Perception of intraphonemic differences by phoneticians, musicians, and inexperienced listeners

James Hillenbrand
Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

Gerald J. Canter and Bruce L. Smith
Department of Communication Sciences and Disorders, Northwestern University, 2299 Sheridan Road, Evanston, Illinois 60208

(Received 27 July 1988; accepted for publication 9 March 1990)

Several studies have shown that extensive training with synthetic speech sounds can result in substantial improvements in listeners' perception of intraphonemic differences. The purpose of the present study was to investigate the effects of listening experience on the perception of intraphonemic differences in the absence of specific training with the synthetic speech sounds being tested. Phonetically trained listeners, musicians, and untrained listeners were tested on a two-choice identification task, a three-choice identification task, and an ABX discrimination task using a synthetic [bi]–[p^hi] continuum and a synthetic [wei]–[rei] continuum. The three-choice identification task included the identification of stimuli with an "indefinite" or "ambiguous" quality in addition to clear instances of the opposing phonetic categories. Results included: (1) All three subject groups showed some ability to identify ambiguous stimuli; (2) phonetically trained listeners were better at identifying ambiguous stimuli than musicians and untrained listeners; (3) phonetically trained listeners performed better on the discrimination task than musicians and untrained listeners; (4) musicians and untrained listeners did not differ on any of the listening tasks; and (5) participation by the inexperienced listeners in a 10-week introductory phonetics course did not result in improvements in either the three-choice identification task or the discrimination task.

PACS numbers: 43.71.Es, 43.71.An, 43.72.Ja

INTRODUCTION

The categorical perception phenomenon has been a focus of phonetic perception research since the first published reports of this kind of experiment at Haskins Laboratories more than three decades ago. The basic idea behind this phenomenon is that, for certain kinds of speech-sound continua, there is a very strong relationship between the ability to discriminate stimuli and the way that stimuli along that continuum are labeled. Given an appropriate set of stimuli that are spaced evenly in physical terms, listeners are much better at discriminating stimulus pairs that are drawn from different phonetic categories than pairs that are drawn from the same phonetic category (Liberman *et al.*, 1957; Liberman *et al.*, 1961a; Liberman *et al.*, 1961b; Bastian *et al.*, 1961; see Repp, 1984; and Harned, 1987, for general reviews of this literature). This kind of performance for stops and many other consonant categories is in contrast to the "continuous" perception that is seen for vowels and most nonspeech continua (e.g., Fry *et al.*, 1962; Eimas, 1963; Bastian and Abramson, 1964; Abramson, 1961; Liberman *et al.*, 1961a; Liberman *et al.*, 1961b). For these stimulus continua, the ability to discriminate stimulus pairs far exceeds the ability to assign absolute labels to individual tokens.

There has been considerable interest in the effects of training on categorical perception. This issue has been approached from two different directions. In one set of experiments attempts have been made to produce the categorical perception effect for nonspeech sounds by training subjects in the use of category labels (Lane, 1965; Cross *et al.*, 1965), while in another series of experiments attempts have been made to attenuate the categorical perception effect for stop consonants by providing subjects with discrimination training on within-category pairs (Strange, 1972; Ganong, 1977; Samuel, 1977; Carney *et al.*, 1977; Edman *et al.*, 1978; Edman, 1979).

The primary contention of the series of experiments by Lane and his colleagues was that the categorical perception phenomenon was not specific to speech and could be obtained simply by training subjects in the use of category labels. Cross *et al.* (1965) trained subjects to use category labels in response to a visual continuum consisting of sectored circles. Subjects were able to use the category labels consistently, and a subsequent ABX discrimination task showed a peak corresponding to the labeling boundary. In a related study, Lane (1965) tested subjects on a nonspeech analog of a voicing continuum. Prior to identification training with the nonspeech analogs, ABX discrimination functions were flat and close to chance. Discrimination functions obtained after training showed a peak at the identification boundary. Similarly, Kopp and Udin (1969) and Kopp and Livermore (1973) found evidence of a discrimination peak following identification training for a continuum of sinusoids varying in frequency.

Several other experiments have failed to find evidence that identification training produces categorical perception. For example, Liberman *et al.* (1965) reported the results of an attempt to replicate Cross *et al.* (1965) that failed to find evidence of discrimination peaks for a visual continuum either before or after identification training. A second failure to replicate Cross *et al.* was reported by Parks *et al.* (1969).

Evidence for the effectiveness of techniques designed to attenuate categorical perception effects through discrimination training seems to be more clear. Hanson (1977) reported a large increase in within-category discrimination performance as a result of instructing subjects to attend carefully to small differences between stimuli and providing feedback throughout a same–different reaction-time task. A report by Repp (1975) suggests that it was primarily the feedback that was responsible for the increase in performance. Repp's listeners were given the same instructions and task but without feedback, and showed no improvement in within-category performance. Strange (1972) reported only a small improvement in performance after extensive discrimination training with feedback. Repp (1984) and Carney *et al.* (1977) attributed the relatively modest performance improvement in Strange's study to the high stimulus uncertainty of the oddity task that was used. In support of this contention, Repp cites the excellent within-category performance reported by Sachs and Grant (1976) and Ganong (1977), who provided feedback to subjects and reduced stimulus uncertainty by presenting same–different trials in blocks that contrasted only two stimuli.

