Date: 5/2/2008 11:51:05 AM

Journal Title: Perception and Psychophysics

Volume: 62
Issue: 8
Month/Year: 2000
Pages: 1668-1680

Article Author: Francis, A.L., Nusbaum, H.C., & Baldwin, K.

Article Title: Effects of training on attention to acoustic cues

Imprint:

Perception & Psychophysics
2000, 62 (8), 1668-1680

Effects of training on attention to acoustic cues

ALEXANDER L. FRANCIS
University of Hong Kong, Hong Kong

and

KATE BALDWIN and HOWARD C. NUSBAUM
University of Chicago, Chicago, Illinois

Learning new phonetic categories in a second language may be thought of in terms of learning to focus one's attention on those parts of the acoustic–phonetic structure of speech that are phonologically relevant in any given context. As yet, however, no study has demonstrated directly that training can shift listeners' attention between acoustic cues given feedback about the linguistic phonetic category alone. In this paper we discuss the results of a training study in which subjects learned to shift their attention from one acoustic cue to another using only category-level identification as feedback. Results demonstrate that training redirects listeners' attention to acoustic cues and that this shift of attention generalizes to novel (untrained) phonetic contexts.

Although initial studies of second language phonological learning by adults reported that training can have very little effect on accuracy (Lisker, 1970; Lisker & Abramson, 1970; Strange & Dittman, 1984; Werker & Tees, 1984; see Strange & Jenkins, 1978, for a review), later studies demonstrated successful second language phonological training (McClaskey, Pisoni, & Carrell, 1983; Pisoni, Aslin, Perey, & Hennessy, 1982; Tees & Werker, 1984).

Even relatively difficult phonological contrasts can be learned with appropriate training techniques (Lively, Pisoni, & Logan, 1991; Logan, Lively, & Pisoni, 1991). One explanation of the importance of training methods in learning different contrasts is that training must be designed to shift attention *to* unfamiliar acoustic cues and *away* from familiar cues. For example, Yamada and Tohkura (1991) found that Japanese listeners who performed poorly at distinguishing between English /r/ and /l/ did so because they were attending to different acoustic cues from those that are used by native English speakers. On the basis of these and related findings, a number of researchers (e.g., Best, 1994; Jusczyk, 1997, pp. 220–221; Nusbaum & Goodman, 1994; Nusbaum & Lee, 1992; Pisoni, Lively, & Logan, 1994) have characterized phonetic category learning as shifting attention to relevant acoustic cues.

Recent work on native phonological development presents converging support for the hypothesis that phonetic categories are learned by shifting attention to differentially weight acoustic cues. Nittrouer and her colleagues (Nittrouer & Crowther, 1998; Nittrouer & Miller, 1997a, 1997b) have demonstrated that, in identifying fricatives (/s/ vs. /sh/), 5- to 7-year-old English-speaking children tend to rely more strongly than adults on vowel formant transition cues, and less strongly than adults on the peak of the frequency spectrum of the fricative noise. Thus, over the course of development, American-English-speaking children appear to shift attention to give more weight to the structure of fricative noise, and less to vowel formant transitions, in classifying fricatives.

Research on learning to understand synthetic speech (in which existing acoustic cues may be uninformative or misleading) also suggests that perceptual learning of speech may involve shifting attention to change the importance of acoustic cues (Nusbaum & Lee, 1992). Since the interpretation given to one acoustic cue affects the interpretation of the others (e.g., Best, Morrongiello, & Robson, 1981; Carden, Levitt, Jusczyk, & Walley, 1981; Repp, 1982), the irrelevant or misleading cues will have to be actively discounted in favor of other acoustic cues also available. However, shifting attention to informative cues and away from misleading cues reduces the number of incorrect hypotheses about the linguistic category of a speech sound, thereby increasing the effective use of cognitive resources (see Nusbaum & Goodman, 1994; Nusbaum & Lee, 1992; Nusbaum & Magnuson, 1997; Nusbaum & Schwab, 1986).

Although there are differences between the tasks of learning to understand synthetic speech and learning to understand a new language, these tasks are sufficiently similar in that both can be characterized as shifting attention to linguistically informative acoustic cues. Nusbaum and Lee (1992) argued that understanding how listeners learn to shift their attention as a result of perceptual learning of synthetic speech may provide insight into the basic attentional mechanisms involved in learning new phonetic

The authors thank John Kingston, Elaine J. Francis, and an anonymous reviewer for their comments on a previous version of this article. This work derives from K.B.'s M.A. thesis. Correspondence should be addressed to A. L. Francis, Department of Speech and Hearing Sciences, University of Hong Kong, Prince Philip Dental Hospital, 5 F, 34 Hospital Road, Hong Kong (e-mail: afrancis@hkusua.hku.hk).

categories. However, most recent research making claims about attention in phonetic category acquisition primarily involves observing the differences between groups of subjects that have different degrees or types of linguistic experience. For example, Nittrouer and her colleagues studied children at different stages of linguistic development, and Yamada and Tohkura compared the performance of native speakers of Japanese with that of native speakers of American English. In both cases the more experienced group of listeners (older children, native American English speakers) learned to attend to the appropriate cues for their native language through years of exposure over the course of their physiological and psychological development from infancy, and it is not clear from these studies how listeners learn to distribute attention during first or second language learning, or even in learning synthetic speech. Indeed, there is currently no evidence to show that listeners are in fact able to learn to shift their attention at the level of acoustic cues given feedback only about the linguistic phonetic category. Although we know the outcome of learning, there is little direct evidence characterizing the specific mechanisms of transition (see Nusbaum & Goodman, 1994).

Phonetic categories, especially stop consonants, are perceived more or less categorically by adults (Liberman, 1970), so it may be difficult to encourage adult learners to shift their attention between acoustic cues since these cues are not directly perceptually available to the listener and may even be perceptually integral (Sussman, Fruchter, Hilbert, & Sirosh, 1998; Sussman & Shore, 1996). Techniques such as perceptual fading (Jamieson & Morosan, 1986, 1989) have been used successfully to highlight the relevant cues for listeners. Other techniques have been used to shift the level of a listener's attention from the phonological category to that of fine phonetic structure (as discussed by Werker, 1994; see also Logan & Pruitt, 1995, for a review of laboratory training techniques). But these approaches require shaping attention to acoustic cue structure by external means (physical highlighting or contextual structure), which may not always be available in language acquisition outside the laboratory. While such perceptually focused instruction may aid in learning difficult non-native contrasts in a laboratory setting isolated from the process of acquiring an entire language system, this kind of instruction is not available in first language acquisition, or in most cases of second language acquisition. However, feedback may well be available to language learners at the level of phonological categories (e.g., "Did you say *rake* or *lake*?") Thus it is important to understand whether feedback at the level of phonological categories can shift attention to relevant acoustic cues.

