Perceptual context effects of speech and nonspeech sounds: The role of auditory categories

Radhika Aravamudhan[a)]
School of Audiology, Pennsylvania College of Optometry, Elkins Park, Pennsylvania 19020

Andrew J. Lotto
Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

John W. Hawks
School of Speech Pathology and Audiology, Kent State University, Kent, Ohio 44242

(Received 27 July 2007; revised 22 April 2008; accepted 13 June 2008)

Williams [(1986). "Role of dynamic information in the perception of coarticulated vowels," Ph.D. thesis, University of Connecticut, Standford, CT] demonstrated that nonspeech contexts had no influence on pitch judgments of nonspeech targets, whereas context effects were obtained when instructed to perceive the sounds as speech. On the other hand, Holt *et al.* [(2000). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722] showed that nonspeech contexts were sufficient to elicit context effects in speech targets. The current study was to test a hypothesis that could explain the varying effectiveness of nonspeech contexts: Context effects are obtained only when there are well-established perceptual categories for the target stimuli. Experiment 1 examined context effects in speech and nonspeech signals using four series of stimuli: steady-state vowels that perceptually spanned from /ʊ/-/ɪ/ in isolation and in the context of /w/ (with no steady-state portion) and two nonspeech sine-wave series that mimicked the acoustics of the speech series. In agreement with previous work context effects were obtained for speech contexts and targets but not for nonspeech analogs. Experiment 2 tested predictions of the hypothesis by testing for nonspeech context effects after the listeners had been trained to categorize the sounds. Following training, context-dependent categorization was obtained for nonspeech stimuli in the training group. These results are presented within a general perceptual-cognitive framework for speech perception research. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956482]

PACS number(s): 43.71.An, 43.71.Es, 43.66.Ba, 43.66.Ki [PEI] Pages: 1695–1703

I. INTRODUCTION

The perceived phonemic category of a speech sound can be influenced by the surrounding phonetic context. One of the earliest demonstrations of the context-dependent nature of speech perception was by Lindblom and Studdert-Kennedy (1967). They presented Swedish listeners with three series of synthetic vowels. The first series consisted of steady-state vowels varying acoustically in the frequency of the second formant (F2) and varying perceptually from /ʊ/ (low F2) to /ɪ/ (high F2). The second and third series embedded the vowels in a time varying formant context that corresponded perceptually to /wVw/ and /jVj/ (where V stands for a vowel from the series). The resulting vowel categorization functions differed depending on the context condition. Relative to the steady-state condition, more /ɪ/ responses were obtained for the /wVw/ context and more /ʊ/ responses were obtained for the /jVj/ context. Similar context-dependent vowel categorization has subsequently been demonstrated by Nearey (1989) using stop consonant contexts (/bVb/ vs /dVd/) and by Nabelek and Ovchinnikov (1997), using isolated vowels versus /wV/ contexts (with no steady-state portion).

The perceptual context shifts obtained for vowels provocatively mirrors an effect of context witnessed in normal speech production. Lindblom (1963) reported that the acoustic realization of vowels is influenced by the surrounding consonants being produced. Because vowel articulation tends to assimilate toward the articulation of the context consonants, the formant frequencies "undershoot" the values that would be obtained when the vowel is produced in isolation. Thus, an /ɪ/ produced in the /wVw/ context results in a lower F2 than if produced in isolation (closer to the low F2 at the onset and offset of /w/). Thus a /wVw/ context in production results in acoustics (lower F2) that are more similar to /ʊ/ produced in isolation. Conversely, Lindblom and Studdert-Kennedy (1967) demonstrated that /wVw/ context results in more /ɪ/ perceptual categorizations. That is, /wVw/ context results in more /ʊ/-like *productions* but more /ɪ/-like *perceptions*. The context-dependent perception appears to compensate for context-dependent changes inherent in production.

The symmetry of speech perception and production suggests that the mechanisms underlying the perceptual context effects may be specific to speech sound processing. Lindblom and Studdert-Kennedy (1967) proposed that their context effect demonstrated "that *categorization of the continuum is adjusted* in the different *environments so as to compensate for an undershoot effect* in the vowel stimuli" (pp. 830 and 842; original italics). Repp (1982) suggested

[a)]Author to whom correspondence should be addressed. Electronic mail: raravamudhan@pco.edu

that phonetic context effects are the result of listeners' "implicit knowledge of this coarticulatory variation (p. 97)." This perceptual compensation or implicit knowledge could be the result of experience with the acoustic effects of undershoot (coarticulation) or could be part of a specific perceptual process that has evolved to handle the complexity of speech production—acoustic mapping [e.g., a phonetic module such as proposed in later descriptions of motor theory (Liberman and Mattingly, 1989)].

An alternative explanation for the perceptual context effects was described by Repp (1982): Context effects result from general-auditory mechanisms (not specific to speech) with "the perception of relevant acoustic cues…somehow affected by the context" (p. 27). Repp (1982) rejected this account because "no plausible mechanisms that create such effects have been suggested, and no similar effects with nonspeech analogs have been reported so far" (p. 97). Subsequently, such an auditory account was tested by Holt *et al.* (2000) using /bVb/ and /dVd/ contexts similar to those utilized by Nearey (1989). Holt *et al.* (2000) first replicated the results of Nearey (1989) and Lindblom and Studdert-Kennedy (1967) by obtaining shifts in the categorization of a vowel series varying from /uh/ (as in but) to /eh/ (as in bet) as a function of consonant context (more /uh/ responses in /dVd/ context). They then replaced the formant transitions corresponding to the consonants with a single sine-wave tone varying in frequency along the path of F2 from the original speech stimuli. The rationale for this manipulation was to test whether context effects required the context to have phonetic content. If the perceptual effects of context are a result of a process that compensates specifically for coarticulation then shifts in categorization should only be obtained when the context is speech that can be coarticulated with the vowel. On the other hand, if perceptual shifts occur with contexts that only share some acoustic similarity with the speech contexts (but no phonetic content) then it indicates that general-auditory interactions may play a role. In fact, Holt *et al.* (2000) obtained a shift in categorization functions for the nonspeech contexts that was similar in size and direction as was obtained with the analogous speech contexts. In a follow-up experiment, Holt *et al.* (2000) also obtained perceptual shifts when the contexts were steady-state sine-wave tones with a frequency matched to the F2 onset frequencies of the transitions for the speech contexts.[1]