The most complete set of data regarding the effects of training on the perception of intraphonemic differences consists of a series of experiments conducted at the University of Minnesota. Carney *et al.* (1977) tested subjects both before and after extensive discrimination training using a fixed-standard, same–different procedure and an oddity paradigm. The results for both procedures showed large improvements in within-category performance following training. Follow-up studies showed that training on a [b]–[p] continuum transferred readily to a [g]–[k] continuum (Edman *et al.*, 1978), and that training on a [b]–[d]–[g] continuum transferred to a [p]–[t]–[k] continuum (Edman, 1979).

Carney *et al.* also showed that providing appropriate feedback to subjects allowed listeners to assign category labels to arbitrary subsets of the stimuli. Similarly, Pisoni *et al.* (1982) showed that native speakers of American English were able to divide a -70 to $+70$-ms voice-onset time (VOT) continuum into categories corresponding to lead, short-lag, and long-lag stops after simply listening to examples of the three categories.

The Minnesota studies all involved extensive periods of discrimination training. Repp (1981) reported the results of a successful training study using fricative-vowel syllables that involved a simple and relatively brief training session. The 25-min training session consisted of presenting subjects with pairs of isolated fricative noises followed by the same pair of noises in vocalic context. The results showed a very substantial improvement in performance for fricative-vowel syllables following this brief training period.

Taken as a whole, the results of these training studies suggest that subjects can show substantial improvement in their ability to perceive intraphonemic differences given the appropriate combination of training, instructions, and psychophysical procedure. Training can improve a subject's ability to discriminate within-category distinctions and can serve to modify a subject's labeling behavior. The purpose of the present study was to determine whether listeners with extensive experience in phonetics and music are better able to perceive intraphonemic differences than inexperienced listeners *in the absence of specific identification or discrimination training on the synthetic speech sounds being tested*. Our primary interest was to examine the effects of general training in phonetics on the perception of intraphonemic differences. Data from the musicians were included to determine whether any performance advantage for the phoneticians might be specific either to speech or to the kinds of listening tasks engaged in by phoneticians.

Some limited information is available on the perception of intraphonemic differences by musicians and nonmusicians. In a study that is described only briefly, Chandler and Strange (1984) tested 20 musicians and 20 nonmusicians on a 10-step [rak]–[lak] continuum. The musicians performed better than the nonmusicians on an oddity discrimination task, and there was a slight indication that the musicians were more reliable in providing labels for subphonemic categories in an open-set identification task.

The present study compared the performance of phonetically trained listeners, experienced musicians, and untrained listeners on two identification tasks and one discrimination task. The stimuli consisted of a [bi]–[p^hi] continuum cued by variations in voice-onset time and a [wei]–[rei] continuum cued by variations in second- and third-formant starting frequencies. The first identification task was a conventional two-alternative forced-choice labeling procedure. The second identification procedure was a three-choice labeling task in which listeners were asked to identify instances of stimuli with an "ambiguous" or "indefinite" quality in addition to clear instances of the opposing phonetic categories. The discrimination task was a two-step ABX procedure.

A secondary purpose of the present study was to determine whether any performance advantage that might be attributable to phonetics training is due to long-term training or, alternatively, whether these skills might be acquired as a result of relatively short-term phonetics training. For this reason, a group of undergraduates who had enrolled in a 10-week introductory phonetics course was selected to serve as the untrained listener group. This provided the opportunity to compare the performance of these subjects both before and after general training in phonetics.

I. METHODS

A. Subjects

1. Phonetically trained listeners

Listeners in the phonetically trained group were 12 subjects with a minimum of five years of experience in speech science or clinical speech pathology. Four subjects (including the first author) were researchers in speech science/experimental phonetics, four subjects were highly experienced master's-level speech pathologists, and four subjects were

Northwestern University doctoral students in speech pathology. The speech researchers had 5–11 years of postdoctoral experience and were originally trained in experimental psychology, linguistics, or speech and hearing science. The speech pathologists had 5–20 years of experience in clinical work and clinical supervision. The doctoral students all had master's degrees in speech pathology, and two of the four had worked in a clinical setting prior to entering the doctoral program. One of the subjects (JH) had extensive experience with synthetic speech, and two other subjects had previously served as listeners in a small number of experiments involving synthetic speech. The remainder of the subjects had little or no experience with synthetic speech.

2. Musically trained listeners

A questionnaire was used to select a group of 15 highly experienced musicians who were musical instructors, professional performers, or graduate students in the Northwestern University School of Music. All of the subjects had begun formal musical training before age 12, and all judged their proficiency to be excellent in at least one instrument. In most cases subjects judged their proficiency to be either good or excellent in two or more instruments. Subjects reported 12–33 years of experience with what they considered to be their primary instrument, and in every case reported at least eight years of experience with a second instrument. Two subjects considered themselves to be primarily vocalists, although both reported extensive instrumental training as well. None of the subjects had received training in phonetics, and none was experienced with synthetic speech.

