted by the spectral characteristics of the speech and nonspeech contexts.

C. Discussion

The percept created by the experiment 1 hybrid nonspeech/speech stimuli is one of rapidly presented tones preceding a bi-syllabic speech utterance. One could easily describe these nonspeech precursors as extraneous to the task of speech categorization and, indeed, listeners were not required to make any explicit judgments about them during the perceptual task. The task in this experiment was *speech perception*. Yet, even in these circumstances nonspeech contexts contributed to speech categorization. Speech does not appear to have a privileged status in producing context effects on speech categorization, even when afforded the benefit of temporal adjacency with the target of categorization.

Although general perceptual/cognitive accounts of speech perception are most consistent with these effects and can account for the directionality of the observed context effects, it is nonetheless surprising even from this theoretical perspective that the effect of nonspeech contexts is so robust. The results run counter to modular accounts that would suggest that there are special-purpose mechanisms for processing speech that are informationally encapsulated and therefore impenetrable to influence by nonlinguistic information (Liberman *et al.*, 1967; Liberman and Mattingly, 1985). The very simple sine-wave tones that comprised the nonspeech contexts are among the simplest of acoustic signals. To consider them information for speech perception by a speech-specific module would require a module so broadly tuned as to be indistinguishable from more interactive processing schemes. The results of Experiment 1 also are difficult to reconcile with a direct realist perspective on speech perception. The direct realist interpretation of the categorization patterns observed in the Speech Only condition is that the speech contexts provide information relevant to parsing the dynamics of articulation (Fowler, 1986; Fowler and Smith, 1986; Fowler *et al.*, 2000). It is unclear from a direct realist perspective why, in the presence of clear speech contexts providing information about articulatory gestures, listeners would be influenced by nonspeech context sounds at all, let alone be more influenced by the nonspeech contexts than the speech contexts in the Conflicting condition. It does not appear that context must carry structured information about articulation to have an impact on speech processing.

III. EXPERIMENT 2

The stimuli created for Experiment 1 were constructed as a compromise among stimuli used in previous experiments investigating speech and nonspeech context effects. The /ga/ to /da/ speech target series of Holt (2005) was chosen for its naturalness in an effort to provide the most conservative estimate of context-dependence (synthesized or otherwise degraded speech signals are typically thought to be more susceptible to contextual influence). The synthetic /al/ and /ar/ contexts were taken from the stimulus materials of Lotto and Kluender (1998) because they produce a robust influence on speech categorization along a /ga/ to /da/ series (see also Stephens and Holt, 2003). Nonetheless, there are stimulus differences originating from the synthetic nature of the /al/ and /ar/ speech contexts of Experiment 1 and the more natural characteristics of the speech targets. This could lead the two sets of speech materials to be perceived as originating from different sources. If this was the case, the independence of the sources should reduce or eliminate articulatory gestural information relevant to compensating for intraspeaker effects of coarticulation (a within-speaker phenomenon) via gestural parsing. Although previous research has provided evidence of cross-speaker phonetic context effects (Lotto and Kluender, 1998), it may nonetheless be argued that Experiment 1 does not provide the most conservative test of nonspeech/speech context effects because of the possible perceived difference in speech source across syllables.

Therefore, Experiment 2 was conducted in the same manner as Experiment 1, but using natural /al/ and /ar/ productions recorded from the same speaker that produced the end point stimuli of the /ga/ to /da/ speech target stimulus series. The experiment thus serves as both a replication of

the findings of Experiment 1 and an opportunity to investigate whether the influence of nonspeech context on speech categorization is robust enough to persist even when speech contexts and targets originate from the same source.

A. Methods

1. Participants

Ten adult monolingual English listeners, none of whom participated in Experiment 1, received a small payment or course credit for volunteering. All participants were recruited from the Carnegie Mellon University community and reported normal hearing.