We carried out a training study in which listeners were given category-level feedback to shift attention from one acoustic cue to another. The two cues used are properties that normally cooperate to specify the place of articulation of consonants: (1) the origin and direction of the formant transitions between consonantal release and a following vowel (Delattre, Liberman, & Cooper, 1955) and (2) the shape of the frequency spectrum at burst onset, approximately the first 25 msec after release (Blumstein & Stevens, 1980; Cole & Scott, 1974). A pretest–posttest design enabled us to gauge subjects' initial cue use and observe changes in their responses resulting from learning. In order to test the generality of any learning that occurred, subjects were trained on only a part of the total stimulus set and were then tested on the entire set. If phonetic category learning is accomplished by adjusting the distribution of attention to acoustic cues, then we would expect training to increase attention to the cue that listeners are trained to use, while also decreasing attention to the untrained cue (because it varies freely in relation to the trained cue). Prior results (Walley & Carrell, 1983) demonstrated that formant transitions determine place of articulation when transition and burst cues are in conflict. We expected that subjects would initially rely more on formant transitions, but that with appropriate feedback they might shift their attention toward the burst-onset spectrum.

METHOD

Subjects

Thirty-six undergraduate and graduate students at the University of Chicago were paid to participate in this experiment. None of the subjects had any reported history of speech or hearing difficulties, and all were native English speakers between the ages of 18 and 35. Subjects were paid $20 for approximately 3 h of participation. Subjects were randomly assigned to be trained on one of two kinds of training, either formant-cued training or onset-cued training.

Stimuli

Two sets of stimuli were constructed following the methods used by Walley and Carrell (1983). *Cooperating-cue stimuli* were syllables in which the spectrum of the burst release and the origin of the formant transitions were mutually consistent: both provided cues to the same initial consonant. Transition and formants cuing /b/, /d/, and /g/ were combined with the vowels /a/, /i/, and /u/ to make nine syllables. *Conflicting-cue stimuli* consisted of syllables in which the release burst cued the perception of a consonant that was different from that cued by the pattern of the formant transitions. Bursts with rising power spectra, flat or falling spectra, and midfrequency peaks (see Blumstein & Stevens, 1979; Walley & Carrell, 1983) were combined with rising, falling, or diverging patterned transitions to make 18 conflicting-cue patterns combined with the vowels /i/, /a/, and /u/. For each of the nine possible consonant-vowel (CV) syllable combinations of /b d g/ and /a i u/, one cooperating-cue version and two conflicting-cue versions were constructed using the Klatt (1980) speech synthesizer in parallel resonance mode.

All three versions of each CV syllable (e.g., /ba/) had the same formant transitions, formant bandwidths, and steady-state vowel portion. The only difference was in the burst release. In the cooperating-cue stimulus, the burst release was appropriate for the syllable /ba/, whereas in the two conflicting-cue stimuli the burst was appropriate for either a /d/ or a /g/, respectively. The duration of voicing for each syllable was 255 msec. The amplitude of voicing was a constant 60 dB for the first 220 msec, falling linearly to 0 dB over the last 35 msec of the syllable. Fundamental frequency increased from 103 to 125 Hz over the first 35 msec. Subsequently $f0$ decreased linearly to 95 Hz over the next 180 msec. Over the last 40 msec of the syllable, $f0$ decreased from 95 to 50 Hz. Formant bandwidths for all stimuli were also identical, as follows: $F1$, 50 Hz; $F2$, 70 Hz; $F3$, 110 Hz; $F4$, 170 Hz; $F5$, 250 Hz. Stimuli were synthesized at a sampling rate of 10 kHz and low-pass filtered at 4.8 kHz. All syllables using the vowels /a/ and /u/ were constructed using the transitions and formant amplitude information taken directly from values reported for the stimuli used by Walley and Carrell (1983), while the /i/ syllables were constructed according to the same principles.

The center frequency of all fourth and fifth formants remained constant throughout the duration of the syllable. $F5$ was set at 4500 Hz. $F4$ frequencies varied depending on the consonant or vowel in the syllable. In /a/ syllables, $F4$ was 3600 Hz, while for /u/ syllables, $F4$ was 3200. F4 for the syllables with /i/ varied depending on the formant transitions. Following the formant transitions of /b/ and /g/, $F4$ was 3600 Hz. $F4$ was set at 3800 Hz for the /i/ syllables with the formant transitions of /d/. The formant amplitudes of each /i/ stimulus were individually set to give the appropriate onset spectrum contours according to linear predictive coding (LPC) analysis (Markel & Gray, 1976). The altered amplitude values remained constant for the first 25 msec of the stimulus and then were linearly interpolated over the following 25 msec to 76 dB, where they remained for the rest of the stimulus. The only difference between versions was in the amplitude of the formants (and thus the shape of the onset spectrum) during the first 50 msec after the burst. Formant amplitudes in the vowel did not differ between syllables. No other aspects of the original stimuli from Walley and Carrell (1983) were changed. All formant values for all syllables are shown in Table 1.

Spectral analysis of each stimulus was performed to verify that all stimuli had the intended spectral characteristics. Spectral information about the burst release was extracted by LPC, with 14 linear prediction coefficients using a preemphasis coefficient of .5 on a 25.6-msec Hamming window centered at the onset of each syllable. The results of this analysis are given in Figures 1–3.

After synthesis, all 27 syllables were checked again against the templates and matching criteria defined by Blumstein and Stevens

Table 1
Parameters Used to Generate All Stimuli Using Klatt Synthesizer

Stimulus	Formant	Time 1 (msec)	Frequency (hz)	Time 2 (msec)	Frequency (hz)
/ba/	$F1$	0–20	220–720	20–255	720
	$F2$	0–40	900–1240	40–255	1240
	$F3$	0–40	2000–2500	40–255	2500
	$F4$	0–255	3600		
	$F5$	0–255	4500		
/da/	$F1$	0–35	220–720	35–255	720
	$F2$	0–40	1700–1240	40–255	1240
	$F3$	0–40	2800–2500	40–255	2500
	$F4$	0–255	3600		
	$F5$	0–255	4500		
/ga/	$F1$	0–45	220–720	45–255	720
	$F2$	0–40	1640–1240	40–255	1240
	$F3$	0–40	2100–2500	40–255	2500
	$F4$	0–255	3600		
	$F5$	0–255	4500		
/bi/	$F1$	0–15	180–330	15–255	330–270
	$F2$	0–40	1800–2000	40–255	2000–2300
	$F3$	0–40	2600–3000	40–255	3000
	$F4$	0–255	3600		
	$F5$	0–255	4500		
/di/	$F1$	0–20	180–330	20–255	330–270
	$F2$	0–40	1800–2000	40–255	2000–2300
	$F3$	0–40	2800–3000	40–255	3000
	$F4$	0–255	3800		
	$F5$	0–255	4500		
/gi/	$F1$	0–30	180–330	30–255	330–270
	$F2$	0–40	2400–2000	40–255	2000–2300
	$F3$	0–255	3000		
	$F4$	0–255	3600		
	$F5$	0–255	4500		
/bu/	$F1$	0–15	180–370	15–255	370–300
	$F2$	0–40	800–1100	40–255	1100–1000
	$F3$	0–40	2000–2350	40–255	2350
	$F4$	0–255	3200		
	$F5$	0–255	4500		
/du/	$F1$	0–15	180–370	15–255	370–300
	$F2$	0–40	1600–1100	40–255	1100–1000
	$F3$	0–40	2700–2350	40–255	2350
	$F4$	0–255	3200		
	$F5$	0–255	4500		
/gu/	$F1$	0–15	180–370	15–255	370–300
	$F2$	0–40	1400–1100	40–255	1100–1000
	$F3$	0–40	2000–2350	40–255	2350
	$F4$	0–255	3200		
	$F5$	0–255	4500		