3. Inexperienced listeners

The inexperienced listeners were 15 Northwestern University undergraduates who had enrolled in a 10-week introductory course in phonetic theory and phonetic transcription. These subjects were tested both before and after the course. The phonetics course consisted of 30 h of lectures in articulatory and acoustic phonetics and 15 h of laboratory practice in phonetic transcription. The first seven h of laboratory practice emphasized broad phonemic transcription and the remaining eight h focused on narrow phonetic transcription, including the use of diacritics to mark features such as vowel length, vowel nasalization, and aspiration. None of the subjects in this group considered themselves to be experienced musicians, although approximately half of the listeners had received some form of musical training as recently as early high school.

B. Stimuli

Two nine-step speech-sound continua were synthesized using an implementation of Klatt's (1980) software formant synthesizer. All stimuli consisted of five formants and were synthesized with a 10-kHz sample frequency and 12 bits of amplitude resolution. One continuum ranged from [bi]–[p^hi] by varying VOT from 0 to 80 ms in 10-ms steps. VOT was manipulated using procedures described by Abramson and Lisker (1965). All of the stimuli began with 5 ms of frication, followed by aspiration for the duration of the VOT interval. The voicing source was initiated immediately following the end of the aspiration interval. Starting and ending frequencies of the first three formants, and the durations of linear transition intervals, are shown in Table I. The fundamental frequency rose linearly from 120–125 Hz over the first 40 ms, then fell linearly to 90 Hz over the remaining 210 ms of the signal. The fourth and fifth formants were set to 3300 and 3700 Hz, respectively, for the entire 250-ms duration of the signals.

Stimuli for a second synthetic continuum varied in nine steps from [wei] to [rei] by varying the onset frequencies of $F2$ and $F3$. Synthesis parameters were taken from Sharf and Benson (1982). Table II shows the contours of the first three formants over the 300-ms course of the two endpoint stimuli. Formant frequencies for $F2$ and $F3$ were changed approximately 5% between individual stimuli on the nine-step continuum. Further details about the stimuli can be found in Sharf and Benson.

C. Procedures

A computer program was used to adjust all stimuli on both continua to the same overall rms intensity. At the output of the D/A converter, stimuli were low-pass filtered at 4 kHz, amplified, attenuated, and delivered binaurally through TDH-49 earphones to subjects seated in a sound-treated booth. A sound-level meter (Bruel & Kjaer 2203) and artificial ear (Bruel & Kjaer 4142) were used to adjust the peak intensity of the speech signals to 75 dBA. Presentation of stimuli, collection of listeners' responses, and measurement of response latencies were controlled by a PDP 11 computer.

Each subject completed a total of six tasks: a two-choice identification task (90 trials), a three-choice identification task (90 trials), and an ABX discrimination task (140 trials) for the [b]–[p] continuum and for the [w]–[r] continuum. The six listening tasks could be completed in a single session lasting 45–50 min. All subjects began testing with a two-choice identification task, followed by a three-choice identification task for the same continuum. For a random half of the subjects, testing began with [b]–[p]; the other half began with [w]–[r]. Following the four identification tasks, subjects participated in ABX discrimination testing on each continuum, with the order of presentation counterbalanced across the two stimulus series.

The identification tasks consisted of ten randomly ordered presentations of the nine stimuli on each continuum. For the two-choice task, subjects indicated whether the stimulus sounded like "b" or "p" (or "w" vs "r") by pressing the

TABLE I. Starting and ending values of the first three formant frequencies and the durations of linear formant transitions (in ms) for the synthetic [bi]–[p^hi] continuum.

Formant	Starting frequency	Ending frequency	Transition duration
$F1$	180	286	20
$F2$	1465	2307	40
$F3$	2180	3026	40

TABLE II. Contours of $F1$-$F3$ for endpoint stimuli on the nine-step synthetic [wei]–[rei] continuum (from Sharf and Benson, 1982). Formant frequencies were changed approximately 5% between adjacent stimuli on the continuum.

Time (ms)	0	40	110	140	235	300
Parameter			[wei] end point			
$F1$	300	300	471	471	330	330
$F2$	700	700	1700	1700	2000	2000
$F3$	2400	2400	2500	2500	2600	2600
			[rei] end point			
$F1$	300	300	471	471	330	330
$F2$	1100	1100	1700	1700	2000	2000
$F3$	1600	1600	2500	2500	2600	2600

appropriate button on a response box. Instructions for the three-choice identification task for the [b]–[p] continuum were as follows:

> You will be hearing the same set of sounds that you were just tested on (in a different order), but this time you will be using three rather than two buttons. Your job is to press button No. 1 if you hear a clear 'b,' button No. 3 if you hear a clear 'p,' and button No. 2 if the sound has an 'indefinite' quality; that is, somewhat like 'b' and somewhat like 'p.'

The ABX task consisted of five randomly ordered presentations of all possible two-step ABB, ABA, BAA, and BAB triads, for a total of 140 trials. The interstimulus interval was 0.5 s and the intertrial interval was 1.0 s, measured from the subject's response. Subjects were told that the first two stimuli in the triad would always be different, and that the third stimulus would match either stimulus one or stimulus two. Listeners were instructed to press the appropriate button on the response box to indicate whether the third stimulus matched the first or second stimulus in the triad. Subjects received no feedback on the accuracy of their responses.