2. Stimuli

Stimulus design was identical to that of Experiment 1, except that the speech context stimuli were digitally recorded (20-kHz sample rate, 16-bit resolution) natural utterances of /al/ and /ar/ spoken in isolation by the same speaker who recorded the natural speech end points of the target stimulus series. The 350-ms syllables were down-sampled to 10 kHz and matched in rms energy to the /da/ end point of the target stimulus series. These syllables served as the speech contexts in the stimulus construction protocol described for Experiment 1.

3. Design and Procedure

The design, procedure, and apparatus were identical to those of Experiment 1.

B. Results

The results of Experiment 2 are shown in the bottom row of Fig. 3. Marginal means are plotted in Fig. 4. Probit boundary values are presented in Table I.

1. Speech Only condition

Consistent with the findings of Experiment 1, there was a significant influence of preceding /al/ and /ar/ on speech target categorization. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA confirmed that listeners categorized speech targets preceded by /al/ as "ga" significantly more often ($M=60.89$, $SE=2.28$) than the same targets following /ar/ ($M=51.00$, $SE=3.4$), $F(1,9)=18.426$, $p=0.002$, $\eta_p^2=0.672$. Thus, natural /al/ and /ar/ recordings matched to the target source produced a significant context effect on categorization of the speech targets.

One potential concern about the use of synthesized speech contexts in Experiment 1 was that a perceived change in talker may have reduced observed effects of speech context. However, comparison of the influence of the synthesized versus naturally produced speech contexts on categorization of the speech targets with a cross-experiment $2\times 2\times 9$ (Experiment$\times$Context$\times$Target Speech Stimulus) mixed model ANOVA with Experiment as a between-subjects factor, did not reveal a significant difference in the context effects produced by the /al/ and /ar/ stimuli of Experiments 1 and 2, $F(1,18)=2.88$, $p=0.11$, $\eta_p^2=0.138$.

2. Cooperating condition

The primary question of interest is whether nonspeech contexts influence speech categorization even in the presence of adjacent speech signals originating from the same source. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA supports what is illustrated in the bottom row of Fig. 2. There was a significant spectrally contrastive effect of the cooperating, spectrally matched, speech and nonspeech contexts, $F(1,9)=76.21$, $p<0.0001$, $\eta_p^2=0.894$, such that listeners more often categorized speech targets as "ga" when High Mean nonspeech precursors and /al/ preceded them ($M=64.00$, $SE=2.19$) than when Low Mean nonspeech precursors and /ar/ preceded them ($M=50.56$, $SE=2.54$).

An additional $2\times 2\times 9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA examined the context effects across the Speech Only and Cooperating conditions of Experiment 2. Of note, although the mean difference between conditions was greater for the Cooperating condition ($M_{(\text{High+/al/})\text{-}(\text{Low+/ar/})}=13.44\%$) than the Speech Only condition ($M_{\text{/al/-/ar/}}=9.89\%$), this difference was not statistically reliable, $F(1,9)=2.52$, $p=0.147$, $\eta_p^2=0.219$. This differs from Experiment 1, for which speech and nonspeech contexts collaborated in the Cooperating condition to produce a greater effect of context on speech categorization than did the speech contexts alone.

3. Conflicting condition

An analogous analysis was conducted across the Speech Only and Conflicting conditions, revealing that the categorization patterns observed for the Conflicting condition were significantly different than those found for the Speech Only condition, $F(1,9)=18.63$, $p=0.002$, $\eta_p^2=0.674$. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA showed that, contrary to the robust effect of speech contexts in the Speech Only condition, there was no effect of hybrid nonspeech/speech contexts in the Conflicting condition, $F(1,9)=3.29$, $p=0.103$, $\eta_p^2=0.267$ ($M_{\text{Low+/al/}}=57.00$, $SE=2.52$ vs $M_{\text{High+/ar/}}=59.11$, $SE=2.52$). The presence of spectrally mismatched nonspeech contexts effectively neutralized the influence of the natural speech precursors.

