Transfer of training of a new linguistic contrast in voicing

CYNTHIA L. McCLASKEY, DAVID B. PISONI, and THOMAS D. CARRELL
Indiana University, Bloomington, Indiana

The present study examined the plasticity of the human perceptual system by means of laboratory training procedures designed to modify the perception of the voicing dimension in synthetic speech stimuli. Although the results of earlier laboratory training studies have been ambiguous, recently Pisoni, Aslin, Perey, and Hennessy (1982) have succeeded in altering the perception of labial stop consonants from a two-way contrast in voicing to a three-way contrast. The present study extended these initial results by demonstrating that experience gained from discrimination training on one place of articulation (e.g., labial) can be transferred to another place of articulation (e.g., alveolar) without any additional training on the specific test stimuli. Quantitative analyses of the identification functions showed that the new perceptual categories were stable and displayed well-defined labeling boundaries between categories. Taken together with the earlier findings, these results imply a greater degree of plasticity in the adult speech processing system than has generally been acknowledged in past studies.

This report is concerned with the perceptual categorization of human speech sounds, particularly the categorization of the voicing dimension in stop consonants. The voicing dimension has been studied extensively in recent years in an attempt to characterize the potential interactions between genetic and environmental influences in speech perception (see Aslin & Pisoni, 1980, for a review). In the word initial English stop consonants /b, p, d, t, g, k/, voice onset time (VOT) is defined as the interval between the onset of vocal cord vibration and the release of articulatory closure. The English phonemes, however, represent only a subset of the possible voicing distinctions used in human languages. By measuring the production of stop consonants in 11 diverse languages, Lisker and Abramson (1964) found three major modes of voicing. The first mode includes long lead stops in which vocal-cord vibration precedes the release. Lead values generally occur at −25 msec VOT or greater. The second category includes short lag stops in which laryngeal and supralaryngeal events are nearly simultaneous. Stops in this category have values of approximately −20 and +25 msec VOT. Finally, the third category includes long lag or voiceless aspirated stops that are characterized by substantial voicing delay until after the release. These stops have VOT values greater than +25 msec. The English stops /p, t, and k/ are exemplars of long-lag stops; /b, d, and

g/ are referred to as voiced stops, although they may display voicing lead or short lag, depending on context and talker. Long-lead and short-lag stops are generally allophonic in English so that, for example, a /b/ produced with −30 msec VOT and a /b/ at 0 msec VOT are both still perceived as the same phoneme. Lisker and Abramson (1964) found that each language they studied used only a subset of the three voicing modes that are possible at each place of articulation.

In subsequent cross-language experiments using synthetic speech stimuli differing in VOT, Lisker and Abramson (1967) found that subjects perceived the stimuli categorically. The identification and discrimination functions both corresponded closely to the phonemic categories observed in each language. The identification or labeling functions were very consistent, displaying steep slopes at the boundaries separating perceptual categories. In addition, subjects showed poor discrimination of pairs of stimuli within the same phonemic category, but excellent discrimination of pairs of stimuli selected from different categories. This combination of labeling and discrimination results suggests that the ability to discriminate between speech stimuli is closely related to their identification (Liberman, Harris, Hoffman, & Griffith, 1957). Thus, although the occurrence of three primary modes of voicing implies that certain regions of the voicing continuum may be naturally discriminable (Streeter, 1976a, 1976b), the perceptual categories used by adults appear to be determined largely through linguistic experience.

The effects of linguistic experience on speech perception have been studied to a lesser extent with other phonemic contrasts. Goto (1971) and Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) have examined the abilities of Japanese and English

subjects to identify and discriminate the liquids /r/ and /l/. These consonants are phonemically distinctive in English but not in Japanese. In both studies, the Japanese subjects showed poorer discrimination of /r/ and /l/ than the English subjects. These findings have been replicated and extended by MacKain, Best, and Strange (1982) with Japanese learners of English. They found that Japanese adults with little conversational instruction in English had great difficulty in discriminating synthetic versions of "rock" and "lock."