The pattern of results reported by Holt *et al.* (2000) is in line with experiments by Lotto *et al.* (1997) and Holt (2005) demonstrating nonspeech contexts affecting speech perception. Holt and Lotto (2006) have proposed that these context effects arise from general perceptual mechanisms that enhance the spectral contrast between neighboring sounds. A redescription of the perceptual results presented above makes this contrastive nature more apparent. In each of the cases, the vowel is categorized more often as the endpoint with the higher F2 frequency (/eh/ or /ɪ/) when the context has a lower-frequency F2 onset (/wVw/, /wV/, /bVb/, or sine waves modeling /bVb/ F2 transitions). Conversely, more low-frequency F2 vowel categorizations (ʊ/ or /uh/) occur in the contexts with high-frequency F2 onsets (/jVj/, /dVd/, or sine waves modeling /dVd/ F2 transitions). That is, the perception of the "relevant acoustic cue" (F2 frequency in the case of these vowel contrasts) is affected contrastively by the surrounding context whether it is speech or nonspeech. These contrastive perceptual patterns have also been demonstrated for speech contexts and nonspeech targets (Stephens and Holt, 2003) and for speech contexts and targets with human infant (Fowler *et al.*, 1990) and avian (Lotto *et al.*, 1997) listeners.

Whereas the results of Holt *et al.* (2000) appear to indicate that it is not critical that the context has phonemic content in order to induce speech categorization shifts, it is still the case that the listeners' task was phonetic identification. Therefore, any putative phonetic module would be active during the task making it difficult to determine if the perceptual effects are really nonspecific to speech perception. Williams (1986) conducted a study that attempted to disentangle phonetic and spectral contents in context effects by manipulating the listeners' task as opposed to the acoustics of the stimuli. Based on the stimulus sets of Lindblom and Studdert-Kennedy (1967), Williams (1986) constructed two sets of stimuli consisting of sine waves with frequencies that tracked the trajectories of the first three formants for /wʊw/-/wɪw/ and for /ʊ-ɪ/. That is, there was a /w/ context sine-wave series and an isolated-vowel sine-wave series with steps varying in the frequency of the middle tone corresponding to F2 in each series. For one task, Williams (1986) asked listeners to categorize these stimuli as exemplars of the vowels /ʊ/ and /ɪ/, taking advantage of the fact that listeners can perceive sine-wave replicas of speech phonetically when instructed to do so (e.g., Bailey *et al.*, 1997; Remez *et al.*, 1981; Best *et al.*, 1981). The categorization results demonstrated a similar context effect to that obtained by Lindblom and Studdert-Kennedy (1967). More /ɪ/ categorizations were obtained from the series modeling /wVw/ than from the series modeling isolated vowels. This result would be predicted by both speech-specific and general-auditory accounts of phonetic context effects. In the second task, listeners were presented the same stimulus sets but were asked to categorize the stimuli in terms of pitch: "high" versus "low." This judgment would presumably be based on the perceived frequency of the second tone (mimicking F2) because this is what varied in the series. If the perception of the frequency of F2 is altered by context, as proposed by Holt and Lotto (2006), then one should obtain a shift in pitch judgments toward more high judgments in the /wVw/ series as opposed to the isolate vowel series (a contrastive shift from the lower frequency of the context tones). In fact, Williams (1986) obtained no shift in relative pitch judgments when comparing context and isolated conditions. These results suggest that context effects require processes specific to phonetic perception. Nearey (1989) suggested that the Williams result (1986) provides strong evidence that these effects are the result of a "speech mode" as opposed to a general-auditorv effect.

The results of Williams (1985) and Holt *et al.* (2000) provide some problems for coherent interpretation. If context effects are specifically the result of phonetic perception, then why would nonspeech contexts have an effect on vowel categorization? One possibility is that the phonetic perceptual processes (or phonetic module) are not narrowly tuned and

nonspeech information can be integrated into phonetic judgments. However, this would not explain why birds show similar context effects with, presumably, no perceptual processes specifically tuned to human speech (Lotto *et al.*, 1997). Another problem with interpreting the Williams results (1986) is that Mullenix *et al.* (1988) in an attempted replication of Williams results (1986) did obtain more high pitch judgments for sine-wave stimuli based on /wVw/ compared to a series based on isolated vowels. Nevertheless, if context effects were the result of general perceptual processes, why would it be so much easier to obtain shifts in phonetic categorization tasks as opposed to more general tasks such as relative pitch judgments? One possible explanation is that the difference between these two tasks extends beyond the fact that one uses phonetic labels and the other pitch labels.