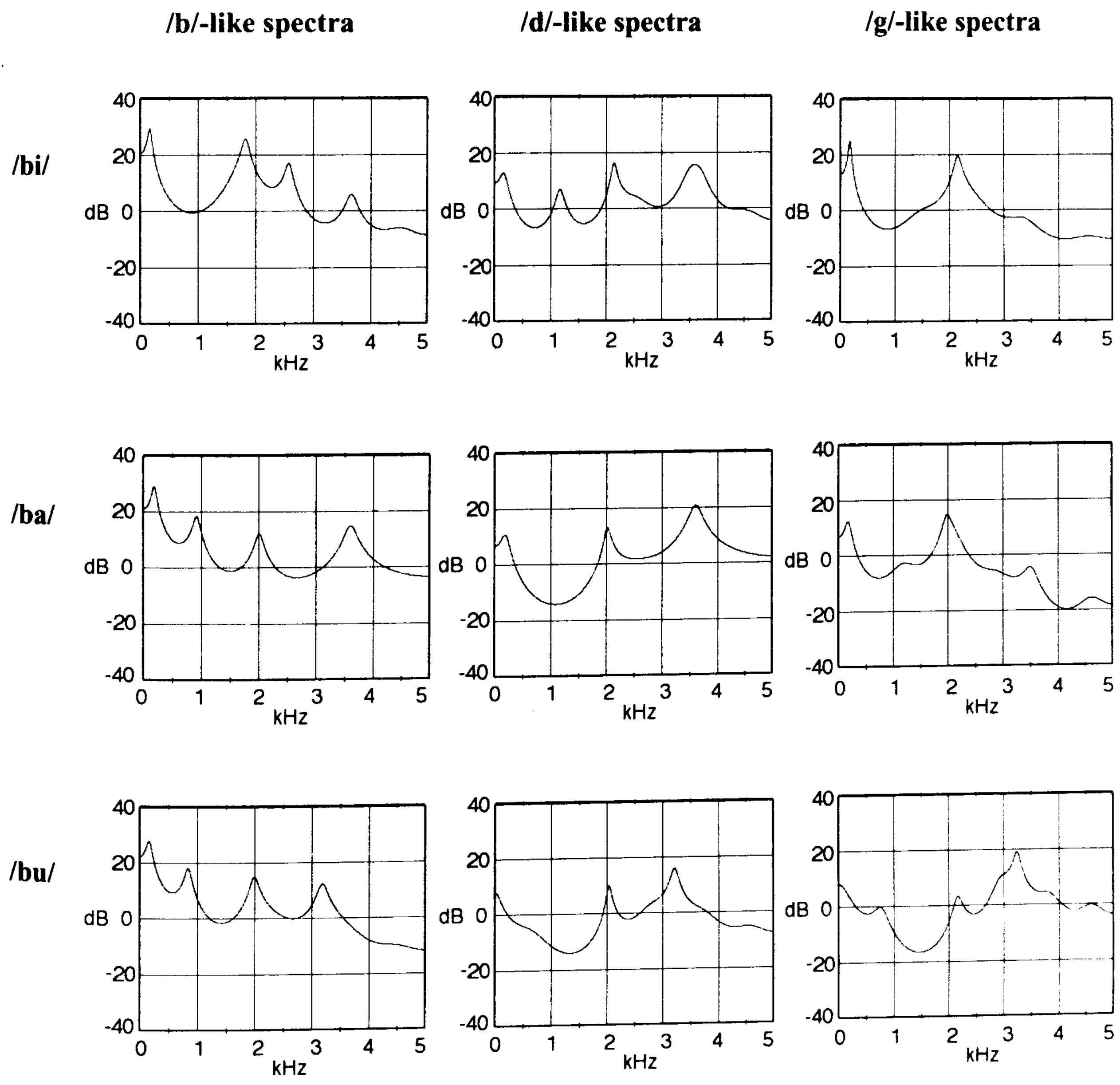

Figure 1. Onset spectra for syllables with /b/ formant transitions, as measured by linear predictive coding using14 coefficients, a preemphasis of 0.5, and a 25.6-msec Hamming window centered at syllable onset. The figure is organized according to the syllable consistent with the formant transitions (rows) and the segment consistent with the burst frequency distribution (columns). In each plot, frequency ranges from 0 to 5000 Hz from left to right and amplitude ranges from −40 to 40 dB SPL from bottom to top.

(1979) and shown in Figure 4. In the cooperating-cue version of each stimulus, the shape of the onset spectrum fit into the template shown by Blumstein and Stevens to be associated with the place of articulation specified by the transitions. The conflicting-cue versions each had an onset spectrum that fit one of the Blumstein and Stevens templates.

Procedure

The experiment was conducted during three 1-h sessions on consecutive days. Subjects were randomly assigned to two different groups, the formant-trained group and the onset-trained group. Sessions were conducted with groups of 1 to 3 subjects, seated in individual sound-attenuated booths in front of a response keyboard and monitor. All stimuli were presented to subjects over Sennheiser HD430 headphones at a comfortable listening level (approximately 66 dB). Stimuli were presented in real time under computer control at a 10-kHz sampling rate through a 12-bit D/A converter and low-pass filtered at 4.8 kHz. Stimuli were randomized within each block and responses were recorded on a computer-controlled keyboard. The assignment of responses to response keys was counterbalanced across subjects.

The experiment consisted of a pretest, followed by training, followed by a posttest, as shown in Table 2. The session on Day 1 consisted of the pretest and the first training session; Day 2 consisted of the second training session; on Day 3, subjects had a third training session and then took the posttest. Subjects were given 10 familiarization trials before the pretest. The stimuli for the familiarization trials were randomly chosen from the set of all stimuli and were