Taken together, these cross-language studies have demonstrated that linguistic experience plays an important role in influencing the discrimination of speech sounds. However, the long-term effects of this environmental influence and the ease with which the perceptual system of mature adults can be modified has been a much more controversial issue. Strange and Jenkins (1978) recently reviewed a number of laboratory training experiments which attempted to modify the perception of the voicing continuum. Although the ability of some adults to discriminate nonnative contrasts in voicing showed improvement, Strange and Jenkins concluded that "significant modification of phonetic perception is not easily obtained by simple laboratory training techniques" (Strange & Jenkins, 1978, p. 153). Moreover, they concluded that the modifications that did occur appeared to be limited to the original test stimuli and the same perceptual tasks that were used in training. The possibility exists, of course, that even those subjects whose perceptions were altered were not learning to modify the voicing continuum as such. Hence, the generalizability of these results is severely restricted. Strange and Jenkins also concluded, as have most researchers over the past two decades, that the adult perceptual system is extremely resistant to change and that selective retuning or modification through the use of laboratory training procedures is an arduous, if not impossible, process.

It is possible that the failure of these previous attempts to produce new linguistic contrasts reflects a genuine absence of the required perceptual abilities. However, other studies indicate that the problem may be more methodological. For example, Carney, Widin, and Viemeister (1977) have demonstrated that discrimination within phonemic categories is possible when the relevant acoustic cues are emphasized during training. In addition, Carney et al. showed that their subjects could change their category boundaries to arbitrary values chosen by the experimenters. However, the subjects used in the Carney et al. study were highly experienced veterans of psychophysical experiments and had been participants in many testing sessions before the data were collected. Moreover, the experimental procedures used by Carney et al. involved a low-uncertainty testing format and many sessions spread out over a period of several months.

In contrast, the subjects used in a recent study by Pisoni, Aslin, Perey, and Hennessy (1982) were naive

undergraduates who were able to modify their perception of VOT in less than 1 h of exposure to the new contrast. By providing immediate feedback during training to emphasize the relevant acoustic cues, Pisoni et al. demonstrated that naive listeners could learn to perceive a new voicing category. However, one weakness of the experimental design employed by Pisoni et al. was first pointed out by Strange (1972) in reference to her own work. Because the subjects were both trained and tested on the same synthetic labial stop consonants, it is not clear whether the subjects were actually learning about the specific place of articulation or about the distinctive acoustic correlates of VOT.

In a recent study dealing with the discrimination of VOT, Edman, Soli, and Widin (Note 1) examined the performance of subjects who transferred discrimination experience gained from one place of articulation to another. The subjects were given initial and final discrimination and identification tests using the /b-p/ and /g-k/ continua of Lisker and Abramson, but they were trained on only one of the series. For a subject who was trained on the labial continuum, the discrimination training effects were substantial for the labial stimuli, as might be expected (see, also, Soli, 1983). However, the subject also demonstrated significant transfer of training in discrimination, suggesting that subjects can learn "VOT *per se* and not some unique properties of the training series" (Edman et al., Note 1, p. 5).

The present investigation was designed to extend the findings on discrimination to absolute identification of a new linguistic category in voicing. Although Pisoni et al. (1982) demonstrated that naive subjects could learn to identify a new linguistic category, they did not establish that the discrimination training and experience with the new linguistic category transferred to other places of articulation. And, although Edman et al. (Note 1) showed transfer of training in discrimination, they did not assess whether the discrimination training also transferred to identification and labeling of a new voicing category.

METHOD

Subjects

Twenty-one subjects were drawn from a paid-subject pool of undergraduate students attending Indiana University. The subjects were originally recruited through advertisements and were paid $3 per hour for each testing session. All subjects were monolingual speakers of English with no history of a speech or hearing problem, as determined by a pretest questionnaire, and all were naive to the experimental procedures and stimuli used in the experiment.

Stimuli

The stimuli consisted of two sets of 15 synthetic stop consonant-vowel syllables corresponding to labial and alveolar places of articulation. Each stimulus set was generated on the Klatt (1980) cascade-parallel software synthesizer as implemented in the Speech Research Laboratory at Indiana University (Kewley-Port, Note 2). The stimuli differed in 10-msec steps of