II. IDENTIFICATION RESULTS

A. Two-choice identification

Figure 1 shows results for the [b]–[p] and [w]–[r] two-choice identification tasks. Results are shown for the two groups of trained listeners, and for the inexperienced listeners before training. (Results for the inexperienced listeners after training will be discussed in a separate section.) It can be seen that the three groups of subjects performed very similarly on the two-choice identification tasks. Category boundaries were calculated for each subject by linear interpolation of the 50% point on the identification function. Two separate one-way ANOVAs showed that the three groups did not differ significantly in the location of two-choice category boundaries ([b]–[p]: $F(2,54) = 2.9$, p n.s.; [w]–[r]: $F(2,54) = 0.1$, p n.s.).

B. Three-choice identification

Middle-category responses from the three-choice identification task for the three groups of listeners are shown in Fig. 2. For both the [b]–[p] and [w]–[r] continua, it can be seen that the phonetically trained listeners produced a well-defined "indefinite" category corresponding closely to the category boundary in the two-choice identification task. The middle category shows a peak of approximately 72% for the [b]–[p] continuum and 75% for the [w]–[r] continuum. The musicians and inexperienced listeners also produced a middle-category peak corresponding to the two-choice category boundary, but the peak heights were substantially smaller than those associated with the phonetically trained listeners.

A difference score was calculated using middle-category responses in an effort to quantify the degree of consistency with which subjects in the three groups identified ambiguous stimuli. The basic idea was to compare responses to stimuli near the two-choice category boundary with responses to stimuli remote from the boundary. The first value in the difference score was a weighted average of responses to the

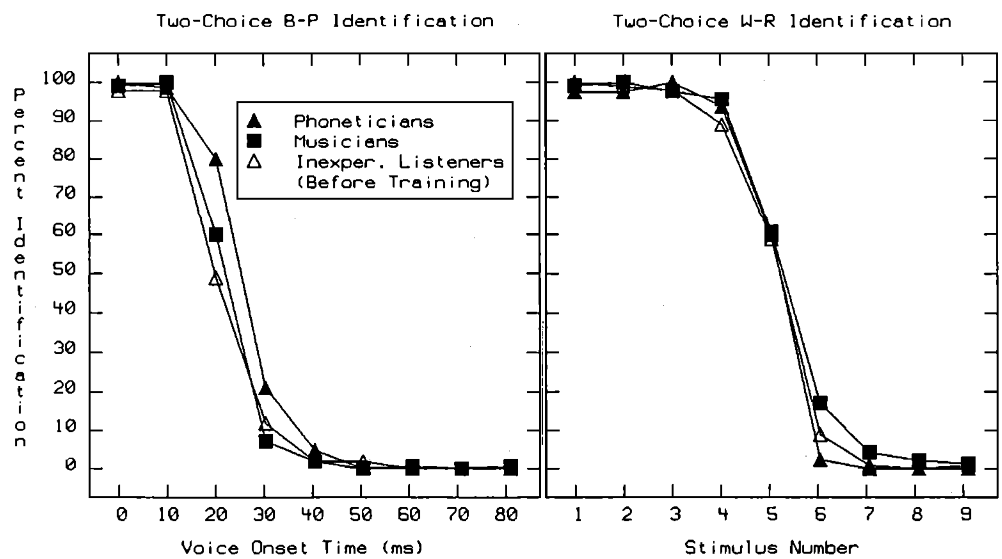

FIG. 1. Identification functions from the two-alternative forced-choice listening task.

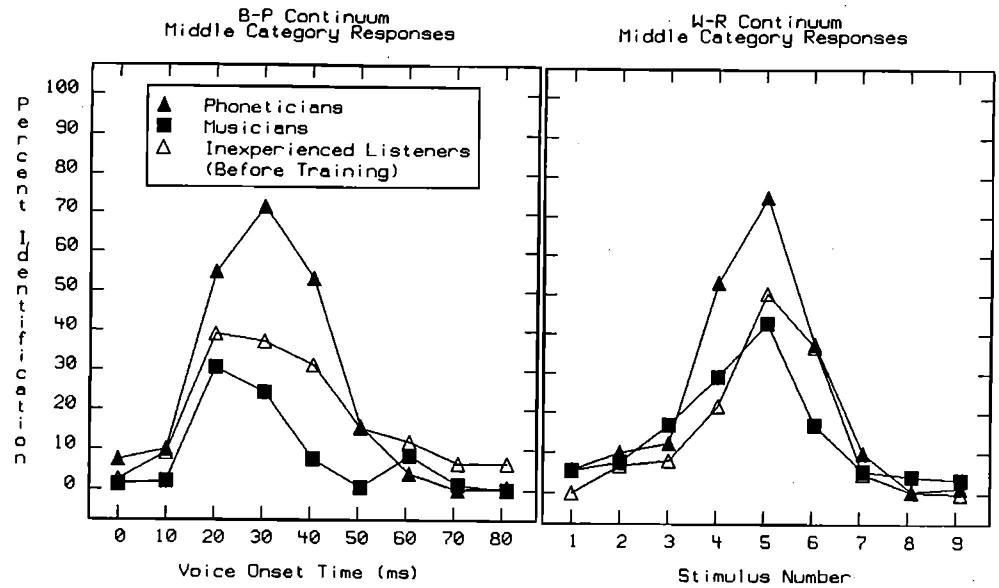

FIG. 2. Middle-category responses from the three-choice identification task. Responses represent trials in which subjects judged stimuli to have an "ambiguous" or "indefinite" quality.

stimulus closest to the subject's two-choice category boundary and the two stimuli on either side of the boundary stimulus. Responses to the boundary stimulus received a weight of 1.0, and responses to the two near-boundary stimuli received a weight of 0.5. The second value in the difference score was the percentage of middle-category responses to stimuli remote from the subject's two-choice category boundary, where "remote" was defined as two or more stimulus units from the category boundary. Mean and standard deviation difference scores for subjects in the three groups are shown in Table III.