4. Comparing Cooperating and Conflicting conditions

A comparison of the patterns of categorization for the hybrid nonspeech/speech context conditions with a $2\times 2\times 9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA revealed a main effect of Context (entered into the analysis in terms of the nonspeech characteristics of the context) such that listeners more often labeled the speech targets as "ga" when the nonspeech context was drawn from a distribution with a High Mean frequency ($M=60.50$, $SE=2.26$) than when it was drawn from a distribution with a Low Mean frequency ($M=54.83$, $SE=2.47$), $F(1,9)=25.61$, $p=0.001$, $\eta_p^2=0.740$. In this analysis, the contribution of the speech contexts to speech target categorization was reflected by a significant Condition by Context interaction, $F(1,9)=99.10$, $p<0.0001$, $\eta_p^2=0.917$ such that listeners more often labeled targets as "ga" when the precur-

sor syllable was /al/ ($M=61.56$, $SE=2.25$) than when it was /ar/ ($M=53.78$, $SE=2.42$). Thus, both speech and nonspeech contexts contributed to the categorization responses observed in the hybrid context conditions of Experiment 2.

C. Discussion

The overall pattern of results of Experiment 2 confirms that speech and nonspeech contexts jointly influenced speech categorization, even when the natural speech contexts were matched to the categorization targets in source. Of note, however, the influence of the nonspeech contexts in the presence of the natural speech contexts was less dramatic than were the effects observed when the same nonspeech contexts were paired with synthesized speech syllables in Experiment 1. Contrary to the findings of Experiment 1, the nonspeech precursors did not collaborate with the natural speech contexts of Experiment 2 to produce a context effect significantly greater than that elicited by the natural speech syllables alone. Moreover, although there was strong evidence of joint nonspeech/speech context effects in the Experiment 2 Conflicting condition, the influence of the nonspeech was not so strong as to overpower the natural speech context and reverse the observed context effect as it did in Experiment 1. These more modest patterns of interaction may be due to the somewhat stronger effect of context elicited by the natural speech syllables. This difference, evident in the shift in mean "ga" responses across speech contexts in the Speech Only conditions (the difference in mean percent "ga" responses for /al/ vs /ar/ contexts was 5.22% for Experiment 1 and 9.89% in Experiment 2) was not consistent enough to be statistically reliable across experiments. Nonetheless, the pattern of effect sizes suggests that the natural speech syllables may have contributed a greater overall influence to target speech categorization. This is simply to say that the speech contexts of Experiment 2 may have contributed more to the resulting target percept relative to the strong influence of the nonspeech contexts than did the synthesized syllables of Experiment 1.

To more closely examine this possibility, an additional statistical analysis was conducted to determine the relative contribution of speech contexts in the hybrid nonspeech/speech conditions across experiments as speech context type (synthesized, natural) varied. A $2\times2\times2\times9$ (Experiment $\times$Condition$\times$Context$\times$Target Speech Stimulus) mixed model ANOVA with Experiment as a between-subjects factor compared the relative influence of speech contexts in the Cooperating versus Conflicting conditions across experiments. A significant difference is reflected by the three-way Experiment$\times$Condition$\times$Context interaction, $F(1,18)=56.83$, $p<0.0001$, $\eta_p^2=0.759$. When nonspeech contexts were present, the relative influence of synthesized versus natural speech contexts differed. Computing the difference in mean "ga" responses in the hybrid nonspeech/speech conditions conditioned on the speech context illustrates why this is so. The categorization shift attributable to the synthesized speech contexts of Experiment 1 ($M_{/al/}-M_{/ar/}=58.06-54.50=3.56$) is significantly less than that of the natural speech contexts of Experiment 2 ($M_{/al/}-M_{/ar/}=61.56-53.78=7.78$). Many factors may have contributed to the relatively greater effect of context produced by the natural syllables including, but not limited to, the richer acoustic characteristics of natural speech, the closer spectral correspondence of the natural syllables with the target speech syllables, perception of the two syllables as originating from the same talker, amplitude relationships of the spectral energy from the two precursors, and auditory grouping by common acoustic characteristics. Whatever caused the natural syllables to be relatively stronger contributors to the effect on speech categorization, the results of the Conflicting condition nevertheless provide strong evidence of perceptual contributions from both nonspeech and speech contexts even for natural speech contexts. Moreover, the statistical analyses of the Experiment 2 Cooperating versus Conflicting conditions provide corroborating evidence that both the speech and nonspeech contexts contributed to the observed pattern of results.