VOT from −70 to +70 msec. The steady-state portion of the stimuli consisted of the vowel /a/, which was 255 msec in duration. The formant values chosen for this vowel were: F1 = 700 Hz, BW1 = 90 Hz; F2 = 1200 Hz, BW2 = 90 Hz; F3 = 2600 Hz, BW3 = 130 Hz; F4 = 3300 Hz, BW4 = 400 Hz; F5 = 3700 Hz, BW5 = 500 Hz. The labial formant transitions were 40 msec in duration and had starting frequencies of F1 = 438 Hz, F2 = 1025 Hz, F3 = 2425 Hz. These frequencies were selected to simulate productions of natural speech as measured by Klatt (Note 3) and measurements of spectrograms of a male speaker in our laboratory. The alveolar formant transitions were 50 msec in duration. The starting frequencies were: F1 = 400 Hz, F2 = 1550 Hz, F3 = 2600 Hz. To simulate the burst that occurs at release of stop closure, a turbulent noise source (AF) was passed through the bypass channel of the parallel branch of the synthesizer. The burst was 10 msec in duration, and the amplitude was carefully chosen on the basis of pilot listening tests. The spectrum of the labial release burst was distributed fairly uniformly across all frequencies, and the energy of the alveolar burst was centered around 3300-3850 Hz. Voicing lead was simulated by setting F1 at 180 Hz with a bandwidth of 150 Hz. The sinusoidal voicing source (ASV) was then passed through F1 to simulate the low-frequency energy. Voiceless stimuli have an aspirated component produced by turbulence at the opening of the vocal cords. Aspiration was simulated by passing a noise source through the cascade branch of the synthesizer. Aspiration noise was also added to the final 35 msec of the CV syllable, and F1 was widened somewhat to make the offsets sound more natural. At the onset of release of stop closure, the pitch contour (F0) rose briefly from 120 to 125 Hz and then fell linearly to 100 Hz and remained there for the duration of the steady-state portion of the vowel.

Procedure

The experiment was conducted with groups of two to six subjects seated in separate testing booths in a small experimental room. The stimuli were presented binaurally over matched and calibrated TDH-39 headphones. Voltage levels were monitored with a VTVM at a constant level of 80 dB SPL during presentation of stimuli. The collection of all responses and presentation of feedback to subjects was controlled on-line in real time by a PDP-11/05 computer. All experimental trials were paced to the slowest subject in each group.

On Day 1, the subjects were presented with four different phases of the experiment, each phase was summarized on a typed page of instructions which was also read aloud to subjects by the experimenter just prior to the presentation of stimuli. In the first phase, the subjects were asked to identify the 15 stimuli (each presented 10 times in random order) into two categories corresponding to the English categories /ba/ and /pa/. The onset of a trial was signaled by the illumination of a cue light on top of the response box.

In the next phase, familiarization, the subjects merely listened to several ordered presentations of the −70-, 0-, and +70-msec VOT stimuli. The subjects did not respond to the stimuli but were instructed to concentrate on listening to the beginning of the consonant-vowel syllable. The third phase consisted of 40 presentations of the three exemplary stimuli (i.e., −70-, 0-, and +70-msec VOT) in a random order. The subjects were now required to use three response categories in labeling the stimuli. Immediately following each identification response, a light above the correct response button was illuminated to indicate the correct response and provide feedback to the subject. Finally, in the last phase of Day 1, all 15 stimuli were again presented 10 times each in random order. However, now the subjects were required to identify stimuli from the entire continuum, using all three response categories without any feedback. Of the original 21 subjects run on the first day of testing, 15 passed the 85% correct training criterion in the third phase and were asked to return for a second day of testing.

On Day 2, the last two phases of the previous day were repeated again, using the labial CV series. When these trials were

completed, the subjects were presented with the alveolar stimuli (i.e., the transfer series), using −70-, 0-, and +70-msec VOT exemplars arranged in order. The subjects did not respond overtly to these stimuli, but merely listened to several presentations of these test signals. The last phase of Day 2, the transfer phase, consisted of 10 randomly ordered presentations of each of the 15 alveolar CV syllables. The subjects were now required to identify these stimuli into the three new categories without feedback for correct responses. One group of subjects (N = 7) received training on the labial stimuli first and were then tested for transfer on the alveolar series. Another group (N = 8) received the stimuli in reverse order; that is, training on the alveolars first followed by transfer to labials.

RESULTS

Figure 1 shows the average labeling functions for the group of subjects, trained first on the labial stimuli and then transferred to alveolar stimuli; Figure 2 shows the functions of the second group, which was trained on alveolar and then transferred to labial stimuli. In each figure, there are four panels corresponding to the four conditions of the experiment. These data are from only the subjects who met the 85% criterion on Day 1. Data from the remaining subjects were excluded from any further analyses.