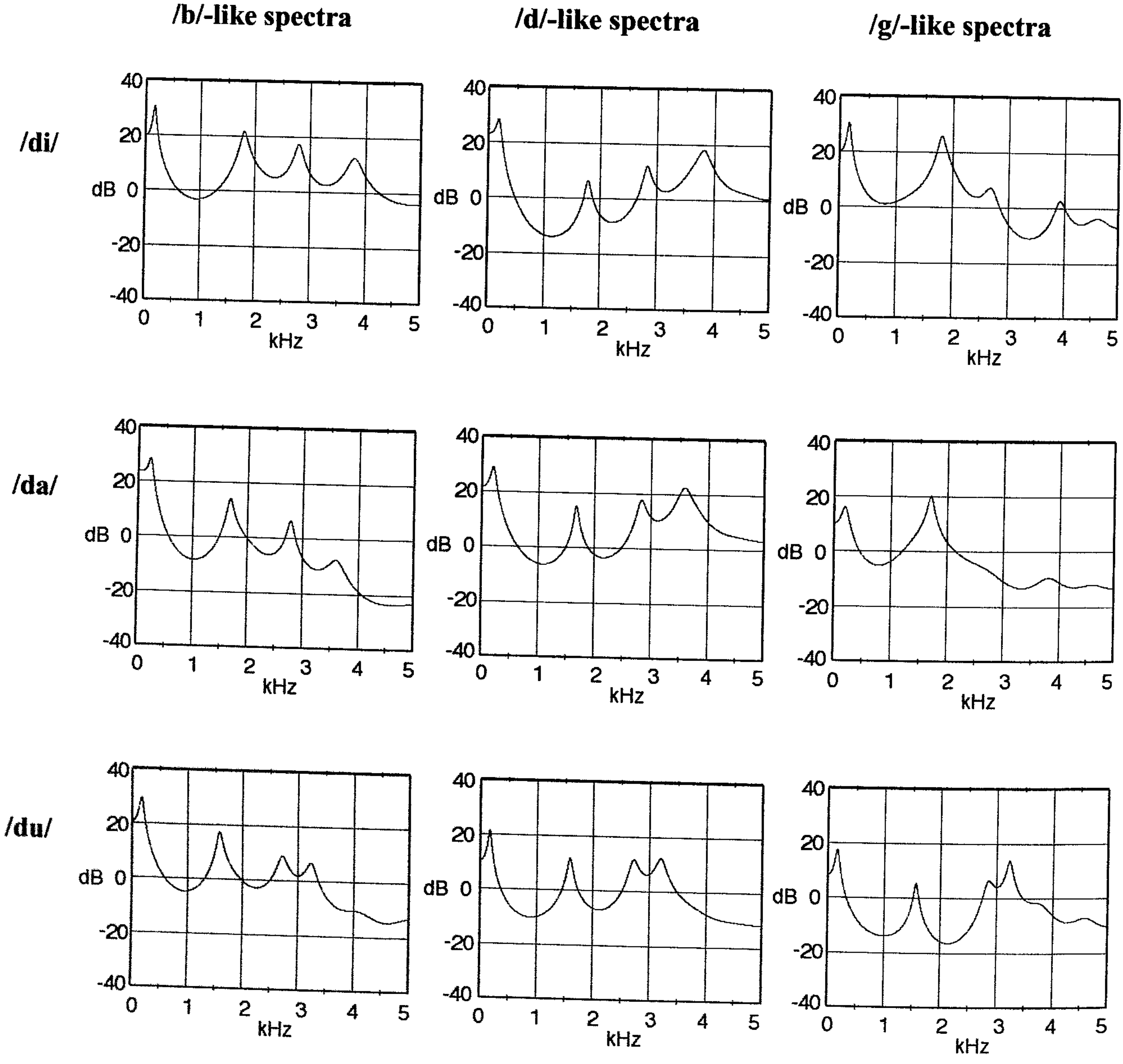

Figure 2. Onset spectra for syllables with /d/ formant transitions (see Figure 1 caption for details).

the same for all subjects. No feedback was given during familiarization trials, and subjects were simply instructed to listen to each sound they heard. The pretest and posttest each consisted of 20 presentations of each of the 27 stimuli (540 trials per test). Each testing session was divided into two blocks of 10 repetitions of each stimulus (270 trials), and stimuli were presented in randomized order within each block. Following the posttest, subjects were also asked to fill out a questionnaire concerning their subjective impressions of the stimuli and how difficult they found the learning task at the beginning and end of the experiment.

The stimuli containing the vowels /a/ and /i/—the *training set*—were presented to subjects during all three phases of the experiment: pretest, training, and posttest. Syllables containing /u/—the *generalization set*—were presented to subjects only during the pretest and posttest.

Each trial within the pretest and posttest phases of the experiment consisted of a single presentation of a syllable and a three-alternative (b, d, or g) forced-choice identification response. If no response was made during a 3-sec interval, a null response was recorded and the next trial started.

After the pretest, familiarization with the 18 /a/ and /i/ stimuli was repeated. Subjects were not required to respond, and were instructed to listen to the syllables. Following familiarization, subjects were trained on a total of 24 repetitions of each /a/ and /i/ syllable. Subjects received six training blocks, each consisting of four presentations of each of the 18 stimuli in random order. The first two training blocks were given on Day 1, immediately following the pretest; the next three training blocks were run on Day 2, and the final training block was presented on Day 3, immediately preceding the administration of the posttest.

During training subjects received visual and auditory feedback. After responding, subjects heard a second auditory presentation of the syllable paired with a printed version of the syllable on the computer display (/ba/, /da/, /ga/, /bi/, /di/, or /gi/). No response was