IV. GENERAL DISCUSSION

A spectral contrast account of context-dependent speech perception makes strong directional predictions about context-dependent speech categorization in circumstances in which both speech and nonspeech contexts are present. Specifically, it is expected that the effect of joint speech/nonspeech context on speech categorization will be dictated by the spectral characteristics of each source of context such that the speech and nonspeech contexts may either cooperate or conflict in their direction of influence on speech categorization as a function of how they are paired. The results of two experiments demonstrate that speech and nonspeech contexts do jointly influence speech categorization. When hybrid nonspeech/speech context stimuli were spectrally matched in Experiment 1, they collaborated to produce a bigger effect of context on speech categorization than did the same speech contexts on their own. A context effect on speech categorization was also observed in this condition in Experiment 2 (for which natural utterances provided speech context), but this effect was not significantly greater than that observed for the natural speech contexts alone.

When the spectra of the hybrid nonspeech/speech contexts were spectrally mismatched such that they predicted opposing influences on speech categorization, the observed context effects differed from the context effect produced independently by the speech contexts. In Experiment 1, the context effect observed in the Conflicting condition was of equal magnitude, but in the opposite direction of that observed for solitary speech contexts. The direction of the context effect was predicted, not by the adjacent speech contexts, but instead by the spectral characteristics of the temporally nonadjacent nonspeech contexts. A qualitatively similar, although less dramatic, effect was observed for the spectrally conflicting speech and nonspeech contexts of Experiment 2; the nonspeech contexts neutralized the effect of speech context such that no context-dependent shift in target speech categorization was observed. Overall, the effects observed for the hybrid context conditions of Experiment 2, with natural speech contexts matched in source to the target

syllables, were relatively more modest than those observed in Experiment 1. This may have been due to the somewhat larger effect of context exerted by the natural speech contexts. Most important to the aims of the study, however, both experiments provided evidence that linguistic and nonlinguistic sounds jointly contribute to observed context effects on speech categorization. The sum of the results is consistent with general auditory/cognitive approaches with an emphasis on the shared characteristics of the acoustic signal and the general processing of these elements, in this case, spectral distributions of energy. The spectral characteristics of the context stimuli, whether the stimuli were speech or nonspeech, predicted the effects upon the speech categorization targets.

Mechanistically, an important issue that remains is whether such general auditory representations common to speech and nonspeech govern the joint effects of speech and nonspeech contexts on speech categorization or whether independent representations of the context stimuli exert an influence on speech categorization at a later decision stage. This is a thorny issue to resolve in any domain. Some theorists suggest that if processes share common resources or hardware, they can be expected to interfere or otherwise interact with one another whereas if they are distinct, they should not. The present results meet this criterion for indication of common resources or hardware, but further investigation will be required to hold this question to a strict test. Nevertheless, whether nonspeech contexts are operative on common representations or integrated at a decision stage, the information that is brought to bear on speech categorization is clearly not dependent on the signal carrying information about articulation *per se*. An account cognizant of the spectral distributions of acoustic energy possessed by the context stimuli, as postulated by a general auditory/cognitive account under the term *spectral contrast* makes the only clear predictions of what happens to speech categorization when speech and nonspeech are jointly present in the preceding input and these predictions are supported by the results.