The two category identification functions are shown in the upper left-hand panel of each figure. The data shown in these panels indicate that two-category labeling
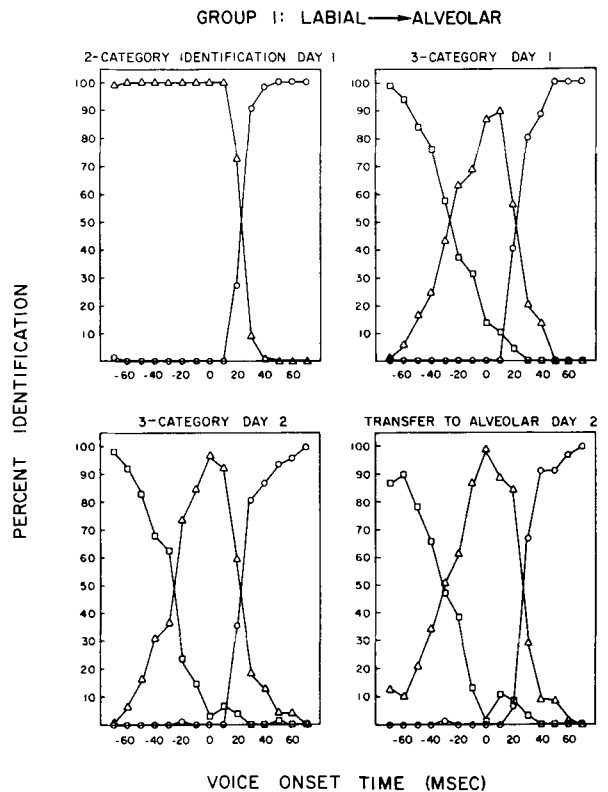

Figure 1. Average identification functions for two- and three-category labeling and three-category transfer labeling for Group 1.
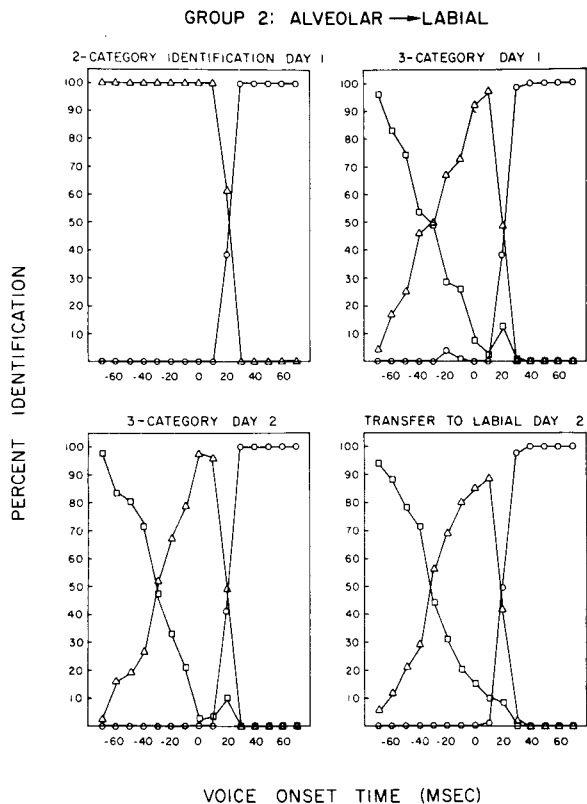
GROUP 2: ALVEOLAR ―→LABIAL

Figure 2. Average identification functions for two- and three-category labeling and three-category transfer labeling for Group 2.

of speech stimuli that are phonemically distinct in English can be done very consistently by native speakers of English. The second panel, in the upper right of each figure, corresponds to the last condition of Day 1. Inspection of these functions reveals that after familiarization and training with one set of stimuli that varied in VOT, the subjects demonstrated consistent identification of an additional (third) category, one that is not functionally distinctive in English. This result was accomplished with less than 1 h of training. Panel 3, in the lower left of each figure, shows a replication of the three-category identification data obtained on Day 2 of the experiment. A visual comparison of individual subjects' data across these two conditions demonstrates the high consistency of the identification responses, although there is some variability across subjects. This consistency is due, in part, to the original criterion used for inclusion in the study. The subjects were required to meet an 85% correct level of performance on the three-category exemplar stimuli in order to participate further in the second phase of the experiment. Subjects who could not identify the three exemplar stimuli (−70, 0, and +70 msec) were excused from testing on Day 2.

The transfer of training data, shown in Panel 4 in the lower right of each figure, closely resemble the data

shown in Panels 2 and 3. From visual inspection of these data, it is obvious that the subjects were able to use three labeling categories in a consistent and highly reliable manner, even though one of the perceptual categories was not phonemically distinctive in their native language. Moreover, these subjects were able to perceive reliably and to identify consistently differences in voicing lead, even though they were tested with stimuli on which they had received no specific training.

