Perception & Psychophysics
1982, 31 (3), 210-218

Some experiments on perceptual learning of mirror-image acoustic patterns

MARY ELLEN GRUNKE and DAVID B. PISONI
Indiana University, Bloomington, Indiana 47405

It is well known that the formant transitions of stop consonants in CV and VC syllables are roughly the mirror image of each other in time. These formant motions reflect the acoustic correlates of the articulators as they move rapidly into and out of the period of stop closure. Although acoustically different, these formant transitions are correlated perceptually with similar phonetic segments. Earlier research of Klatt and Shattuck (1975) had suggested that mirror image acoustic patterns resembling formant transitions were not perceived as similar. However, mirror image patterns could still have some underlying similarity which might facilitate learning, recognition, and the establishment of perceptual constancy of phonetic segments across syllable positions. This paper reports the results of four experiments designed to study the perceptual similarity of mirror-image acoustic patterns resembling the formant transitions and steady-state segments of the CV and VC syllables /ba/, /da/, /ab/, and /ad/. Using a perceptual learning paradigm, we found that subjects could learn to assign mirror-image acoustic patterns to arbitrary response categories more consistently than they could do so with similar arrangements of the same patterns based on spectrotemporal commonalities. Subjects respond not only to the individual components or dimensions of these acoustic patterns, but also process entire patterns and make use of the patterns' internal organization in learning to categorize them consistently according to different classification rules.

In natural speech, a single phonetic segment may have many different acoustic representations, depending on the context in which it is spoken. An extensive body of research over the last 30 years has been directed towards trying to identify acoustically invariant properties of phonemes that can mediate speech perception. For example, stop consonants that follow a vowel are roughly the mirror image in time of their counterparts that precede the same vowel—the /b/ in the syllable /ba/ typically has rising formant transitions into the vowel following release, whereas the /b/ in /ab/ has falling formant transitions into the period of stop closure. A similar situation has been observed with /d/ in the syllables /da/ and /ad/—the /d/ is characterized by falling second and third formant transitions into the vowel and rising formant transitions out of the vowel.

A child who is acquiring language somehow learns to recognize consonants that occur in different positions as members of the same phonetic category. Can this acquisition process occur through acoustically based similarity between mirror-image patterns? That is, do mirror-image acoustic patterns share perceptually salient features that cause them to sound alike or to be classified together by listeners?

One approach to answering these questions has been to use a selective adaptation paradigm (e.g., Ades, 1974; Pisoni & Tash, 1975; Wolf, 1978). In general, this research has failed to find evidence that acoustical invariants can provide a basis for identification of stop consonants in different syllable positions. Ades (1974) reported that repeated presentation of a CV syllable had an adapting effect on a CV syllable continuum but not on a mirror-image VC continuum. On the other hand, results of Pisoni and Tash's (1975) experiments on CV and VC syllable showed adaptation effects across syllable position and suggested that there might be auditory property detectors that respond to rising or falling spectral changes in the speech signal. In the absence of other information, rise-fall detectors could provide one way of categorizing the same consonant in pre- and postvocalic position. However, in a more recent study, Wolf (1978) considered a variety of acoustic properties, including identical release bursts, mirror-image formant transitions, and similar onset and offset spectra, and found none to be the basis for the invariant perception of place of articulation in initial and final syllable position, at least as revealed through selective adaptation techniques.

In addition to selective adaptation experiments, another approach to the question of the perceptual

This research was supported by NIMH Research Grant MH-24027, NIH Research Grant NS-12179, an NIMH postdoctoral fellowship to Mary Ellen Grunke, and a fellowship from the Guggenheim Foundation to David B. Pisoni. An earlier report of some of these findings was presented at the Acoustical Society of America meetings in Cambridge, Massachusetts, June 1979, by Mary Ellen Grunke. We thank Robert E. Remez for his helpful comments on an earlier version of the paper.