With respect to spectral contrast, there is an element of these experiments that may seem puzzling. Considering that previous research has demonstrated that *adjacent* nonspeech context influences speech categorization (e.g., Lotto and Kluender, 1998), one may wonder why the nonspeech contexts of the present experiments exerted their influence on the nonadjacent speech *targets* rather than the adjacent speech *contexts*. To understand why this should be so, it is useful to think about speech categorization as drawing from multiple sources of information.[2] Context is merely one source of information; the acoustic signal corresponding to the target of perception is another. If the acoustic signal greatly favors one speech category alternative over another then context exerts very little effect. This is the case, for example, for the more limited effects of context that emerge (here, and in other experiments) at the end points of the target speech categorization stimulus series where acoustic information is unambiguous with respect to category membership. However, when acoustic signals are partially consistent with multiple speech categories context has a role in categorization. In the present experiments, the speech target syllables were acoustically manipulated to create a series varying perceptually from /ga/ to /da/. Thus, by their very design the intermediate stimuli along this series were acoustically ambiguous and partially consistent with both /ga/ and /da/. Context was thus afforded an opportunity to exert an influence. On the contrary, the acoustic structure of the speech context stimuli in the present experiments overwhelmingly favored either /al/ or /ar/; they were perceptually unambiguous and context therefore could exert little influence. The results of the present experiments demonstrate that when the speech contexts are acoustically unambiguous, they contribute to the effects of context rather than reflect the influence of the nonspeech precursors. Although it may seem surprising that the nonspeech context stimuli should influence perception of nonadjacent speech targets, recent research has demonstrated that the auditory -system is willing to accept context information as evidence by which to shift a categorization decision even when it occurs more than a second prior and even when multiple acoustic signals intervene (Holt, 2005). By these standards, the influence of the nonadjacency of the nonspeech contexts with the speech targets in the present experiments is relatively modest.

In sum, the joint influence of speech and nonspeech acoustic contexts on speech categorization can most simply be accounted for by postulating common general perceptual origins. Previous research has highlighted parallels between phonetic context effects and those observed between purely non-speech sounds (e.g., Diehl and Walsh, 1989), but these results have been challenged on the grounds that perception of nonspeech analogs to speech cannot be directly compared to speech perception, since speech has a clear, identifiable environmental source whereas nonspeech analogs to speech (pure tones, for example) do not (Fowler, 1990). A response to this challenge is that nonspeech contexts influence perception of *speech* (e.g., Lotto and Kluender, 1998). This is a stronger test in that it identifies the information sufficient for influencing speech categorization; when nonspeech stimuli model limited acoustic characteristics of the speech stimuli that produce context effects on speech targets these nonspeech sounds likewise elicit contexts effects on speech categorization. The present experiments introduce a new paradigm to test the joint effects of speech and nonspeech context stimuli on speech categorization. This paradigm is perhaps even stronger in that it allows investigation of the influence of nonspeech signals on speech categorization in the presence of speech context signals that also exert context effects. The present results demonstrate the utility of this tool in pursuing the theoretical question of how best to account for the basic representation and processing of speech.

ACKNOWLEDGMENTS

The author thanks Dr. A. J. Lotto, Dr. R. L. Diehl, Dr. C. Fowler, and Dr. T. Wade for helpful discussions of these results and C. Adams for her essential role in conducting the research. The work was supported by grants from the James S. McDonnell Foundation (Bridging Mind Brain and Behavior, 21st Century Scientist Award) and the National Institutes of Health (2 RO1 DC004674-04A2).

[1]Although dichotic presentation of single tones and speech targets has been shown to produce context effects (Lotto *et al.*, 2003), investigation of the influence of multiple-tone acoustic history contexts on speech categorization under dichotic presentation conditions has not been reported to date. However, the long time course (>1 s) over which effects of tonal acoustic histories on speech categorization persist and the observation that tonal acoustic histories influence speech categorization even when as many as 13 neutral tones intervene between the acoustic history and speech target argue that central (i.e., not purely sensory) auditory mechanisms play an important role (Holt, 2005).