One of the salient differences between speech perception and other perceptual tasks studied by psychoacousticians is that adult listeners have extensive experience with speech sounds and that they have fairly well-defined perceptual categories for these sounds. Assigning high or low pitch labels to sounds is not a very typical task and, in our experience doing nonspeech studies, is not one that naive participants perform particularly well with no practice. In addition, it is unlikely that listeners have any well-established perceptual categories for novel sine-wave stimuli such as those used by Williams (1986).[2] It is possible that the lack of experience with the labeling task and existing pitch categories for the stimuli interfered with the ability to obtain context-dependent shifts in the pitch judgment task. With no established referent categories for pitch, the listeners may have adjusted their criteria for the labeling task to the perceptual range presented by the stimulus sets. That is, even if a general-auditory contrast effect was present for the context stimuli, the criterion for pitch judgments could be set relative to the shifted perceptual range because the criterion is not anchored to any predetermined category structure. The experiments presented below were designed to test the hypothesis that context effects can be obtained for nonphonetic categorization tasks if the listener has established perceptual categories specific to the labeling task. The comparisons made here are between two speech series and two nonspeech series (sine-wave analogs) that are modeled on acoustic characteristics of the speech series. In keeping with Lindblom and Studdert-Kennedy (1967), Williams (1986), and Mullenix *et al.* (1988), the context effects are measured as differences in response to isolated-vowel stimuli (/*u*/–/ɪ/) versus vowels in the context of /w/. However, in the current study, the context condition corresponds to /wV/ as opposed to /wVw/. These stimuli are modeled in part after parameters used by Nabelek and Ovchinnikov (1997) in their extensive study of the acoustic parameters that affect the size of context effect shifts. They demonstrated that preceding the vowel with /w/ results in significant shifts in vowel categorization but following the vowel with /w/ does not.

In experiment 1, participants were asked to label the nonspeech series with arbitrary labels (group 1 or group 2) with minimal practice and then were asked to label the speech series with the phonetic category for the vowel. Based on the results of Williams (1986), it was predicted that categorization shifts between context and isolated series only would be witnessed for the phonetic task. Experiment 2 provides participants with the same stimuli and tasks except that the listeners from one group go through extensive training categorizing the isolated-vowel sine-wave stimuli as groups 1 and 2 (with no experience with the context sine-wave stimuli). The prediction derived from the hypothesis presented above is that subsequent to training listeners will demonstrate a shift in categorization when the sine-wave stimuli are presented in /wV/ modeled context. This would indicate that the null effect for pitch judgments in the work of Williams (1986) may be due to the lack of well-formed perceptual categories for this task, as opposed to the nonphonetic nature of the task. However, if a context effect is not obtained for the nonspeech series in experiment 2 then this will provide further evidence [from better matched tasks than used by Williams (1986)] that context effects are specific to phonetic categorization.

II. EXPERIMENT 1

Experiment 1 is designed to investigate whether context effects can be elicited with a nonphonetic task. Context effects will be operationally defined as shifts in the overall percentage of a particular category response between an isolated (steady-state) series and a series with context. Categorization responses were collected for four series (two speech and two nonspeech sine-wave analogs). The four series were (1) isolated vowels synthesized to vary from /ʊ/ to /ɪ/ (V series), (2) vowels synthesized in the /wV/ context (wV series), (3) sine-wave analogs with steady-state frequencies at the nominal formant frequencies for the isolated vowels (V-SW series; vowel-sine wave), and (4) sine-wave analogs of the wV speech series (wV-SW series). Based on previous work by Nabelek and Ovchinnikov (1997) and Williams (1986), it is predicted that the phonetic categorization of the speech series will result in a context effect, but that the categorization of the nonspeech series will elicit no context-moderated shift.

A. Method

1. Subject

Twenty native English speaking adults, within the age range of 20–40 years, participated in return for course credit at Kent State University. None of the listeners were phonetically trained. All the subjects were screened for normal hearing (thresholds <20 dB Hearing Level (dBHL), 250–8 kHz).

2. Stimuli

a. Speech series Isolated vowel (V) and context (wV) series were constructed, each with 150 steps of variation in nominal F2 frequency. All speech stimuli were synthesized using a digital formant synthesizer in cascade mode (Klatt, 1980) with 16 bit resolution and a sampling rate of 10 kHz. The V series consisted of 200 ms steady-state formants and a rising fundamental frequency contour from 100 to 120 Hz. The frequency and bandwidth of F1 were fixed at 400 and 60 Hz, respectively, for all tokens. The frequencies of F4 and