0031-5117/82/030210-09$01.15/0

relatedness of mirror-image acoustic patterns has been to use nonspeech sounds in a similarity judgment task. Here, again, the data have not provided very convincing support for the hypothesis that mirror-image acoustic patterns are perceptually similar for a listener. Klatt and Shattuck (1975) and Shattuck and Klatt (1976) had listeners judge the similarity of brief pure-tone frequency glissandos. They found that similarity judgments of two-component patterns—diverging, converging, both rising, or both falling—were based primarily on the direction of the lower glissando component.

The approach adopted to the issue of mirror-image perceptual relatedness in the current experiments was to use nonspeech acoustic patterns of a slightly more complex nature than those used earlier by Klatt and Shattuck. Our patterns contained a steady-state constant frequency (CF) interval in addition to a rapid frequency modulation (FM) at either onset or offset. We also used a perceptual learning task in which listeners were trained to sort four different acoustic patterns into two discrete response categories. The categories were defined according to: (1) the direction of the frequency transition, (2) the temporal occurrence of the frequency transition relative to the steady state, or (3) a phonetic-like classification (i.e., mirror-image patterns paired with the same category). The question of principal interest was which kind of categorization would produce fewest errors during acquisition. In order to determine which perceptual features of the patterns were most salient to listeners, a perceptual learning task was used in Experiments 1 and 2 with a variety of sets of acoustic patterns. In a third experiment, listeners were asked to assign either acoustic or phonetic labels to the same stimulus patterns. This experiment provided a way of determining how accurately the patterns could be heard as speech or, alternatively, as nonspeech frequency-modulated glissandos. Finally, in the last experiment, similarity judgments were collected in order to compare our results directly with those obtained earlier by Klatt and Shattuck, who had used simpler signals.

EXPERIMENT 1
PERCEPTUAL LEARNING TASK

The first experiment used a perceptual learning task and compared acquisition performance among three categorization conditions based on: (1) mirror-image relation, (2) the direction of glissandos, or (3) the relative temporal positions of a glissando and a steady-state frequency. Three sets of acoustic patterns, containing a single tone or double- or triple-tone combinations were used as test signals.

Method

Stimuli. Three sets of stimuli containing four stimuli per set, were generated using a program that combines sine waves to form complex tones (Kewley-Port, Note 1). The stimuli are shown schematically in Figure 1.

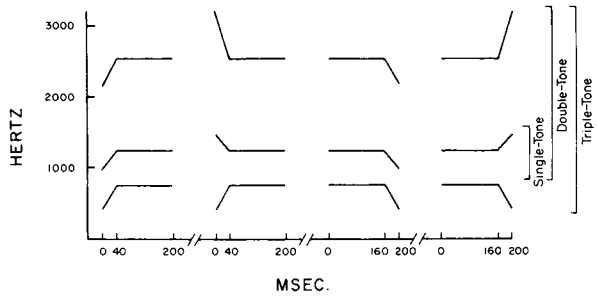

Figure 1. Schematic spectrographic patterns of the nonspeech stimuli used in Experiment 1.

Each stimulus component consisted of a 40-msec linear rise or fall in frequency (FM) and a 160-msec constant-frequency (CF) segment. The four stimuli within a set differed in whether the frequency transitions were rising or falling and in whether the transition preceded (initial) or followed (final) the steady-state portion of the pattern.