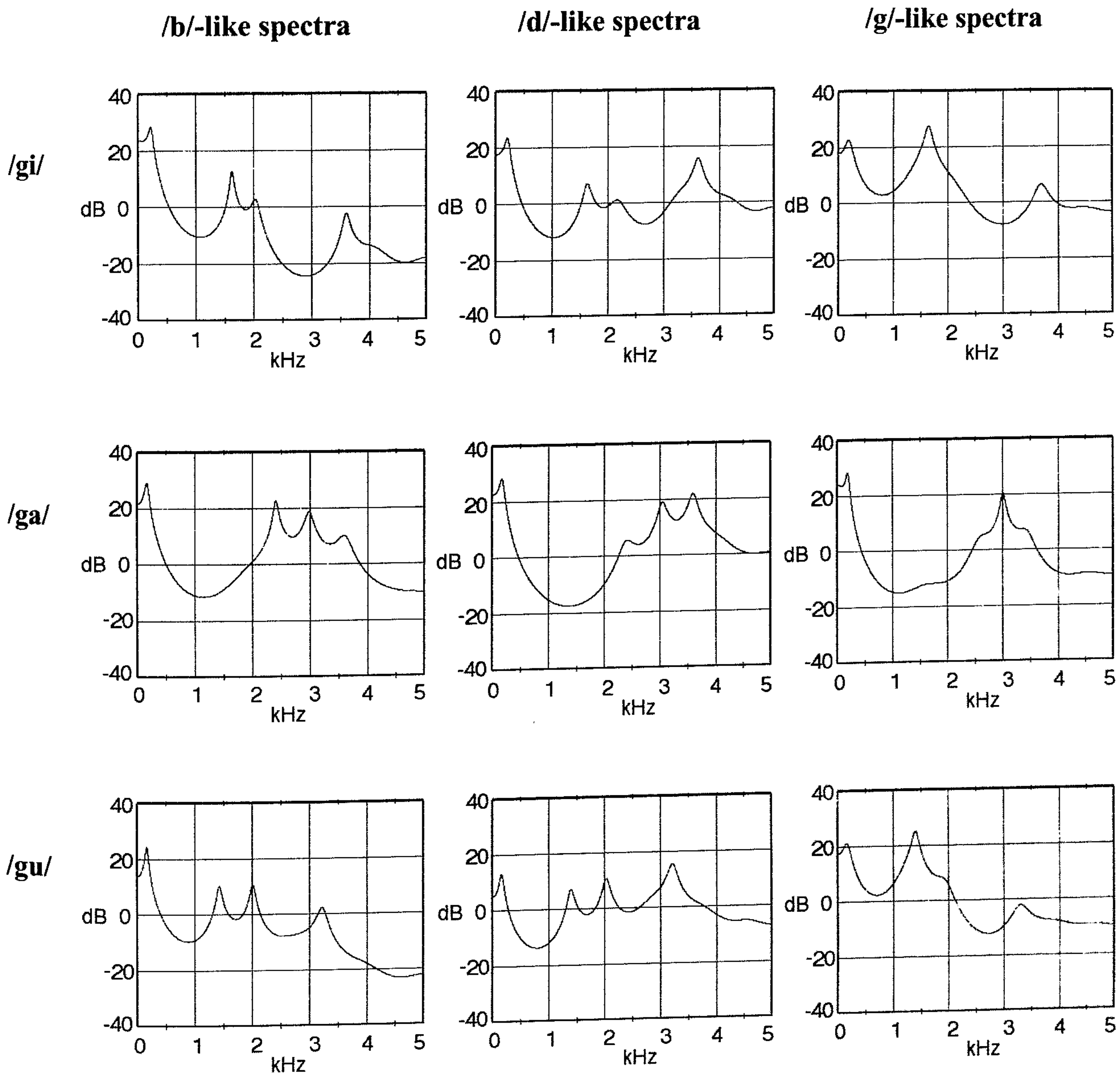

Figure 3. Onset spectra for syllables with /g/ formant transitions (see Figure 1 caption for details).

required to this second presentation. Only the identity of the printed syllable varied between experimental groups. For the formant-trained group, the printed consonant was the consonant cued by the formant transitions. In contrast, for the burst-trained group, the printed consonant was the consonant cued by the spectrum of the burst. Subjects were specifically encouraged to make use of any cues they found helpful in learning the correct identification of each syllable, though no reference was made to any acoustic qualities of the stimuli. They were also given frequent encouragement to keep trying in the face of slow progress.

RESULTS

Each identification response for the conflicting-cue stimuli was scored according to its match with one of three categories—trained, untrained, or other. Trained responses were those in which the subject's response was consistent with use of the cue on which the subject was trained. For example, if a subject in the formant-trained group responded "b" to a token in which the formant cue was consistent with /b/ but the burst cue was consistent with /d/, then that response would be scored as a *trained* response. In contrast, *untrained* responses were those in which the response was consistent with the place of articulation cued by the cue on which the subject was *not* trained. So, in our example, if the formant-trained subject had responded *d* (consistent with the burst cue, but inconsistent with the formant cue on which he/she had been trained), then this response would have been scored as *untrained*. Responses consistent with neither cue were classified as *other*.

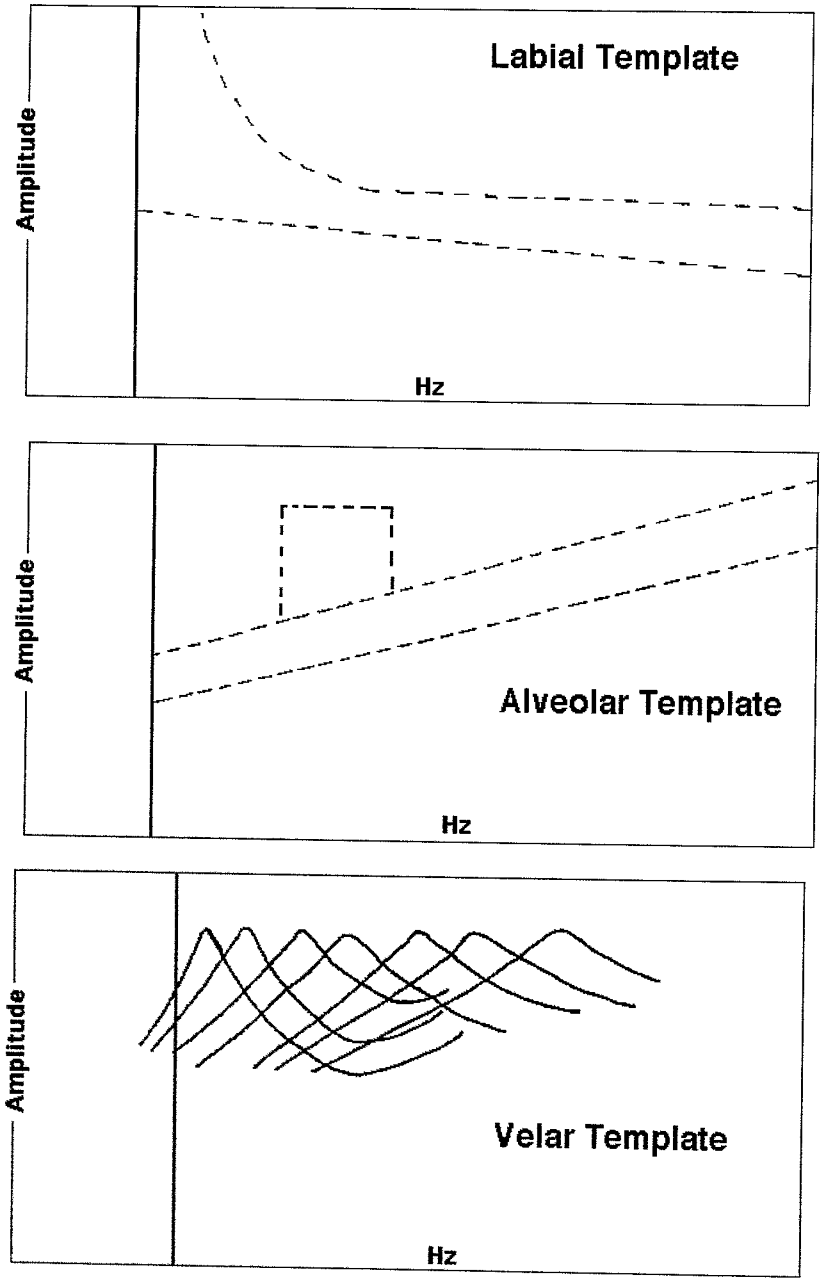

Figure 4. Templates constructed according to parameters described by Blumstein and Stevens (1979). Spectra of burst releases at each place of articulation fit within the dashed lines of the respective template (solid/curved lines in velar template) when aligned with the solid vertical line. Labial burst spectra (top) are characterized by a diffuse falling spectrum (multiple distinct peaks falling in amplitude from left to right). Alveolar burst spectra (middle) are characterized by a diffuse rising spectrum (multiple distinct peaks rising in amplitude from left to right). Velar burst spectra (bottom) are characterized by a compact spectrum (one dominant peak). The actual fitting process is complex, and interested readers are referred to Blumstein and Stevens (1979). Note: due to formatting constraints, templates are not to scale with spectra in Figures 1–3.