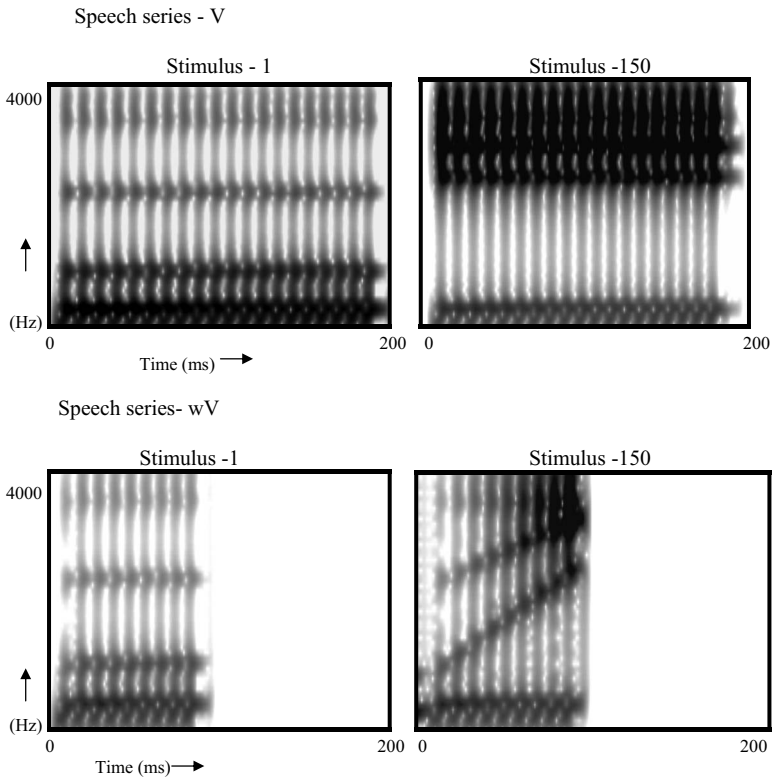


FIG. 1. Spectrograms showing endpoint stimuli for the speech series with isolated vowel (V; top) and vowel in /w/ context (wV; bottom).

F5 were fixed at 3300 and 3800 Hz with bandwidths of 250 and 300 Hz, respectively. The steps of the series varied in the frequency of F2 and F3. The frequency of F2 increased in 10 Hz steps from 1000 to 2500 Hz. These endpoint values matched the stimuli used by Nabelek and Ovchinnikov (1997). Lindblom and Studdert-Kennedy (1967) and Williams (1986) varied their F3 frequency as a function of F2. For the present stimuli, F3 frequency was calculated from F2 frequency using rule described by Nearey (1989).[3] The series varied perceptually from /ʊ/ to /ɪ/. The stimuli for the wV series started with a 20 ms steady-state portion for the formants appropriate for an initial /w/ with F1=250 Hz, F2 =800 Hz, and F3=2200 Hz and bandwidths of 60, 100, and 110 Hz, respectively (Nabelek and Ovchinnikov, 1997). The remaining 75 ms of the stimulus consisted of formant transitions from the initial /w/ values to the values described for the V series above. Thus, there was no steady-state portion for the vowels, only a transition to the final vowel target. This comparison of isolated steady-state stimuli with context stimuli lacking a steady-state portion mimics the characteristics of the stimuli set used by Lindblom and Studdert-Kennedy (1967). Nabelek and Ovchinnikov (1997) in a series of parametric studies determined that the frequency difference between initial and final F2 was the main determinant of shifts in category responses. Spectrograms of representative V and wV series stimuli are presented in Fig. 1.

b. Nonspeech series Sine-wave analogs mimicking the speech stimuli described above were constructed using the multispeech program from Kay-Elemetrics (model 2700, version 2.3). For the V-SW series, 200 ms steady-state sine waves were created with frequencies that matched the frequencies for F1, F2, and F3 of each V series token. For the wV-SW series, the sine waves varied in frequency to match the frequency patterns of the first three formants from each exemplar of the wV series. The intensity of each band was the same as that of the corresponding speech formant. For both series, the sine waves were added together with random starting phase using CoolEdit Pro (version 1.2, Syntrillium Corp.). The rms amplitude of each stimulus varied by less than ±0.6 dB.

3. Procedure

All testing was conducted in a sound-isolated booth with stimuli presented binaurally through headphones (Sennheiser HD25SP) at a comfortable listening level [~70 db sound pressure level (A)]. Stimulus presentation and response collection were controlled by a computer. Responses were made by mouse click on labeled boxes presented on a monitor. The response time was self-paced and participants could choose to listen to each stimulus as many times as they wanted before responding. The experiment consisted of four blocks of 150 stimuli. The order of the blocks was fixed for all participants, whereas the order of stimuli within a block was randomized. Each block contained stimuli from one of the four stimulus series. The order of presentation was V-SW, wV-SW (two nonspeech series), V and wV (two speech series). The nonspeech series were always presented first in order to dissuade listeners from perceiving the sounds as analogs of speech. They were told that they would hear tones that should be labeled as members of "group 1" and "group 2." No instructions were given as to the basis of this labeling difference. Prior to the testing on the V-SW series, participants were presented with a short practice set consisting of ten tokens of the endpoints from the V-SW series (five repetitions of each endpoint). Participants labeled these practice stimuli and were provided visual feedback ("correct" or "wrong"). Group 1 was mapped to the V-SW endpoint with the lowest second tone frequency (analog to F2), whereas group 2 was mapped to the stimulus with the highest second