The three stimulus sets also differed in the number of component tones, one, two, or three. Frequency values of these tones were selected to correspond to the values of the first, second, and third formants in the synthetic syllables /ba/, /da/, /ab/, and /ad/. The patterns in the single-tone set were in the second-formant region. The steady-state portion was set at 1,230 Hz with transition endpoints of 995 Hz for the "b" and at 1,465 Hz for the "d" stimuli. The component frequencies in the double-tone set corresponded to the second and third formants. The steady-state portion of the tone corresponding to the third formant was set at 2,525 Hz; transition endpoints were 2,180 and 3,195 Hz for the "b" and "d" stimuli, respectively. The lower tones of the double-tone stimuli were identical in frequency to the tones used in the single-tone set. All three formants were represented in the triple-tone set, although the frequency transition corresponding to the first formant always rose when it preceded the steady state and fell when it followed the steady state, in accordance with the formant motions observed in CV syllables in natural speech. The two upper components of the triple-tone stimuli were the same as the two components in the double-tone set; the lowest tone consisted of a 770-Hz constant-frequency portion and a frequency modulation rising from or falling to 400 Hz.

Procedure. Stimulus presentation and data collection for all experiments reported here were controlled on-line in real time by a PDP-11/05 computer. The stimuli were presented to subjects at a comfortable listening level of about 80 dB SPL via TDH-39 headphones. The subjects were seated in front of a response panel in a sound-attenuated cubicle. In Experiment 1, the response panel contained two response buttons, a feedback light above each response button, and a cue light that signaled the presentation of each tone 1 sec prior to its onset. The two response buttons were labeled simply "R1" and "R2."

The subjects were required to learn to sort the four acoustic patterns in a given stimulus set into two response categories (R1 and R2) according to one of the three mapping or categorization rules outlined earlier. An experimental session consisted of 12 alternating study and test periods. During study periods, five repetitions of each stimulus were presented. First, all "R1" repetitions were heard in random order; these were followed by all "R2" repetitions. The feedback lights indicated the correct response for each stimulus after it was presented. Test periods consisted of 20 trials, 5 per stimulus. On each trial, a stimulus was presented and the subjects were required to respond by pressing one of the two buttons. After each response, the feedback lights came on and indicated the correct response category. The order of the four stimuli was randomized throughout the test periods.

Subjects were assigned to one of three categorization conditions.

In Condition 1, *the mirror-image condition*, stimuli with a rising transition preceding the steady-state (/———) or a falling transition following the steady-state (———\) were assigned to one response, while stimuli with a falling transition preceding the steady-state (\———) or a rising transition following the steady-state (———/) were assigned to the other response. In Condition 2, *the rise-fall condition*, the stimuli with rising transitions that either preceded or followed the steady-state (/———/) were assigned to one response and the two stimuli with falling transitions were assigned to the other response. Finally, in Condition 3, *the temporal position condition*, the stimuli were assigned to responses according to the temporal position of the transition, either preceding or following the steady-state portion. The subjects were never explicitly told the particular rules defining category membership, although in some cases the arrangements were very obvious after listening to a few trials. Nine separate conditions were formed by the factorial combination of three stimulus sets (single, double, and triple tones) and three categorization rules (mirror-image, rise-fall, and temporal position).

Subjects. Ten subjects, recruited from introductory psychology courses at Indiana University, were assigned randomly to each of the nine experimental conditions. In this and all subsequent experiments, we used only subjects who reported no history of a speech or hearing disorder at the time of testing.

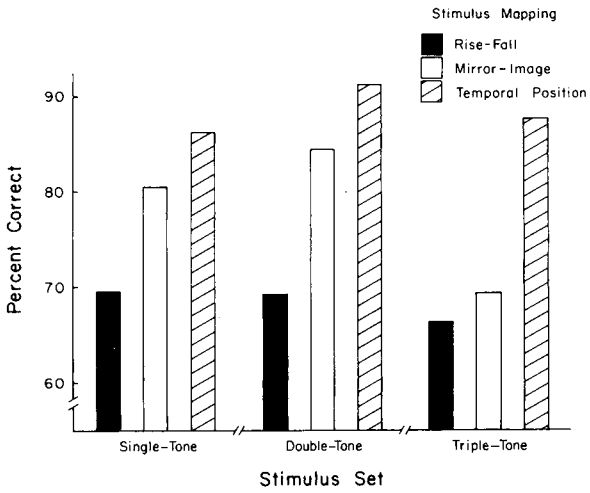

Figure 2. Percent correct performance pooled across subjects and individual learning trials for each stimulus set. The data are displayed separately for each response-mapping (i.e., categorization) condition.