[2]This analysis is consistent with the work of a rational Bayesian decision maker whereby the optimal policy is to combine information from different sources to assign posterior probabilities to possible interpretations of the input and choose the alternative with the highest posterior probability. This approach is amenable to speech perception in that stochastic versions of the TRACE model of speech perception (McClelland, 1991) implement optimal Bayesian inference (Movellan and McClelland, 2001). Moreover, recent theoretical discussions have highlighted how Bayesian analysis may be fruitfully applied to issues in speech perception (Geisler and Diehl, 2002, 2003).

Delgutte, B. (**1996**). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 505–538.

Diehl, R. L., and Walsh, M. A. (**1989**). "An auditory basis for the stimulus-length effect in the perception of stops and glides," J. Acoust. Soc. Am. **85**, 2154–2164.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, Cambridge, MA).

Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phonetics **14**, 3–28.

Fowler, C. A. (**1990**). "Sound-producing sources as objects of perception: Rate normalization and nonspeech perception," Percept. Psychophys. **88**, 1236–1249.

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.

Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as 'vector analysis:' An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 123–139.

Geisler, W. S., and Diehl, R. L. (**2002**). "Bayesian natural selection and the evolution of perceptual systems," Philos. Trans. R. Soc. London, Ser. B **357**, 419–448.

Geisler, W. S., and Diehl, R. L. (**2003**). "A Bayesian approach to the evolution of perceptual and cognitive systems," Cogn. Sci. **118**, 1–24.

Holt, L. L. (**1999**). "Auditory constraints on speech perception: An examination of spectral contrast," Diss. Abstr. Int., B **61**, 556.

Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.

Holt, L. L., and Kluender, K. R. (**2000**). "General auditory processes contribute to perceptual accommodation of coarticulation," Phonetica **57**, 170–180.

Holt, L. L., and Lotto, A. (**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**, 971–990.

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**, 59–69.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lindblom, B. E. F., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.

Lotto, A. J. (**2004**). "Perceptual compensation for coarticulation as a general auditory process," in Proceedings of the 2003 Texas Linguistics Society Conference, edited by A. Agwuele, W. Warren, and S-H. Park Cascadilla Proceedings Project, Sommerville, MA, pp. 42–53.

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

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

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

McClelland, J. L. (**1991**). "Stochastic interactive processes and the effect of context on perception," Cogn. Psychol. **23**, 1–44.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.

Movellan, J. R., and McClelland, J. L. (**2001**). "The Morton-Massaro law of information integration: Implications for models of perception," Psychol. Rev. **108**, 113–148;

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., and Lang, J.M., (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Smith, R. L. (**1979**). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," J. Acoust. Soc. Am. **65**, 166–178.

Stephens, J. D. W., and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of non-speech sounds," J. Acoust. Soc. Am. **114**, 3036–3039.

Sutter, M. L., Schreiner, C. E., McLean, M., O'Connor, K. N., and Loftus, W. C. (**1999**). "Organization of inhibitory frequency receptive fields in cat primary auditory cortex," J. Neurophysiol. **82**, 2358–2371.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.

Ulanovsky, N., Las, L., and Nelken, I. (**2003**). "Processing of low-probability sounds by cortical neurons," Nat. Neurosci. **6**, 391–398.

Wade, T., and Holt, L. L. (**2005**). "Effects of later-occurring non-linguistic sounds on speech categorization," J. Acoust. Soc. Am. **118**, 1701–1710.

Watkins, A. J., and Makin, S. J. (**1994**). "Perceptual compensation for speaker differences and for spectral-envelope distortion," J. Acoust. Soc. Am. **96**, 1263–1282.

Watkins, A. J., and Makin, S. J. (**1996a**). "Some effects of filtered contexts on the perception of vowels and fricatives," J. Acoust. Soc. Am. **99**, 588–594.

Watkins, A. J., and Makin, S. J. (**1996b**). "Effects of spectral contrast on perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **99**, 3749–3757.