A two-way ANOVA calculated on the difference scores yielded a significant effect for Group [$F(2,79) = 14.2$, $p < 0.01$], a nonsignificant effect for Stimulus Continuum [$F\ (1,79) = 1.04$, p n.s.], and a nonsignificant Group by Stimulus Continuum interaction [$F\ (2,79) = 0.9$, p n.s.]. *Post hoc* analyses showed that the group effect was significant for both [b]–[p] [$F\ (2,39) = 5.7, p < 0.01$] and [w]–[r] [$F(2,39) = 4.1$, $p < 0.05$]. Further tests using Newman–Keuls multiple comparisons showed that, for both continua, the phonetically trained listeners performed better than the musicians and the inexperienced listeners, and that the musicians and inexperienced listeners did not differ on either contrast.

Despite the significant group differences, it is important to note that there were several individual musicians and inexperienced listeners who performed quite well on the three-choice task. Based on informal visual examination of the data, approximately 30% of the musicians and inexperienced listeners produced functions on the three-choice task that were indistinguishable from those of the phonetically trained listeners. The converse of this was not true, however; that is, none of the phonetically trained listeners performed poorly on the three-choice task.

III. DISCRIMINATION RESULTS

One possible interpretation of the three-choice identification results is that phonetically trained listeners are more sensitive to intraphonemic differences than listeners without training in phonetics. However, it is also possible that the effect is related primarily to criterion differences between phonetically trained listeners and the other subjects. The ABX discrimination tests were designed to determine whether the groups differed in their sensitivity to small stimulus differences along the two continua.

Discrimination results for the three groups are shown in Fig. 3. For the inexperienced listeners, data are shown for subjects tested *prior* to their participation in the introductory phonetics course. The data shown in the figure are d' measures calculated from the ABX discrimination responses (Macmillan, Kaplan, and Creelman, 1977). As the figure shows, the same basic trends that were seen in the three-choice identification results are reflected in the d' discrimination measures. For both speech-sound continua, discrimination performance was best for the phonetically trained listeners, and little or no difference was seen between the

TABLE III. Difference scores based on middle-category responses for the three groups of listeners for [bi]–[p^hi] and [wei]–[rei] continua (standard deviations are shown in parentheses). The scores represent responses to boundary and near-boundary stimuli minus responses to stimuli remote from the subject's two-choice category boundary.

	B–P	W–R
Phoneticians	55.4 (22.7)	50.3 (13.2)
Musicians	18.2 (20.9)	23.8 (23.6)
Inexperienced listeners (before training)	10.2 (34.5)	24.9 (38.7)

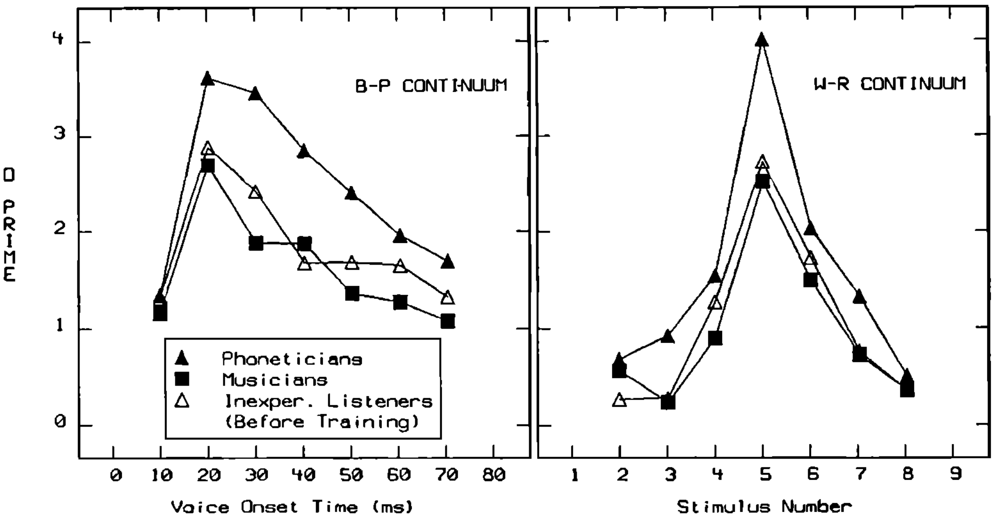

FIG. 3. Discriminability measures from the ABX discrimination task.

musicians and the inexperienced listeners.