Examination of the individual subjects' data shows that the subjects were able to identify three distinct categories along the voicing continuum. Inspection of these figures also shows the occurrence of substantial transfer effects in perception of VOT from one place of articulation to another. In order to quantify these observations more precisely, several analyses of the identification data were carried out to obtain estimates of the strength of the transfer of training effects. A measure of the consistency of responses at each of the 15 stimulus values was first computed and then averaged across the stimulus values to obtain an overall consistency score for each subject (Attneave, 1959). These values are shown in Table 1 for each subject in Conditions 1-4 along with the mean values for each condition.[1]

The values of this index, the index of response uncertainty, range from .999 (very consistent) to .593 (fairly consistent). In general, the response consistency values of each subject are higher for Condition 1, requiring two-category identification, than for any other condition. This result reflects the fact that these two categories of voicing are typically identified by monolingual speakers of English, whereas performance in identifying three categories is a product of less than 2 h of laboratory training.

Table 1
Response Consistency Values for Identification

	Conditions			
Subjects	1	2	3	4
Group 1: Labial to Alveolar Series				
1	.999	.791	.821	.842
2	.919	.712	.773	.737
3	.940	.813	.813	.870
4	.999	.747	.757	.879
5	.813	.800	.865	.593
6	.934	.728	.794	.739
7	.951	.695	.845	.706
Means	.936	.755	.815	.767
Group 2: Alveolar to Labial Series				
1	.934	.807	.846	.849
2	.999	.726	.726	.678
3	.940	.838	.850	.806
4	.967	.714	.768	.736
5	.951	.786	.742	.847
6	.940	.752	.882	.721
7	.951	.834	.979	.861
8	.999	.719	.706	.712
Means	.960	.772	.821	.776

The next analysis that we carried out was designed to measure the sharpness or steepness of the labeling functions. Individual slopes and crossover points were calculated for each labeling function by fitting a normal ogive to the 15 identification data points for each subject via the methods outlined in Woodworth (1938). Thus, for the two-category identification phase, there is only one slope for each subject because only the voiced/voiceless distinction is phonemic in English. For the three-category conditions, two slopes were computed, one for the prevoiced/voiced boundary and the other for the voiced/voiceless boundary.

Tables 2 and 3 present the slope values for the labeling functions. Low values represent steeper slopes in this analysis than do high values. The slope values are consistently smaller for the voiced/voiceless boundary (Wilcoxon test, p < .01). Thus, the labeling functions have steeper slopes for the voiced/voiceless than for the prevoiced/voiceless boundary. Other analyses, across conditions and between groups, revealed no other reliable differences in the slopes of the labeling functions.

A third analysis was also carried out to quantify the precise location of the category boundaries on the VOT continua. A crossover point was obtained from the fitted ogives used to calculate the slopes for each boundary. In addition, a midpoint for each boundary was obtained by inspection directly from the graphs by selecting the point at which each function first straddled the 50% crossover value. The two values for the voiced/voiceless boundary are given in milliseconds in Table 4; the prevoiced/voiced boundary values are given in Table 5.

Scores obtained in these three analyses (response consistency, slope, and crossover) were compared across the

four conditions for each group by means of the Wilcoxon matched pairs test (Siegel, 1956). Comparisons of response consistency across the four conditions indicated that only two conditions differed significantly from each other (p < .01, two-tailed). For both groups, the consistency of responses during identification of two categories on Day 1 showed reliable differences from the consistency of responses in the transfer test on Day 2. These conditions correspond to the data shown in Panels 1 and 4 in each figure. Although the values of response consistency for the data shown in Panel 4 are fairly high, with a mean value of .767 for Group 1 and a mean value of .776 for Group 2, this result was expected due to the lack of specific experience with the transfer stimuli. Another cross-condition consistency comparison that was significant occurred only in Group 1. A significant change (p < .01, two-tailed) in consistency of responses was found from Day 1 to Day 2 on the labial three-category identification stimuli shown in Panels 2 and 3. It appears that the brief amount of practice provided on Day 2 improved the consistency of the responses to labial stimuli (Group 1), but apparently had no comparable effect on the alveolar stimuli (Group 2). The reason for this asymmetry in performance across the two place series is not clear to us at the present time.