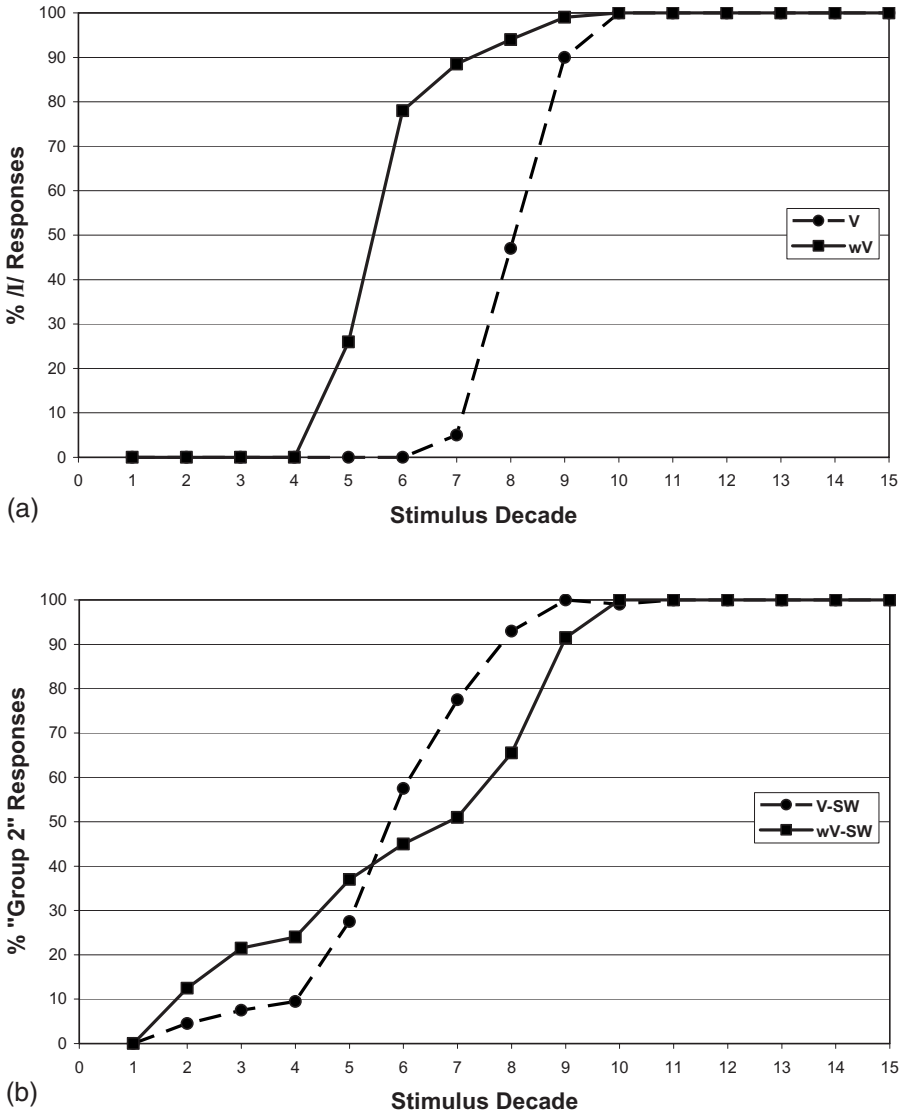


FIG. 2. Categorization functions for experiment 1. (a) Speech series: mean percentage of /ɪ/ responses by stimulus step decades (groups of ten stimuli: 1–10, 11–20, etc.) for isolated (V) and context (wV) conditions. (b) Nonspeech series: mean percentage of group 2 responses by stimulus decade for isolated (V-SW) and context (wV-SW) conditions.

tone frequency. This mapping corresponds to the low versus high pitch judgment utilized by Williams (1986) but avoids pre-existing labels. This short practice session was sufficient to help participants develop a basis for grouping of the nonspeech stimuli. No feedback was presented during any of the subsequent testing blocks. Following the V-SW block, participants labeled the wV-SW stimuli using the same group 1 and group 2 labels. The speech series blocks were then presented with new instructions to label the sounds as containing the vowel in "hood" (/ʊ/) or "hid" (/ɪ/). A short break was provided between each block and the entire session lasted approximately 40 min.

B. Results

The dependent variable of interest is the percentage of stimuli receiving responses corresponding to high-frequency stimuli (i.e., /ɪ/ and group 2 for speech and nonspeech series, respectively) collapsed across the 150 stimuli of each series. A context effect would be defined as a significant difference in percent categorization of /ɪ/ or group 2 between isolated (V or V-SW) and context condition (wV or wV-SW). The predicted direction of this effect would be an increase in "high-frequency" responses in the context conditions (counteracting the effects of vowel undershoot or a spectral contrast with the lower-frequency formants or tones in the context).

The first prediction to be tested is that speech stimuli will elicit a context effect. A paired-sample t-test revealed a significant difference in percent /ɪ/ responses obtained in the V (49.5%) and wV (65.6%) series [$t(19)=12.34$, $p<0.001$]. Figure 2(a) provides a graphic demonstration of the categorization shift. Categorization functions were computed for each subject by averaging the percent /I/ responses across each decade of stimuli (i.e., 1–10, 11–20, etc.) These individual categorization functions were averaged together to create the functions in Fig. 1(a). One can see a clear shift in the phonetic boundary due to context. These results replicate previous demonstrations of /w/ context-induced shifts on vowel categorization (Nabelek and Ovchinnikov, 1997).

The second prediction, based on the results of Williams (1986), was that nonspeech analogs would fail to induce a context effect. This prediction was supported by nonsignificant differences between the V-SW and wV-SW conditions in % of group 2 responses [65.1% vs 63.4%, $t(19)=2.01$, $p=0.06$]. (Note that the trend toward a significant effect for %

group 2 responses is actually in the opposite direction predicted for a context effect with more group 2 responses in the isolated condition.) This null effect is generally in line with the results of Williams (1986) with a slightly different stimulus set and a different task (arbitrary categorization as opposed to pitch judgments). However, the inability of nonspeech contexts to elicit categorization shifts in nonspeech targets is difficult to reconcile with demonstrations of nonspeech contexts influencing speech target categorization, such as those of Holt *et al.* (2000). Experiment 2 was designed to examine whether training on the nonspeech categorization task, with the presumed development of stable representations for these categories, will result in context-moderated categorization for nonspeech stimuli.

III. EXPERIMENT 2

In experiment 2, the same participants from experiment 1 were divided into "control" and "training" groups. The training group received additional exposure and feedback on the nonspeech categorization task. (In experiment 1, training consisted of only ten trials, five from each series endpoint). After training, the same categorization tasks were performed as in experiment 1 with the same dependent variables. The data from experiment 1 served as a comparison for any observed shifts subsequent to training in experiment 2. The hypothesis being examined here is the following: Well-formed target categories are necessary in order to obtain context effects. The first prediction resulting from this hypothesis is that, with no subsequent training, the control group will continue to fail to display context-moderated shifts in categorization performance for nonspeech stimuli. The second prediction is that the training group will show significant nonspeech context effects following training (whereas they showed no context effect in experiment 1).