Statistical analysis consisted of separate Group by Stimulus Contrast ANOVAs for each continuum. Results for both continua showed a significant effect for Group ([b]–[p]: F (2,39) = 10.6, $p<0.01$; [w]–[r]: F (2,39) = 39.3, $p<0.01$) and Stimulus Contrast ([b]–[p]: F (6,34) = 26.3, $p<0.01$; [wei]–[rei]: F (6,34) = 77.4, $p<0.01$). For both continua, Newman–Keuls *post hoc* analyses showed that the phonetically trained listeners performed better than both groups of nonphonetically trained listeners, and that the musicians and inexperienced listeners did not differ from one another. The Group by Stimulus Contrast interaction was significant only for the [w]–[r] continuum [F (68,12) = 4.2, $p<0.01$]. *Post hoc* analyses indicated that the interaction was attributable to larger group differences near the peak in the discrimination function.

It should also be noted that all of the listeners, including the phoneticians, showed the prominent peak at the two-choice labeling boundary that is typical of categorical perception. This contrasts with the findings of Carney *et al.* (1977), who noted that the performance of their listeners following discrimination training was more consistent with a *noncategorical* perception model. This discrepancy between their findings and ours is almost certainly related to the low stimulus-uncertainty procedure used in the Carney *et al.* study as compared with the high-uncertainty procedure used in the present study (see also Kewley-Port *et al.*, 1988; Sachs and Grant, 1976). The noncategorical functions obtained by Carney *et al.* may also be related in part to the specific training that their subjects received using the synthetic stimuli being tested.

IV. EFFECTS OF PHONETICS TRAINING

The performance of subjects before and after participation in a 10-week introductory phonetics course is shown in Fig. 4 (three-choice identification, middle-category responses) and Fig. 5 (d' measures calculated from the ABX discrimination responses). For both tasks, it can be seen that subjects' performance following the 10-week phonetics course was very similar to their performance prior to training. For the three-choice identification data, a two-way repeated-measures ANOVA was calculated for Stimulus Continuum ([b]–[p] vs [w]–[r]) and Training using the difference scores described previously. The analysis produced nonsignificant effects for both factors (Stimulus Continuum: F (1,14) = 2.8, p n.s.; Training: F (1,14 = 0.7, p n.s.). For the d' measures, separate two-way ANOVAs showed significant effects for Stimulus Contrast ([b]–[p]: F (6,84) = 4.08, $p<0.01$; [w]–[r]: F (6,84) = 44.9, $p<0.001$) but no effects for Training ([b]–[p]: F (1,14) = 0.2, p n.s.; [w]–[r]: F (1,14) = 0.9, p n.s.).

V. DISCUSSION

The main finding of this study is that phonetically trained listeners were better than either musicians or inexperienced listeners at identifying speech sounds whose stimulus values were intermediate between two phonetic categories. The important point to be made about this result is that the performance advantage for the phonetically trained listeners was obtained *in the absence of specific training with the stimuli.* The same pattern of performance differences among the groups was seen for an ABX discrimination task, suggesting that the group effect was not limited to any differences that might have existed in decision criteria. It is important to note, however, that the two groups of listeners without phonetics training showed some ability to identify ambiguous stimuli. This finding suggests that the effect of long-term phonetics training was to enhance an ability that was already present in these subjects. In fact, as was noted previously, the three-choice labeling performance of some of the individual listeners in the nonphonetically trained groups was indistinguishable from that of the phonetically trained listeners.

Comparison of the performance of the inexperienced listeners before and after participation in an introductory phonetics course showed no improvement in either the absolute identification of ambiguous stimuli or the ability to discriminate stimuli. This finding indicates quite clearly that the

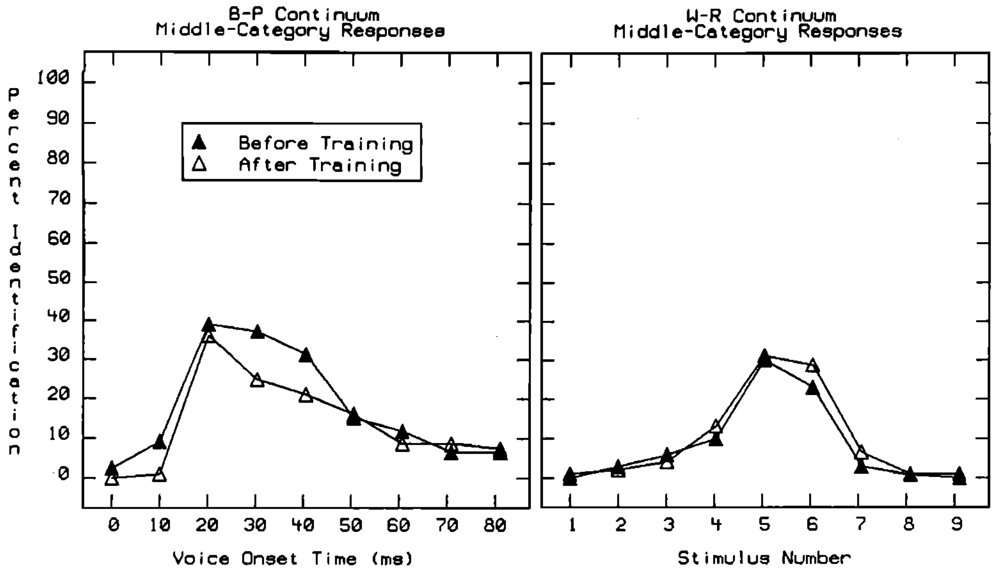

FIG. 4. Middle-category responses from the three-choice identification task for subjects in the inexperienced listener group before and after their participation in an introductory phonetics course.