Analysis of cross-series differences in terms of the slope of the labeling functions showed that, in general, the slopes did not change significantly across conditions. However, there were two exceptions to this observation. For Group 1, the voiced/voiceless boundary differed significantly (p < .01) for the two-category identification condition (Panel 1), as compared with the

| Table 2 |
| Slope Values for Voiced/Voiceless |
| Boundary in Identification |

Subjects	Conditions			
	1	2	3	4
Group 1: Labial to Alveolar Series				
1	20.82	21.87	20.82	20.82
2	21.25	26.24	35.60	21.52
3	19.87	19.87	19.94	20.39
4	20.82	21.04	21.23	21.49
5	24.13	20.64	20.90	24.30
6	19.94	19.87	20.05	20.39
7	20.26	19.78	20.91	32.36
Means	21.01	21.33	22.78	23.04
Group 2: Alveolar to Labial Series				
1	19.94	19.78	20.91	20.17
2	20.82	20.39	20.82	19.98
3	19.87	19.28	19.28	19.07
4	20.39	20.82	20.81	21.71
5	20.26	20.39	20.02	19.66
6	30.17	21.28	20.39	20.26
7	20.26	20.39	20.82	20.39
8	19.28	20.26	19.28	19.66
Means	21.37	20.32	20.29	20.11

| Table 3 |
| Slope Values for Prevoiced/Voiced |
| Boundary in Identification |

Subjects	Conditions		
	2	3	4
Group 1: Labial to Alveolar Series			
1	−22.07	−22.80	−22.74
2	−25.92	−20.92	−24.13
3	−24.27	−22.56	−22.70
4	−27.33	−30.23	−22.42
5	−24.19	−21.30	−35.78
6	−27.21	−26.79	−26.18
7	−26.95	−36.35	−73.07
Means	−25.42	−25.84	−32.43
Group 2: Alveolar to Labial Series			
1	−28.54	−23.89	−23.49
2	−35.39	−27.64	−27.76
3	−32.80	−40.06	−42.89
4	−32.54	−29.63	−30.00
5	−24.85	−27.23	−28.37
6	−24.99	−22.77	−24.80
7	−22.18	−18.76	−20.28
8	−32.30	−29.33	−30.13
Means	−29.20	−27.43	−28.47

Table 4
Midpoint and Crossover Values for Voiced/Voiceless Boundary in Identification

| | Conditions | | | | | | | |
| | 1 | | 2 | | 3 | | 4 | |
Subjects	Midpt	Mean	Midpt	Mean	Midpt	Mean	Midpt	Mean
			Group 1: Labial to Alveolar Series					
1	56	18.67	58	22.77	56	18.67	55	18.67
2	26	29.68	24	37.43	47	64.39	30	20.50
3	17	13.56	17	13.56	19	31.97	25	16.37
4	25	18.67	25	18.86	26	19.57	26	21.29
5	27	22.07	20	16.74	25	18.13	37	31.01
6	18	13.97	18	13.56	16	10.79	24	16.37
7	24	15.67	16	13.08	18	16.23	53	46.59
Means	27.57	20.86	25.42	19.42	29.57	23.11	35.72	24.40
			Group 2: Alveolar to Labial Series					
1	18	13.97	15	13.08	17	16.23	23	15.17
2	25	18.67	24	16.37	24	18.67	19	14.14
3	17	13.56	15	10.37	16	10.37	14	8.30
4	25	16.37	25	18.67	26	18.67	26	22.14
5	23	15.67	23	16.37	20	14.36	16	12.42
6	22	15.17	24	12.01	25	16.37	24	15.67
7	24	15.67	26	16.37	25	18.67	26	16.37
8	16	10.37	16	14.70	15	10.37	16	12.42
Means	21.55	14.93	21	14.74	21	15.46	20.50	14.58

Table 5
Midpoint and Crossover Values for Prevoiced/Voiced Boundary in Identification