Results and Discussion

Of the three categorization conditions, temporal position showed the most accurate performance, 88.4% correct, pooled across all test trials. This condition required only that listeners be able to distinguish the relative temporal order of the steady-state and frequency-modulated portions of the stimuli. Of the two categorization conditions that required selective responding to particular properties of the frequency glissando, performance was better in the mirror-image condition (78.1% correct) than in the rise-fall condition (68.3% correct). With respect to the three stimulus sets, the triple-tone set produced slightly poorer performance (74.5%) than either the double-tone (81.6%) or single-tone (78.7%) sets. Mean percent correct for each stimulus set and categorization (i.e., mapping) condition are shown graphically in Figure 2.

An analysis of variance confirmed the statistical significance of the main effects of mapping condition [$F(2,81) = 35.37$, $p < .001$] and stimulus set [$F(2,81) = 4.55$, $p < .05$]. Follow-up Scheffé tests of pairwise comparisons were performed using an overall significance level of .05. These post hoc tests indicated that all three categorization conditions differed reliably from one another. However, with respect to stimulus set, only the double tones were significantly better than the triple tones in response accuracy.

The overall analysis of variance failed to show a significant interaction between stimulus set and categorization condition ($p > .10$). However, the magnitude of the increase of the mirror-image arrangement over the rise-fall arrangement was 16% and 22% for the single and double tones, respectively, but only 5% for the triple tones. In subsequent experiments replicating these conditions, we observed very similar percentage increases in performance, that is, a relatively small improvement of the mirror-image arrangement over the rise-fall arrangement with triple tones and a considerably larger improvement with the double tones.

In considering accuracy levels for correct categorization of the individual stimuli, performance for transition-initial stimuli (69.0%) was markedly poorer than it was for transition-final stimuli (87.2%) [$F(1,81) = 167.02$, $p < .001$]. In Figure 3, percent correct performance for the "ba," "da," "ab," and "ad" stimuli is shown separately for each of the three mapping arrangements.

Subjects in all three categorization conditions had

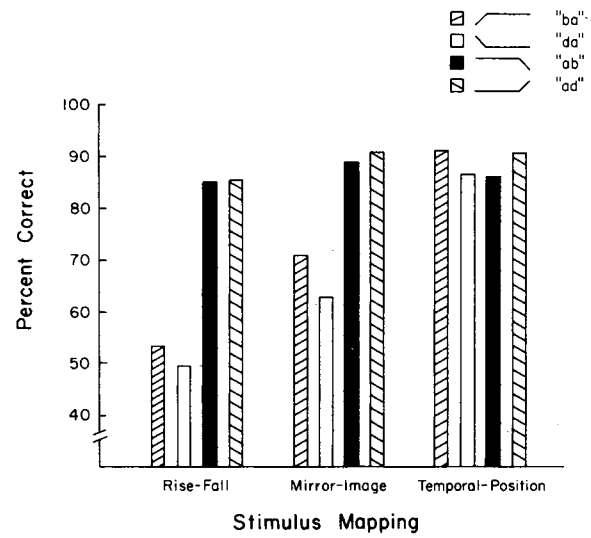

Figure 3. Percent correct performance for individual stimuli under the three mapping (i.e., categorization) conditions of the experiment.

little difficulty in assigning the final-falling ("ab") and final-rising ("ad") tones to the appropriate response category. Likewise, the higher performance level with temporal position mapping indicates that subjects could easily distinguish initial-transition stimuli from final-transition stimuli, a result that is not entirely surprising, since the individual components of the pattern do not need to be analyzed.