identification and discrimination abilities that were shown by the phonetically trained listeners did not emerge as a result of relatively short-term training. It has long been known that very extensive training is required for subjects to perform well on certain kinds of listening tasks. For example, Bryan and Harter's (1899) study of the acquisition of Morse code showed that subjects were far from asymptotic performance after 40 weeks of daily training, and Meyer (1899) showed that 24–28 weeks were required to reach asymptote on a musical pitch identification task. In reviewing this literature, Watson (1980) concluded that the conditions that are associated with very long training intervals involve complex stimuli, listening tasks that require subjects to identify rather than detect or discriminate stimuli, and tasks in which the number of response categories is large. Narrow phonetic transcription obviously meets all of these criteria. Viewed in this light, it seems very likely that the 10 weeks of training in phonetic transcription left the introductory phonetics students well short of asymptotic performance on these kinds of tasks.

One practical implication of the results from the phonetically trained listeners concerns the widespread use of narrow phonetic transcriptions in evaluating speech disorders and in conducting research on speech disorders. The presence of the categorical perception phenomenon might be used to argue that many of the fine details that are recorded in a narrow phonetic transcription represent within-category variations that should not be discernible to a listener. An especially strong case might be made against the widespread practice of classifying speech sounds as "distorted." The present results suggest that experienced phoneticians and clinical speech pathologists can, in fact, make fine-

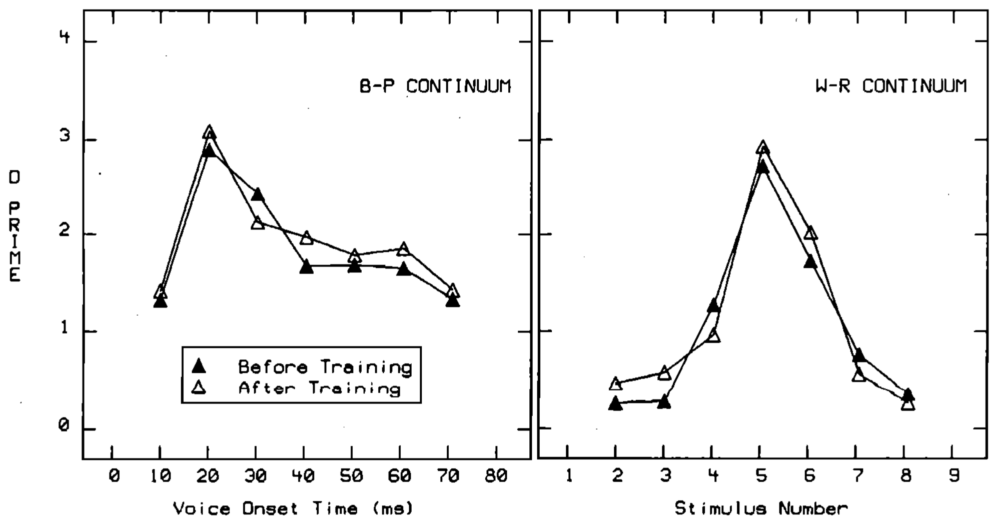

FIG. 5. Discriminability measures from the ABX discrimination task for subjects in the inexperienced listener group before and after their participation in an introductory phonetics course.

grained phonetic distinctions that represent acoustic variations within a phonetic category.

The results from the musically trained listeners suggest that the performance advantage that was seen for the phonetically trained listeners is not related to any general auditory ability that might be common to musicians and phoneticians. In other words, in the absence of data from the musicians, it might have been argued that the phoneticians were simply more attentive listeners, or that they had developed a more-or-less general-purpose analytical listening ability. It seems a safe assumption, however, that highly trained musicians have learned to listen attentively to sound and have had extensive experience in making critical judgments about complex acoustic signals. The failure to find any difference between the musicians and the untrained listeners suggests that the performance of the phonetically trained listeners can be attributed either to: (1) specific experience with speech sounds or (2) general experience with the kinds of absolute labeling tasks that are associated with phonetic transcription.

ACKNOWLEDGMENTS

This research was supported in part by NIH Grant No. 1-R01-NS-22234 to Northwestern University and by NIH Grant No. 5-R01-NS23703 to Western Michigan University. We are very grateful to Thea Burton, Jan Wasowicz, Thomas Ridley, and Jeffrey Reynolds for their help with data collection and data analysis, and to the research subjects who donated their time to this project. We would also like to thank Charles Watson for his help with the literature on long-term auditory learning, James Flege and Michael Clark for their advice on a variety of aspects of the interpretation of the experiments, and Bruno Repp and Kenneth Grant for their thoughtful reviews of this manuscript.

Abramson, A. S. (**1961**). "Identification and discrimination of phonemic tones," J. Acoust. Soc. Am. **33**, 842 (abs).

Abramson, A. S., and Lisker, L. (**1965**). "Voice onset time in stop consonants: Acoustic analysis and synthesis," in *Proceedings of the Sixth International Congress on Acoustics*, edited by D. E. Commins (Imp. G. Thone, Liege, Belgium).