A. Method

1. Subjects

The participants from experiment 1 were randomly assigned to the training and control groups. These groups did not differ on percent /I/ or percent high responses or boundary locations for any of the conditions in experiment 1 (tested with t-tests on each condition and dependent variable separately).

2. Stimuli

The stimulus set for the training was the V-SW (only steady-state sine-wave analogs were used) series from experiment 1. The post-training tasks included all four stimulus sets from experiment 1.

3. Procedure

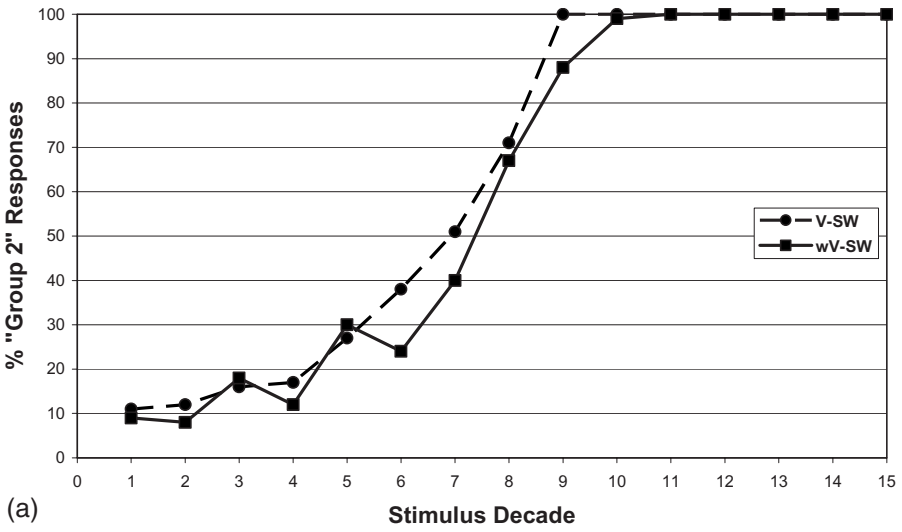
Members of the training group returned on a separate day from experiment 1 to begin training on the nonspeech categorization task. These participants were only trained on the V-SW stimuli. That is, they only received experience and feedback with the steady-state isolated sine-wave stimuli. A training session consisted of a single block during which each of the 150 stimuli was presented once in random order. All stimulus presentation details were equivalent to the description for experiment 1. After each stimulus presentation, the participant categorized the sound as a member of group 1 or group 2 by using a mouse to click a labeled response box on the computer monitor. After the response, feedback was provided visually (correct or wrong). The feedback was determined from a criterion that was specified separately for each participant as equal to their categorization boundary from experiment 1. These boundaries were defined as the stimulus number at which there was a change in response to at least three consecutive responses of the alternative category (remembering that each stimulus received only one categorization response from each participant). Correct responses were group 1 responses to stimulus steps less than their calculated boundary stimulus from experiment 1.

Members of the training group participated in three training sessions on consecutive days. Immediately following the third training session, a categorization test was performed. The test consisted of 30 stimuli randomly drawn from the 150 stimuli of the V-SW series. Participants categorized these stimuli with no feedback. If their performance was 90% correct or greater (correct relative to their experiment 1 boundary) then they were done with the training. If performance was lower than 90% then the participants returned for a subsequent training session. Testing followed training for every session starting at session 3. Training continued until 90% correct performance was obtained on the test or six training sessions were completed. Six of the ten subjects were able to obtain 90% correct identification in four sessions, while the four remaining subjects required five sessions to obtain 90% correct identification.

When training was completed for each participant in the training group they returned the following day (after training) to complete the categorization tasks for all speech and nonspeech sessions. Participants of the control group were asked to return the next day from the day they participated in their first set of categorization task to complete the categorization tasks for all (isolated and context) speech and nonspeech series. This session was identical to the experiment 1 session.

B. Results

The first prediction to be tested was that there would be no context effects for the nonspeech stimuli for the control group. These participants received no categorization training and it was expected that they would categorize the nonspeech sounds similarly to how they had in experiment 1 (no effect of context). There was a small but significant context-moderated difference in % group 2 responses [63.7% vs 60%, $t(9)=2.83$, $p<0.05$], but the shift was in the opposite direction obtained in normal context effects, as was the insignificant shift in experiment 1. The categorization functions for the no-training group are presented in Fig. 3(a). These results support the expectation that the control group would show a little change in categorization performance from experiment 1 to experiment 2.