Note that response percentage is not a measure of accuracy, since the stimuli are composed of conflicting cues and subjects are no more or less accurate for choosing to attend to one cue over another. However, increasing the percentage of responses that are consistent with the training condition can be considered learning, because it indicates that listeners have learned to respond more frequently on the basis of training. In contrast, subjects' percent correct scores on the cooperating-cue stimuli can be considered a measure of accuracy. When both

Table 2
Description of Experimental Procedure

Day	Session	Block	Stimuli	Total Trials	Response	Feedback
1	Pretest	Familiarize	One each of 10 random tokens	10	none	none
		Two test blocks	Per block, 10 each of 27 tokens (conflicting and cooperating cue)	540	3 AFC (b, d, or g)	none
	Training	Familiarize	One each of 18 training tokens (/a/ and /i/ syllables): 6 cooperating cue, 12 conflicting cue	18	none	none
		Two training blocks	Per block, 4 each of 12 training and 6 cooperating cue syllables	144	3 AFC (b, d, or g)	b, d, or g, with trained cue
2	Training	Three training blocks	Per block, 4 each of 12 training and 6 cooperating cue syllables	216	3 AFC (b, d, or g)	b, d, or g, with trained cue
3	Training	One training block	Per block, 4 each of 12 training and 6 cooperating cue syllables	72	3 AFC (b, d, or g)	b, d, or g with trained cue
	Posttest	Familiarize	One each of 10 random tokens	10	none	none
		Two test blocks	Per block, 10 each of 27 tokens (conflicting and cooperating cue)	540	3 AFC (b, d, or g)	none

cues agree, subjects should be able to identify the consonant correctly according to either or both of the cues.

Learning

Performance on the training-set stimuli for the pretest and posttest demonstrates that subjects in both groups clearly learned to respond more appropriately based on their training. Before training on the training set of syllables, formant-trained and burst-trained subjects initially responded according to their respective trained cue at a rate of .40, whereas after training their rate of response was .52 [$t(35) = 7.014, p < .001$].[1] Furthermore, subjects clearly *decreased* their reliance on the cue on which they were not trained, as shown by the decrease in untrained responses from .39 (pretest) to .28 (posttest) [$t(35) = 5.630, p < .001$]. Both of these response patterns suggest that phonetic learning involves increasing the weight given to useful cues while decreasing the weight given to less useful cues. These changes are shown for both groups in Figure 5 for the conflicting-cue stimuli.

Subjects in the formant-trained condition decreased their proportion of untrained (burst) responses by 11 percentage points, from .23 to .12 [$t(17) = 6.923, p < .001$], and subjects in the burst-trained condition similarly decreased their proportion of untrained (formant) responses by 11 percentage points, from .56 to .45 [$t(17) = 2.672$, $p = .02$]. In contrast, the increase in trained responses was not as large in the burst-training condition as it was in the formant-training condition. Formant-trained listeners increased their trained (formant) responses by 18 percentage points, from .58 to .76 [$t(17) = 9.333, p <$ .001], while subjects in the burst-training condition increased their trained (burst) responses by only 6 percentage points, from .22 to .28 [$t(17) = 2.906, p = .01$].

On the cooperating-cue stimuli, formant-trained listeners showed an overall improvement in recognition of 10 percentage points, from .71 (SE = .03) to .81 (SE = .03) [$t(17) = -3.614, p = .002$]. This suggests that listeners in this condition were able to use their improved attention to formant cues to facilitate their recognition of stimuli. In contrast, burst-trained listeners showed much less improvement (2 percentage points) on cooperating-cue stimuli, from .66 (SE = .03) to .68 (SE = .03), and this difference was not significant [$t(17) = -.435, p =$ .669]. This pattern of responses may indicate that the burst cue is simply less salient.

Generalization

To assess whether successfully learning to attend to particular cues in one context affects the distribution of attention to those cues in a different context, further analyses were performed only on data from those subjects who actually demonstrated learning (an increase in the trained response percentage). According to this criterion, 29 subjects were *learners* (17 of 18 in the formant-trained group and 12 of 18 in the onset-trained group). These

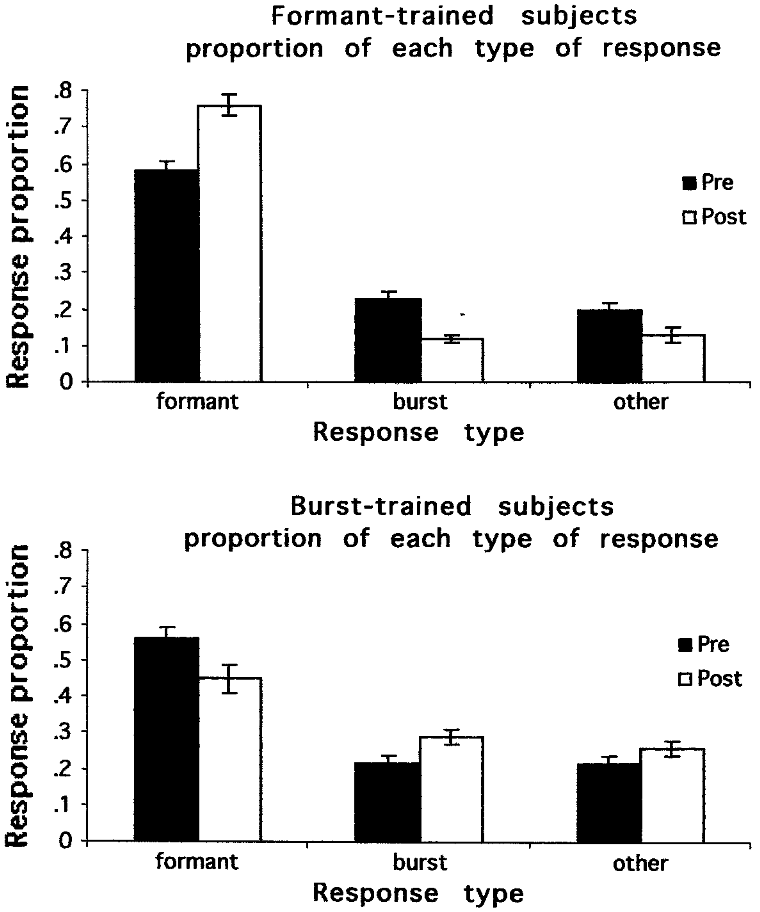

Figure 5. Response percentage for training-set syllables. Scores are displayed for subjects in each training condition (formant and burst) and indicate the percentage of responses that subjects made that were consistent with the cue on which subjects were trained, consistent with the cue on which they were not trained, and not consistent with either cue (other). Error bars indicate standard errors. All subjects.

learners' pretest and posttest scores for the generalization set are displayed in Figure 6. As a group, nonlearner subjects showed little change between their pretest and posttest response proportions. On the training set, stimuli responses were (pretest/posttest) as follows: trained, .34/.33; untrained, .46/.49; other, .19/.19. Generalization set scores (pretest/posttest) as follows: trained: .33/.30; untrained, .42/.44; other, .24/.26.

As shown in Figure 6, subjects showed smaller changes in response percentages for the generalization stimuli than for the training stimuli. However, all the changes are in the predicted direction. That is, though subjects show a smaller increase in their percentage of trained responses on the generalization stimuli than they do on the training stimuli, they still show an increase from .40 to .44 [$t(28) = -2.300$, $p = .03$].[2] Similarly, subjects showed the predicted decrease in untrained responses, from .36 to .30 [$t(28) = 2.504$, $p = .02$], and other responses did not change significantly (from .24 to .26) [$t(28) = -1.318$, $p = .198$]. Thus, listeners who learned to distribute attention more strongly to the cues they were trained on (and away from misleading cues) also extended this listening strategy to vowel contexts on which they were not trained.

DISCUSSION

The results of this experiment demonstrate that it is possible to induce a change in the way subjects attend to the acoustic cues to consonantal place of articulation using only category-level feedback. Subjects learned to withdraw attention from one set of cues and focus selectively on cues that were more useful for responding based on feedback. There are, however, two observations that must still be accounted for.