The major source of difficulty in this task was learning to respond selectively to the initial-rising ("ba") and initial-falling ("da") signals. The superiority of mirror-image over rise-fall mapping demonstrates that it was easier to learn to assign "ba" (or "da") and "ab" (or "ad") tones to a common response category than it was to learn that "ba" went with "ad" (both rising transitions) and "da" with "ab" (both falling transitions). Thus, despite earlier reports by Klatt and Shattuck (1975), our results demonstrate that mirror-image patterns are easier to group together than nonmirror-image patterns in a perceptual learning task.

EXPERIMENT 2
RELEVANT STIMULUS PROPERTIES

In view of the advantage of the mirror-image condition over the rise-fall condition, it was of some interest to determine the particular acoustic features or properties that listeners extracted from the stimuli in the mirror-image condition. That is, which perceptually salient properties served as reliable discriminative cues for learning the mapping rules in the mirror-image condition? Two properties of these patterns were investigated in Experiment 2. The first, short-term spectral composition of the frequency transitions, is specific to only the transitional portion of each stimulus pattern. The second, frequency averaged across both steady state (CF) and transition (FM), is a configural property of each signal as a whole. Changes in the average frequency of these stimulus patterns can be heard as changes in the perceived pitch of the signal as a whole. For the purposes of this study, we equate overall average frequency with perceived pitch.

In the acoustic patterns used in Experiment 1, both the average frequency and the transition endpoints were the same for signals within mirror-image pairs but were different across pairs. In contrast, transition endpoints and average frequency differed within rise-fall pairs, and thus this dimension could be considered an irrelevant cue for learning a rise-fall categorization rule. For Experiment 2, two additional stimulus sets were generated in which either transition endpoints or average frequency were held constant. Thus, these dimensions would not be available as discriminative cues to help the listener sort the stimuli into the correct categories. Would the advantage of the mirror-image over rise-fall categorization condition disappear under either of these new stimulus conditions? If elimination of a particular discriminative cue resulted in attenuation or reversal of the advantage of mirror-image over rise-fall mapping, then it could be argued that this cue contributes to, or underlies, the advantage of mirror-image mapping when the cue is present.

Method

The experimental procedure was basically the same as that of Experiment 1, but only the double-tone stimuli and two stimulus arrangements, mirror-image and rise-fall, were used here. Subjects were trained with one of three stimulus sets in which the steady-state frequencies, transition endpoints, or average overall frequency were or were not the same for all four stimuli in a given set.

Stimuli. The constant-steady-state set was the same as the double-tone set used in Experiment 1. Likewise the "ba" and "ab" stimuli were identical across the three stimulus sets. Only frequency values for the "da" and "ad" stimuli were modified in the remaining stimulus sets. In the constant-transition set, transition endpoints for each stimulus pattern were 995 and 1,230 Hz for the lower tone and 2,180 and 2,525 Hz for the higher tone. Thus, steady-state frequencies were 995 and 2,180 Hz for the "da" and "ad" stimuli and 1,230 and 2,525 Hz for the "ba" and "ab" stimuli. In the constant-average-frequency set, the steady-state frequencies for the two "d" stimuli were 1,183 and 2,456 Hz for the lower and upper tones, respectively; in the transition portion, the frequency increased or decreased to endpoints of 1,418 and 2,801 Hz.[1] Subjectively, the effect of equalizing average frequency across stimuli was to eliminate obvious differences in overall pitch among stimulus patterns.

Subjects. Sixty undergraduates who had not participated in Experiment 1 were recruited from the introductory psychology subject pool. Ten subjects were assigned randomly to each of six experimental conditions formed by factorially combining the two categorization rules and the three stimulus sets.