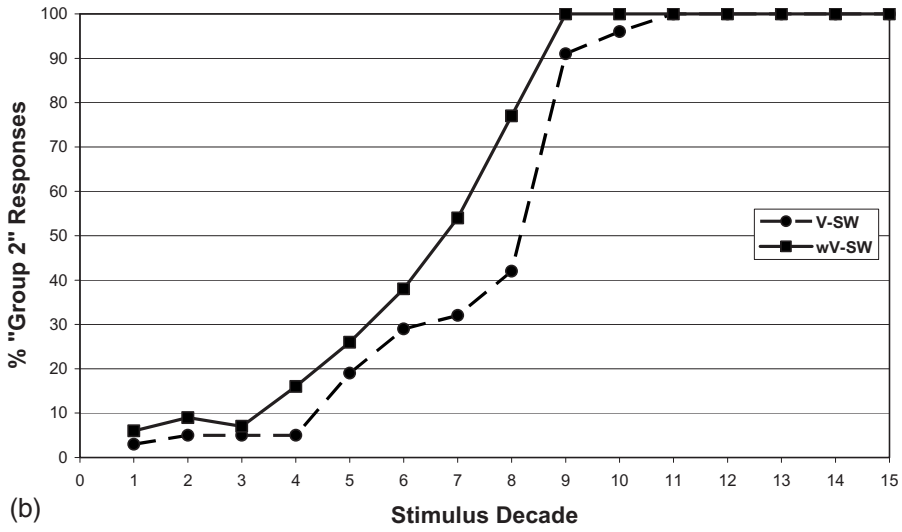


FIG. 3. Nonspeech series categorization functions for experiment 2. Mean percentage of group 2 responses are presented by stimulus decade for isolated (V-SW) and context (wV-SW) conditions. (a) Nontraining group; (b) training group.

Of greater importance for the hypothesis is the prediction that after training with nonspeech categorization, members of the training group would provide significant context shifts for the nonspeech stimuli. This prediction was supported by a significant context effect for % group 2 responses [55.1% vs 62.2%, $t(9)=4.88$, $p<0.001$]. The categorization functions for the training group are presented in Fig. 3(b). These results demonstrate that nonspeech contexts can shift the perception of nonspeech targets when there is adequate experience with the categorization task.

No explicit predictions were made regarding the speech series following training, but it was presumed that speech context effects would be maintained for both the training and control group. The shift in % /I/ as a function of context /w/ was maintained for both groups. [V: 48% vs wV: 66.6%, $t(9)=7.73$, $p<0.000$ (training group) and V:47.9% versus wV: 63.7%, $t(9)=7.67$, $p<0.000$ (control group)].

IV. DISCUSSION

The goal of the current study was to test a hypothesis that may reconcile the seemingly contradictory results of Williams (1986) and Holt *et al.* (2000). Williams (1986) demonstrated that nonspeech contexts had no influence on pitch judgments of nonspeech targets, but that context effects were obtained when the listeners were instructed to perceive the sounds as speech. On the other hand, Holt *et al.* (2000) reported that nonspeech contexts with no phonemic content could shift the categorization of speech targets. Why would nonspeech contexts be capable of influencing the perception of speech sounds but not the perception of similar nonspeech sounds? One hypothesis is that the well-developed categories for the target speech sounds (vowel phonemes) allowed the effects of context to become manifest, whereas the novel nonspeech stimuli and pitch judgment task of Williams (1986) did not have the necessary well-established perceptual categories. The two experiments described in this paper were designed to test predictions arising from this hypothesis.

Experiment 1 once again demonstrated that speech sound categorization is influenced by context, whereas novel nonspeech categorization tasks are not. Listeners categorized vowels (/ɪ/ versus /ʊ/) presented as isolated steady states and in the context of a preceding /w/. As demonstrated previously by Nabelek and Ovchinnikov (1997), the percentage of /ɪ/ responses increases in /w/ context. In contrast, the categorization of sine-wave complexes as group 1 and group 2 did

not differ between isolated and context (sine-wave transitions analogous to the formant transitions of /w/) conditions.

Experiment 2 examined whether training on the nonspeech categorization task would result in context-sensitive perception of the nonspeech targets. Participants were trained to categorize the isolated nonspeech targets based on a criterion (placed at their individual categorization boundaries from experiment 1) or received no subsequent training or exposure to the stimuli. The control group continued to show disparate patterns of categorization for speech and nonspeech stimuli. The training group, on the other hand, demonstrated context-dependent shifts in nonspeech categorization that were in the same direction as the speech context effects.

The results of experiment 2 demonstrate that context effects can be obtained for nonspeech stimuli. Holt *et al.* (2000) demonstrated the effects of nonspeech contexts on the categorization of speech sounds. Stephens and Holt (2003) obtained effects of speech contexts (/al/ and /ar/) on the discrimination of nonspeech sounds (distinguishing F2-F3 combinations filtered from /da/-/ga/ stimuli). The present study adds to this literature by exhibiting an effect of nonspeech context on nonspeech categorization. All of these context effects can be considered contrastive in the manner described by Holt and Lotto (2006). In the case of the current study, the frequency of the second tone, which models F2 in the speech stimuli, was the dimension of importance for the categorization of the isolated steady-state stimuli. The context sound consisted of tones increasing in frequency to the values of the steady-state tones. That is, the context included energy at a lower frequency than the steady-state second tone. The result according to Holt and Lotto (2006) would be that the critical tone would be perceived as effectively higher in frequency. This would lead to more group 2 (trained to correspond to the stimuli with higher second tone frequencies) responses in the context condition, which is exactly what was obtained.

The fact that these contrastive context effects were obtained with nonspeech contexts and targets are in line with the proposition that contrast is the result of the operating characteristics of the general auditory system (Holt and Lotto, 2002; Lotto *et al.*, 2003). If contrast is a consequence of the way that the auditory system encodes complex sounds then it likely plays some role in context effects, such as those presented by Lindblom and Studdert-Kennedy (1967), Nearey (1989), and Nabelek and Ovchinnikov (1997). A general auditory basis for this effect makes it less surprising that phonetic context effects occur in infants (Fowler *et al.*, 1990), have been demonstrated in birds (Lotto *et al.*, 1997), and appear to be independent of the native language of the listener (Mann, 1986; Gow and Im, 2004).