First, one explanation of any phonetic learning study using synthetic stimuli is that subjects are learning to attach linguistic labels to anomalous auditory stimuli, without hearing these stimuli as speech. The observation that

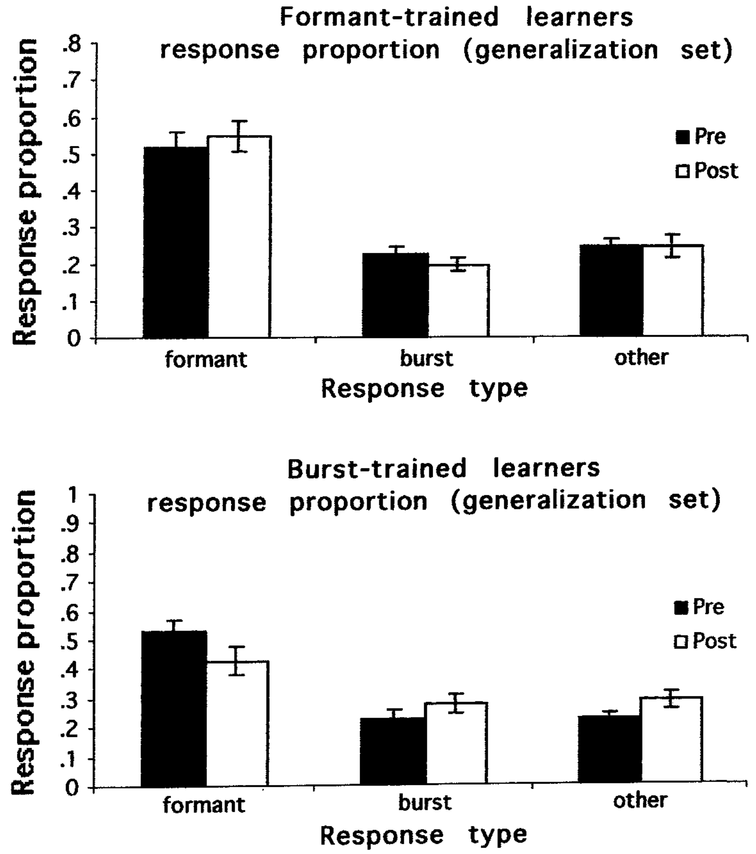

Figure 6. Response percentage for generalization-set syllables. Scores are displayed for subjects in each training condition (formant and burst) and indicate the percentage of responses that subjects made that were consistent with the cue on which subjects were trained, consistent with the cue on which they were not trained, and not consistent with either cue (other). Error bars indicate standard errors. Learners only (see text).

subjects are able to generalize their training to improve on stimuli on which they were not trained suggests that subjects were not simply learning labels for individual and unrelated acoustic patterns, but were applying their knowledge of acoustic–phonetic cue structure to these stimuli.

Furthermore, if subjects were not listening to these stimuli as speech, they would have to learn to identify each stimulus by its individual and idiosyncratic acoustic properties. The difficulty of the task could then be explained in terms of the strong acoustic similarities between the stimuli. Such an analysis would predict that all subjects should have equal difficulty in learning to assign linguistic labels to nonspeech sounds. However, the two experimental groups (formant-trained and burst-trained) had very different degrees of success with their respective learning task, and subjective ratings of task difficulty correlate with the observed greater difficulty in increasing the percentage of trained responses in the burst training condition. This suggests that the learning task presented to the burst-trained subjects was in some way harder than that presented to the formant-trained group. Since the two groups of subjects were exposed to the same stimuli for the same number of trials, any differences in response patterns cannot be due to task difficulty if learning depends only on the (nonspeech) perceptual discriminability of the stimuli.

Thus, the difference in success of learning between formant-trained and burst-trained subjects indicates that the two groups were learning to attend differently to the acoustic cues present in the sounds to which they were equally exposed—acoustic cues that are modeled on the cues actually present in natural speech. The strong tendency for subjects to initially interpret the stimuli according to the formant-transition cue, as reported by Walley and Carrell (1983), further supports our claim that subjects were bringing preexisting knowledge about cue structures in natural speech to bear on the interpretation of these stimuli.

Differences Between Training Groups

However, this difference in response patterns between the two training groups highlights the fact that the two groups of subjects in this experiment were trained on different cues, which previous research has shown do not have an equivalent salience as cues to place of articulation (Walley & Carrell, 1983). The difference between formant and burst cues is evident in subjects' initial response preferences. On the pretest, subjects responded according to the formant cue almost three times as often as they did using the burst cue (.57 vs. .22), and Figure 6 shows a similar pattern of response preference.

This observed preference of subjects for using formant cues seems to affect learning. Burst-trained listeners found the learning task more difficult than did formant-trained listeners, as indicated by their answers on pretest and posttest questionnaires. On each questionnaire, subjects rated the difficulty of the identification task on a scale of 1 to 7, where 1 was easiest and 7 was hardest. Burst-trained subjects gave a mean overall rating of 4.5, while formant-trained subjects' mean overall rating was 3.8, though both groups rated the task as having been harder at the beginning than at the end.

Burst-trained listeners did not exhibit nearly as much learning on the training set as did formant-trained listeners, but generalization appears stronger for the burst-trained listeners. Burst-trained learners improved by .11 on burst cues, while formant-trained learners improved by .19 on formant cues. The effect of training was similar for the learners on the generalization set, with burst-trained learners improving by .06 and formant-trained listeners improving by .03. Overall, there was a drop in learning from the training set to the generalization set, but the drop is more obvious for the formant-trained listeners (.19 to .03) than for the burst-trained listeners (.11 to .06).

One possible explanation for this is that those subjects who learned to attend to bursts learned to do so comparatively well (though few did indeed learn), while many of the formant-trained subjects learned to attend to formant cues only a little bit more than before. Recall that 6 of the 18 original burst-trained listeners failed to show any effect of learning, while only 1 of the 18 original formant-trained listeners was excluded from the learners group. This disproportionate ratio suggests that there is something about learning bursts that is more difficult, or less likely, across listeners.

Burst-trained listeners may have primarily learned a negative lesson. Feedback designed to yield positive evidence for the usefulness of burst cues provides evidence against using formant cues. Listeners may only learn the negative implication (ignore the formant cue) without learning the corresponding positive implication (attend more to the burst cue). If burst-trained subjects followed this learning strategy, we would expect to find two changes in their response patterns: (1) a decrease in untrained responses as listeners learned not to use the cue that conflicted with the cue they were being trained on, and (2) an equal increase in both trained and other responses as listeners divided up more of their responses evenly between these other two possibilities.

The results displayed in Figure 5 suggest that this is only partially the case. While burst-trained listeners did indeed increase their other responses by 4 percentage points, from .22 to .26, as negative learning predicts, this difference is only marginally significant [$t(17) = 1.980$, $p = .06$]. Similarly, their percentage of trained responses did increase more than the percentage of other responses (.07 vs. .04), suggesting that listeners learned not only to ignore the untrained cue but also to respond more often according to the trained cue; this difference is not significant [$t(17) = .494$, n.s.], however. Thus, we are left with the equivocal conclusion that either burst-trained listeners did not learn to attend more to the burst, or attending to the burst simply demands more training to show a reliably greater increase in burst-based responses. In contrast, the responses of the formant-trained subjects clearly indicates that formant-trained listeners learned both to increase their attention to formant cues (thereby increasing their trained responses) and to decrease their reliance on burst cues. This change resulted in a decrease in both untrained and other responses, contrary to the predictions of negative learning. Formant-trained listeners clearly learned to attend more strongly to formant cues, and less strongly to burst cues.