Bastian, J., and Abramson, A. S. (**1964**). "Identification and discrimination of phonemic vowel duration," in *Speech Research and Instrumentation* (Haskins Laboratories, New York) (Tenth final report).

Bastian, J., Eimas, P., and Liberman, A. M. (**1961**). "Identification and discrimination of a phonemic contrast induced by silent interval," J. Acoust. Soc. Am. **33**, 842 (abs).

Bryan, W. L., and Harter, N. (**1899**). "Studies in the physiology and psychology of the telegraphic language. The acquisition of a hierarchy of habits," Psychol. Rev. **6**, 345–375.

Carney, A. E., Widin, G. P., and Viemeister, N. F. (**1977**). "Noncategorical perception of stop consonants differing in VOT," J. Acoust. Soc. Am. **62**, 961–970.

Chandler, A. E., and Strange, W. (**1984**). "Intraphonemic perception of a synthetic /r/-/l/ series by musicians and nonmusicians," J. Acoust. Soc. Am. Suppl. 1 **76**, S27–S28 (A).

Cross, D. V., Lane, H. L., and Sheppard, W. C. (**1965**). "Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception," J. Exp. Psychol. **70**, 63–74.

Edman, T. R. (**1979**). "Discrimination of intraphonemic differences along two places of articulation continua," in *Speech Communication Papers*, edited by J. J. Wolf and D. H. Klatt (Acoustical Society of America, New York), pp. 455–458.

Edman, T. R., Soli, S. D., and Widin, G. P. (**1978**). "Learning and generalization of intraphonemic VOT discrimination," J. Acoust. Soc. Am. Suppl. 1 **63**, S19.

Eimas, P. D. (**1963**). "The relation between identification and discrimination along speech and nonspeech continua," Lang. Speech **6**, 206–217.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (**1962**). "The identification and discrimination of synthetic vowels," Lang. Speech **5**, 171–189.

Ganong, W. F. III (**1977**). "Selective adaptation and speech perception," Unpublished doctoral dissertation, Massachusetts Institute of Technology (Cambridge, MA).

Hanson, V. L. (**1977**). "Within-category discrimination in speech perception," Percept. Psychophys. **21**, 423–430.

Harned, S. (**1987**). *Categorical Perception: The Groundwork of Cognition* (Cambridge U. P., Cambridge, MA).

Kewley-Port., D., Watson, C. S., and Foyle, D. C. (**1988**). "Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli," J. Acoust. Soc. Am. **83**, 1133–1145.

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

Kopp, J., and Livermore, J. (**1973**). "Differential discrimination or response bias? A signal detection analysis of categorical perception," J. Exp. Psychol. **101**, 179–182.

Kopp, J., and Udin, H. (**1969**). "Identification and discrimination functions for pure tone frequencies," Psychonom. Sci. **16**, 95–96.

Lane, H. L. (**1965**). "Motor theory of speech perception: A critical review," Psychol. Rev. **72**, 275–309.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **53**, 358–368.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., and Bastian, J. (**1961a**). "An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance," Lang. Speech **54**, 175–195.

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. (**1961b**). "The discrimination of relative onset time of the components of certain speech and nonspeech patterns," J. Exp. Psychol. **61**, 379–388.

Liberman, A. M., Studdert-Kennedy, M., Harris, K. S., and Cooper, F. S. (**1965**). "A reply to 'Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception' by Cross, Lane, and Sheppard," Haskins Laboratories Status Rep. on Speech Research, SR-3, 3.1–3.14.

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). "The psychophysics of categorical perception," Psychol. Rev. **84**, 452–471.

Meyer, M. L. (**1899**). "Is the memory of absolute pitch capable of development by training?," Psycholog. Rev. **6**, 514–516.

Parks, T., Wall, C., and Bastian, J. (**1969**). "Intercategory and intracategory discrimination for one visual continuum," J. Exp. Psychol. **81**, 241–245, 686–696.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). "Some effects of laboratory on identification and discrimination of voicing contrasts in stop consonants," J. Exp. Psychol.: Human Percept. Performance **8**, 297–314.

Repp, B. H. (**1975**). "Categorical perception, dichotic interference, and auditory memory: A 'same-different' reaction time study" (Haskins Laboratories, New Haven, CT) (unpublished).

Repp, B. H. (**1981**). "Two strategies in fricative discrimination," Percept. Psychophys. **30**, 217–227.

Repp, B. H. (**1984**). "Categorical perception: Issues, Methods, Findings," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), Vol. 10, pp. 243–335.

Sachs, R. M., and Grant, K. W. (**1976**). "Stimulus correlates in the perception of voice onset time (VOT): Discrimination of speech with high and low stimulus uncertainty," J. Acoust. Soc. Am. Suppl. 1 **60**, S91.

Samuel, A. G. (**1977**). "The effect of discrimination training on speech perception: Noncategorical perception," Percept. Psychophys. **22**, 321–330.

Sharf, D. J., and Benson, P. J. (**1982**). "Identification of synthesized /r-w/ continua for adult and child talkers," J. Acoust. Soc. Am. **71**, 1008–1015.

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

Watson, C. S. (**1980**). "Time course of auditory perceptual learning," Ann. Otol. Rhinol. Laryngol. **89**, 96–102.