| | Conditions | | | | | |
| | 2 | | 3 | | 4 | |
Subjects	Midpt	Mean	Midpt	Mean	Midpt	Mean
			Group 1: Labial to Alveolar Series			
1	−17	−12.74	−22	−15.75	−15	−17.65
2	− 6	−19.61	−21	− 5.36	−29	−22.55
3	−30	−26.62	−26	−23.41	−30	−23.94
4	−35	−28.26	−53	−41.03	−30	−23.08
5	−25	−23.76	−23	−11.40	−34	−31.24
6	−40	−30.23	−20	−33.25	−15	−21.36
7	−20	−17.15	− 4	−66.41	−60	−158.58
Means	−24.71	−23.63	−21.14	−28.09	−34.63	−42.63
			Group 2: Alveolar to Labial Series			
1	−36	−40.47	−26	−27.22	−33	−21.68
2	−47	−47.24	−35	−29.90	−26	−22.87
3	−46	−51.92	−46	−73.56	−57	−78.05
4	−40	−34.05	−40	−32.29	−32	−28.71
5	−20	−22.45	−26	−29.72	−30	−41.01
6	−30	−26.55	−32	−23.74	−35	−18.80
7	− 3	− 7.67	− 5	− 5.13	− 9	− 8.15
8	−44	−31.59	−60	−29.60	−34	−34.06
Means	−33.25	−32.74	−33.75	−31.40	−32	−31.67

transfer condition (Panel 4). Although this result indicates a steeper slope for the two-category condition, a comparison of the slope values in Table 2 indicates that, with the exception of Subject 7, most values are fairly consistent across the two conditions. Also for Group 1, the differences in the location of the prevoiced/voiced VOT boundary were significant. When three-category identification on Day 2 was compared with the transfer function (Panel 3 vs. 4), the slopes of the identification function for the transfer series were less steep than the slopes of the functions obtained with the training stimuli.

Turning to comparisons involving the crossover values, no significant differences were observed for comparisons of Group 2, indicating that the location of the category boundaries did not change across the various conditions of the experiment. These results were further confirmed by analysis of the midpoints for

TRANSFER OF TRAINING CONTRAST 329

Group 2 which also showed no change across condition. Group 1, however, showed a significant difference in the location of the voiced/voiceless boundary, which shifted slightly from the two-category identification to the transfer condition (Panels 1 vs. 4). The precise reason why a shift was observed in one series and not the other is also not clear to us at this time.

In order to determine if there were any between-group differences, the labeling data were compared by means of Mann-Whitney U tests for each condition (Siegel, 1956). A comparison of the U values showed no significant differences between Group 1 and Group 2 for three of the four conditions: two-category identification (Condition 1), three-category identification Day 1 (Condition 2), and three-category identification Day 2 (Condition 3). This was true for analyses of both the slope and response consistency measures. However, the fourth condition, the transfer phase, showed one statistically significant difference between Group 1 and Group 2 (U = 6). This result was found in the voiced/voiceless region and reflects the finding that Group 1 (labial stimuli transferred to alveolar stimuli) had a slightly steeper slope than Group 2 (alveolar stimuli transferred to labial stimuli).

In summary, the results of all three analyses of the labeling data—slope, response consistency, and crossover point—show that three perceptual categories could be reliably identified by our subjects in less than 2 h of laboratory training. Furthermore, when presented with stimuli on which they had not received specific training, our subjects generalized the experience they had gained with VOT during training to identification of the new stimuli. A reliable and consistent perceptual category emerged clearly for these subjects, with relatively simple laboratory procedures and in a short period of time.

DISCUSSION

The present experiment demonstrates that a new perceptual category can be acquired in a relatively short amount of time through relatively simple laboratory training procedures. The subjects consistently identified the VOT stimuli into three perceptual categories and showed strong evidence of transfer of training from the place of articulation on which they had originally been trained to another place of articulation. Such findings demonstrate clearly that the perceptual capacities required to learn or relearn these contrasts are still present in experimentally naive monolingual adult speakers of English. English speakers typically divide stop consonants differing in VOT into only two categories at each place of articulation corresponding to /b-p/, /d-t/, and /g-k/. In the present study, after training on stimuli from one place of articulation, our subjects were able to generalize their experience of the three-way labeling contrast to another place of articulation without additional training. The present findings on transfer of *identification* are therefore complementary to the results

reported earlier by Edman et al. (Note 1), who demonstrated improved intraphonemic *discrimination* of stop consonants after training (see, also, Soli, 1982). Edman et al. also showed that discrimination training generalized from one place of articulation during training (e.g., /b-p/) to a second place of articulation at the time of final testing (e.g., /g-k/). However, these investigators failed to obtain any identification data from their subjects. Transfer of training was examined only for discrimination.