Conclusions

The ability to shift attention from less informative cues to more informative cues based on explicit feedback must depend on a basic cognitive mechanism that can adaptively remap acoustic patterns onto phonological categories. This is just the kind of mechanism needed to account for previous research on perceptual learning of synthetic speech (e.g., Greenspan, Nusbaum, & Pisoni, 1988; Nusbaum & Lee, 1992; Schwab, Nusbaum, & Pisoni, 1985). Moreover, this mechanism could also play a role in second-language acquisition and in normal language development.

With respect to second language acquisition, the two tasks examined in the present experiment, formant-learning and burst-learning, can be thought of as examples of learning Best's (1994, 1995) two-category (TC) and single-category (SC) contrasts, respectively. In learning a TC contrast, listeners must learn to distinguish between two foreign categories that assimilate to different native categories. In learning a SC contrast, listeners must also learn to distinguish between the two foreign categories, but in this case both foreign categories assimilate more or less equally well to a single native category. Our listeners were asked to learn to perceive two stimuli as different, even though both shared one cue for place of articulation and differed according to another cue. This task was comparatively easy for the formant-trained listeners because syllables that differ according to formant transitions map easily onto different categories in English (a TC contrast) because the formant structure is a very salient cue to distinguishing place of articulation. Learning was more difficult for the burst-trained subjects because bursts are less salient (Walley & Carrell, 1983). Thus, tokens that differed only in terms of the burst (but shared formant cues) were both still assimilated to the single category cued by the formant transitions (a SC contrast).

Even though the architecture of the human vocal tract makes it impossible for any real human language to exhibit a divergence of these particular two cues, the principle still holds. In phonetic learning, listeners must relate acoustic pattern structure to phonological categories in a context-sensitive manner. Although the present results do not directly examine learning in these cases, where new phonological categories must be learned at the same time as acoustic patterns, it seems plausible that a mechanism that shifts attention between acoustic cues may be an important part of this process. Indeed, just this kind of mechanism for shifting attentional focus among acoustic cues has been invoked either explicitly or implicitly by many researchers to account for facts of both first and second language acquisition (e.g., Best, 1994; Jusczyk, 1994, 1997; Pisoni et al., 1994; Polka, 1991; Strange, 1995; Werker, 1994).

Research on the acquisition of foreign language contrasts suggests that a listener's ability to learn a new phonetic contrast may require restructuring existing knowledge (cf. Cheng, 1985). In the process of phonetic learning, listeners must discover which cues are important in which contexts, and then shift their attention to those cues in those contexts. The data presented here constitute evidence for the existence of a cognitive mechanism for performing precisely this kind of reanalysis of familiar cues on the basis of category-level feedback.

REFERENCES

Best, C. T. (1994). The emergence of language-specific phonemic influences in infant speech perception. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171-206). Baltimore: York Press.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, **66**, 1001-1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, **67**, 648-662.

Carden, G., Levitt, A. G., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics*, **29**, 26-36.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, **92**, 414-422.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, **81**, 348-374.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, **27**, 769-773.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 421-433.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/–/θ/ contrast by francophones. *Perception & Psychophysics*, **40**, 205-215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, **43**, 88-96.

Jusczyk, P. (1994). Infant speech perception and the development of the mental lexicon. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 227-270). Cambridge, MA: MIT Press.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge. MA: MIT Press.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

Liberman, A. M. (1970). Some characteristics of perception in the speech mode. *Perception & Its Disorders*, **48**, 238-254.

Lisker, L. (1970). *On learning a new contrast.* (Status Report on Speech Research SR-24). New Haven, CT: Haskins Laboratories.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.

Lively, S. E., Pisoni, D. B., & Logan, J. S. (1991). Some effects of training Japanese listeners to identify English /r/ and /l/. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 175-196). Tokyo: Ohmsha.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, **89**, 874-886.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 351-378). Baltimore: York.

Markel, J. D., & Gray, A. H. Jr. (1976). *Linear prediction of speech*. New York: Springer-Verlag.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, **34**, 323-330.

Nittrouer, S., & Crowther, C. S. (1998). Examining the role of auditory sensitivity in the developmental weighting shift. *Journal of Speech, Language, & Hearing Research*, **41**, 809-818.

Nittrouer, S., & Miller, M. E. (1997a). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America*, **102**, 572-580.

Nittrouer, S., & Miller, M. E. (1997b). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, **101**, 2253-2266.

Nusbaum, H. C., & Goodman, J. (1994). Learning to hear speech as spoken language. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 299-338). Cambridge, MA: MIT Press.

Nusbaum, H. C., & Lee, L. (1992). Learning to hear phonetic information. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 265-274). Tokyo: Ohmsha.

Nusbaum, H. C., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109-132). San Diego: Academic Press.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception* (pp. 113-158). San Diego: Academic Press.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982).

Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 297-314.

Pisoni, D. B., Lively, J. S., & Logan, S. E. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, **89**, 2961-2977.

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

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, **27**, 395-408.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 3-45). Baltimore: York.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, **36**, 131-145.

Strange, W., & Jenkins, J. (1978). The role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick (Eds.), *Perception and experience* (pp. 125-169). New York: Plenum.

Sussman, H. M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: The orderly output constraint. *Behavioral & Brain Sciences*, **21**, 241-299.

Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics*, **58**, 936-946.

Tees, R., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability discriminate non-native speech sounds. *Canadian Journal of Psychology*, **38**, 579-590.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **73**, 1011-0122.

Werker, J. F. (1994). Cross-language speech perception: Developmental change does not involve loss. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 93-120). Cambridge, MA: MIT Press.

Werker, J. F., & Tees, R. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, **75**, 1866-1878.

Yamada, R. A., & Tohkura, Y. (1991). Perception of American English /r/ and /l/ by native speakers of Japanese. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 155-174). Tokyo: Ohmsha.

NOTES

1. All tests using residual mean squares are based on arcsine-transformed percentages to insure that block and treatment effects are additive (Kirk, 1995).

2. Analyses on the generalization set stimuli are one tailed because results from the training set stimuli indicate (but do not determine) the likely direction of change on the generalization stimuli.

(Manuscript received August 17, 1998;
revision accepted for publication January 26, 2000.)

Notices and Announcements

Psychonomic Society Journals on Line

The September issue of *Psychonomic Bulletin & Review* is now available on line free of charge, as well as in print. Beginning in January 2001, all of the Psychonomic Society journals will be on line. On-line subscriptions will be available at no additional cost to all who subscribe to the printed editions of the journals. Others will be able to download individual articles for a fee. Tables of contents and abstracts of current and past issues from January on will be available to all at no cost. For information updates, readers should check the Psychonomic Society Publications web site: www.psychonomic.org.