Results and Discussion

For both the constant-steady-state and constant-transition stimulus sets, the mirror-image categorization condition showed more accurate performance than the rise-fall condition: 82.0% correct for the mirror-image condition vs. 68.0% for the rise-fall condition in the constant-steady-state set and 96.8% for mirror image vs. 64.9% for rise-fall in the constant transition set. However, when average frequency was held constant for all stimuli in a set, effectively reducing and neutralizing all salient pitch cues, the mirror-image mapping condition (68.3%) was not significantly different from, and was in fact slightly poorer than, the rise-fall mapping condition (73.2%). An analysis of variance on total correct responses established the statistical significance of the interaction between stimulus set and mapping condition [$F(2,54) = 13.45$, $p < .001$].

With regard to performance on individual test stimuli, the subjects again showed higher response accuracy to transition-final stimuli than to transition-initial stimuli, but only for the constant-steady-state and constant-average-frequency stimulus sets. For the constant-steady-state set pooled over categorization conditions, response accuracy was 59.46% for

transition-initial stimuli and 90.50% for transition final stimuli. For the constant-frequency set, response accuracy was 62.12% for transition-initial stimuli vs. 79.38% for the transition-final stimuli. When the frequency of the transition endpoints was held the same for all four stimuli in the set (i.e., the constant-transition stimuli), response accuracy for transition-initial signals (82.46%) did not differ reliably from transition-final signals (79.25%).

These results demonstrate the perceptual salience of the overall frequency (i.e., pitch) of the entire stimulus as a perceptual cue in learning to categorize acoustic patterns according to a mirror-image relationship. The present findings indicate that listeners did not respond simply to the individual components of these patterns, such as the direction or extent of the transition or steady-state segments, but instead attended to a configural property of the entire stimulus—its overall frequency during perceptual learning of the categorization rules. Average frequency, as defined with these stimulus patterns, is correlated with the overall perceived pitch of the patterns and subjects were using this property to control their responses.

EXPERIMENT 3
LABELING TASK

The first two experiments used a perceptual learning task to study the acquisition of mirror-image acoustic patterns resembling speech. In Experiment 3, we used a labeling or identification task to examine another related question concerning how these patterns are perceived: To what extent could these nonspeech signals, which were patterned after certain features of speech sounds, also be heard as speech, or alternatively, as frequency-varying tones? These stimulus patterns are speech-like in the sense that they share certain important properties in common with speech, but not in the sense that they would be mistaken for actual speech tokens. Our interest was in determining whether listeners could process the signals as if they were speech and attach appropriate phonetic labels to them. In addition, we also wanted to know if listeners could process the signals simply as auditory tones and identify the direction of the frequency change within each signal.

Method

Subjects were required to identify acoustic patterns with either acoustic or phonetic labels. No feedback was provided in this experiment, since we wanted to measure subjects' ability to categorize these acoustic patterns solely on the basis of the acoustic or phonetic attributes implicit in these signals. Thus, we were not interested in the subjects' ability to learn an arbitrary sound-to-label association in the context of a particular test situation as in the previous experiments.

Stimuli and Procedure. The stimulus patterns were the single-, double-, and triple-tone signals used earlier in Experiment 1. Two conditions were examined. In the "phonetic" condition, subjects were told that the stimuli were distorted tokens of natural speech. The response labels provided to subjects were the syllables "ba," "da," "ab," and "ad," which were placed under four separate buttons on the response panel. In the "acoustic" condition, the subjects were told that the stimuli were frequency-modulated tones generated by a computer and that they consisted of a short interval with constant pitch, preceded or followed by a very rapid rise or fall in pitch. The response labels were schematic line drawings of the time course of the frequency change of each stimulus: (/——), (——/), (\——), (——\).

A factorial design was used again; the between-subject factors were stimulus set (single, double, or triple tones) and label condition (acoustic or phonetic). Ten naive subjects were recruited from the same source as in Experiments 1 and 2 and were assigned at random to the six experimental conditions. Sixty repetitions of each of the four stimuli were presented in random order to the subjects, for a total of 240 labeling responses per subject.