The results of the present experiment extend the earlier findings of Pisoni et al. (1982) by demonstrating a robust transfer of training effect from one synthetic stimulus continuum to another. In the Pisoni et al. study, subjects were trained and tested on only a labial continuum differing in VOT. By demonstrating three-category identification for VOT stimuli that were not used in training, the present study has established that naive subjects are able to acquire very specific and detailed knowledge about the complex temporal-spectral dimension of voicing as cued by VOT. The knowledge gained is not just about the distinctive correlates of the specific test stimuli used in the original training.

The findings obtained in this study raise several questions about the conclusions previous investigators have drawn concerning the effects of laboratory training in speech perception, namely, that laboratory training procedures appear to have little, if any, effect on modifying speech perception in adult listeners. In this regard, several methodological aspects of the present study are worth emphasizing here. First, we were able to control the subjects' attentive processes to focus selectively on those distinctive aspects of the stimuli that were relevant to successful identification. This was accomplished by familiarization and training with immediate feedback on exemplars or "prototypes" of the three perceptual categories. Although our verbal instructions also directed the subjects' attention to the initial part of each stimulus, the use of on-line computer techniques for discrimination training and immediate feedback was also an important aspect of the success of the training procedures used in the present study.

Second, we imposed a fairly strict training criterion on subjects which effectively reduced intrasubject variability. Although the subjects who were eliminated from testing on Day 2 did not identify the stimuli into three categories at the required performance levels, it should be noted here that they did respond well above chance. While not a specific concern of the present study, it would be of some interest in future studies to measure the amount of additional training that would be required for these noncriterion subjects to reach a comparable level of performance.

In their review of laboratory training studies, Strange and Jenkins (1978) concluded that modification of the VOT dimension is not easily accomplished by the use of laboratory training methods in a short period of time. Their conclusions have broad implications for the

plasticity of the adult perceptual system and the acquisition of new linguistic contrasts at the phonological level. The data obtained in this study suggest that previous failures to acquire a new linguistic contrast do not appear to be related in any way to permanent changes in the sensory, perceptual, or cognitive mechanisms used by mature adult listeners. Our listeners did not appear to have "lost" any of their speech-processing abilities due to environmental exposure or realignment of their perceptual mechanisms. In light of the positive results obtained in this study, there is strong impetus to extend the use of these discrimination training procedures to other linguistic contrasts involving English /r/ vs. /l/ in Japanese subjects and various tonal distinctions in certain dialects of Chinese. Cross-linguistic comparisons of this kind are currently underway in our laboratory, and reports of this work will be forthcoming.

REFERENCE NOTES

1. Edman, T. R., Soli, S. D., & Widin, G. P. *Learning and generalization of intraphonemic VOT discrimination*. Paper presented at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island, May 1978.

2. Kewley-Port, D. *KLTEXC: Executive program to implement the KLATT software speech synthesizer* (Research on Speech Perception Progress Report No. 4). Bloomington: Indiana University, 1978.

3. Klatt, D. H. *Analysis and synthesis of CV syllables in English.* Unpublished manuscript, M.I.T., Cambridge, Massachusetts, 1978.

REFERENCES

Aslin, R. N., & Pisoni, D. B. Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Perception and production.* New York: Academic Press, 1980.

Attneave, F. *Applications of information theory to psychology.* New York: Holt, Rinehart and Winston, 1959.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, 62, 961-970.

Goto, H. Auditory perception by normal adults of the sounds "l" and "r." *Neuropsychologia*, 1971, 9, 317-323.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, 54, 358-368.

Lisker, L., & Abramson, A. S. A cross language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384-422.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences.* Prague: Academia, 1967.

MacKain, K. S., Best, C. T., & Strange, W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 1982, 2, 369-390.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, 1975, 18, 331-340.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, 8, 297-314.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Soli, S. D. The role of spectral cues in the discrimination of voice onset time. *Journal of the Acoustical Society of America*, 1982, 73, 2150-2165.

Strange, W. *The effects of training on the perception of synthetic speech sounds: Voice onset time.* Unpublished doctoral dissertation, University of Minnesota, 1972.

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

Streeter, L. A. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 1976, 259, 39-41. (a)

Streeter, L. A. Kikuyu labial and apical stop discrimination. *Journal of Phonetics*, 1976, 4, 43-49. (b)

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

NOTE

1. To obtain a quantitative measure of response consistency in identification, we also calculated the amount of informational uncertainty (H) for each stimulus value along the synthetic continuum with the methods described by Attneave (1959). These H values were then converted to informational redundancy values (1−Rel H) so that complete response diversity (i.e., two or three equiprobable responses) would register as 0 and complete response consistency would equal 1.000.