One of the pieces of evidence against the general perceptual nature of these effects was the lack of a context effect for the nonspeech pitch judgments in Williams, 1986. The results of the current study suggest that the failure to find context effects was not because the sounds were perceived as nonspeech, but because there were not well-established perceptual categories on which to base the pitch judgments. When listeners are trained to categorize the stimuli based on a fixed criterion (as in experiment 2), nonspeech context can affect nonspeech categorization as predicted by a spectral contrast account. This raises the question of what aspect of the training was responsible for the manifestation of context-sensitive categorization. We speculate that when listeners are asked to do a relative judgment task such as high or low pitch or group 1 versus group 2 with no established categories, they place their decision boundary to around the middle point of the (perceptual) stimulus range. Such a default strategy has been witnessed in other studies on the categorization of nonspeech sounds based on center frequency of noises (Sullivan *et al.*, 2005a, 2005b). Because this criterion is a function of the auditory (as opposed to acoustic) range of stimuli, it is free to shift with changes in the perceived range. In the context condition, according to the spectral contrast account, the effective pitch values of the critical second tones are increased relative to the lower-frequency context. If the entire range is affected then the midpoint of that range in terms of the acoustic value (stimulus step) is unchanged. The result is no shift in judgments due to context.

On the other hand, if one receives training in the form of feedback relative to a fixed acoustic boundary, the result is that the boundary is fixed to the corresponding auditory value, independent of the entire range. Sullivan *et al.* (2005a, 2005b) have shown that listeners can learn to shift their categorization criterion based on feedback rather quickly for novel nonspeech sounds. In the subsequent context condition, the shift in effective second tone pitch results in more stimuli being perceived as larger than the fixed boundary value. The result is more group 2 judgments and a significant context effect.

The results presented here are another indication of the fruitfulness of examining the roles of both the general-auditory and learning/cognitive systems when attempting to explain the perception of complex sounds such as speech (Holt *et al.*, 2004; Diehl *et al.*,. 2004). Lindblom (1996) had discussed the Williams (1986) study as demonstrating that perception is not "merely" about responding to an auditory signal, but is a function of "signal-plus-knowledge." That is, our interpretation of an acoustic stimulus is always a function of our peripheral representation of that signal and the application of a knowledge base relative to that signal. The present results support this point even more strongly than the original Williams (1986) study. The differences in perception of speech and nonspeech are not inherent in the nature of the signal; they are the result of our experience with and of the functional classification of these signals.

[1]This result allayed concerns that the effect was the result of auditory "capture" of some harmonics by the context tone (Dannerbring, 1976; Darwin, 1984; Darwin and Sutherland 1984).

[2]We have referred to the Williams (1986) tasks as phonetic categorization and pitch judgment to make the point about the differences in these two tasks.

[3]Nearey's (1989) rule for F3 calculation: For $F2 > 1500$, $F3_{front} = 0.522 \times F1 + 1.197 \times F2 + 57$; $F2 <= 1500$, $F3_{back} = 0.7866 \times F1 - 0.365 \times F2 + 2341$.

Bailey, P. J., Summerfield, A. Q., and Dorman, M. (**1977**). "On the identification of sine wave analogues of certain speech sounds," Haskins Laboratory Status Report on Speech Research, Report No. SR-51052, Yale University, New Haven, CT.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys. **29**, 191–211.
Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.
Darwin, C. J., and Sutherland, N. S. (**1984**). "Grouping frequency components of vowels: When is a harmonic not a harmonic?," Q. J. Exp. Psychol. A **36**, 193–208.
Dannerbring, G. L. (**1976**). "Percieved auditory continuity with alternately rising and falling frequency transitions," Can. J. Psychol. **30**, 99–115.
Diehl, R. L., Lotto, A. J., and Holt, L. L. (**2004**). "Speech perception," Annu. Rev. Psychol. **55**, 149–179.
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.
Gow, D. W., Jr., and Im, A. M. (**2004**). "A cross-linguistic examination of assimilation context effects," J. Mem. Lang. **51**, 279–296.
Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.
Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the neural mechanisms of speech context effects," Hear. Res. **167**, 156–169.
Holt, L. L., Lotto, A. J., and Diehl, R. L. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.
Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–990.
Liberman, A. M., and Mattingly, I. G. (**1989**). "A specialization for speech perception," Science **243**, 489–494.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.
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.
Lindblom, B. (**1996**). "Role of articulation in speech perception: clues from production," J. Acoust. Soc. Am. **99**, 1683–1692.
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 of 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.
Mullennix, J. W., Pisoni, D. B., and Goldinger, S. D. (**1988**). "Some effects of time-varying context on the perception of speech and nonspeech sounds," Report No. 14, Indiana University, Bloomington, IN.
Nabelek, A. K., and Ovchinnikov, A. (**1997**). "Perception of nonlinear and linear formant trajectories," J. Acoust. Soc. Am. **101**, 488–497.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.
Stephens, J. D. W., and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of nonspeech," J. Acoust. Soc. Am. **114**, 3036–3039.
Sullivan, S. C., Lotto, A. J., Newlin, E. T., and Diehl, R. L. (**2005a**). "Sensitivity to stimulus distribution characteristics in auditory categorization," J. Acoust. Soc. Am. **117**, 2596.
Sullivan, S. C., Lotto, A. J., Newlin, E. T., and Diehl, R. L. (**2005b**). "Sensitivity to changing stimulus distribution characteristics in auditory categorization," J. Acoust. Soc. Am. **118**, 1896.
Williams, D. R. (**1986**). "Role of dynamic information in the perception of coarticulated vowels," Ph.D. thesis, University of Connecticut, Stanford, CT.