Results and Discussion

Responses were scored as correct or incorrect depending on whether the indicated label was the most appropriate cue for the presented stimulus. Percent correct performance for both labeling conditions across the three stimulus sets is displayed in Figure 4.

For the single- and double-tone stimuli, the subjects were able to use acoustic labels more accurately than phonetic labels (single tones: 49.8% vs. 36.6% correct for acoustic and phonetic labels, respectively; double tones: 61.0% vs. 42.2%). However, with triple tones, which contained energy in the first formant region, listeners assigned phonetic labels more accurately than acoustic labels (62.7% correct for phonetic labels compared with 42.6% for acoustic labels). Subjects in the phonetic labeling condition were able to hear these triple-tone patterns as speech, whereas subjects in the acoustic labeling condition had difficulty in focusing their attention on the individual components of the patterns. The decrement in performance for the acoustic labeling group can

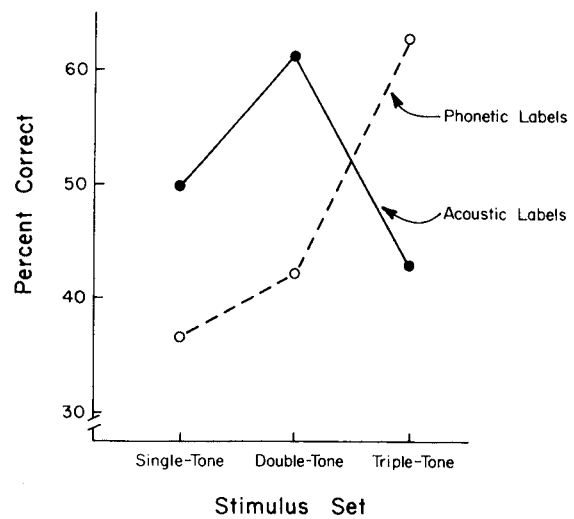

Figure 4. Percent correct identification in the labeling task for acoustic and phonetic labels. The data are shown separately for single-, double-, and triple-tone stimuli.

be accounted for by the presence of conflicting information in the F1 transition for half the stimuli. In the triple-tone patterns, the F1 always rises in initial position and falls in final position. For stimuli containing rising transitions in initial position (i.e., /ba/) or falling transitions in final position (i.e., /ab/), the information in F1 is redundant with regard to the direction of the movements of F2 and F3, thus facilitating performance. However, for stimuli containing falling transitions in initial position (i.e., /da/) and rising transitions in final position (i.e., /ad/), the F1 component is in conflict with the direction of the transitions in the remainder of the pattern, thus contributing to a decrement in performance. The interaction between response label type and stimulus set was statistically reliable at the .01 level [$F(2,54) = 6.78$]. Thus, listeners could attend to either the phonetic or acoustic properties of these signals with better-than-chance accuracy, but their overall accuracy in doing so varied with the complexity of the signal and the specific stimulus properties attended to under phonetic or acoustic labeling instructions.

Examination of the labeling performance for the four separate stimuli indicated that, for acoustic labels, response accuracy was greater with transition-final signals ("ab," 69.44%; "ad," 62.83%) than transition-initial signals ("ba," 38.94%; "da," 33.39%). This outcome was also observed in the previous perceptual learning experiments. Interestingly, however, when listeners assigned phonetic labels to these same signals, the differences between transition-initial and transition-final signals were reduced substantially to nonsignificant levels ("ba," 45.17%; "da," 43.00%; "ab," 51.28%; "ad," 49.17%). These data are presented in Figure 5, where they have been pooled across single, double, and triple tones, since the observed overall pattern of responses was essentially the same for each stimulus set. These results demonstrate a clear dissociation between auditory and phonetic categorization of the same acoustic signals. Moreover, the present findings suggest that the subjects in our earlier experiments did not perceive the patterns as speech signals upon which to make their judgments, but rather responded to them as complex auditory patterns, since the same asymmetrical pattern of responding was observed across initial and final positions.

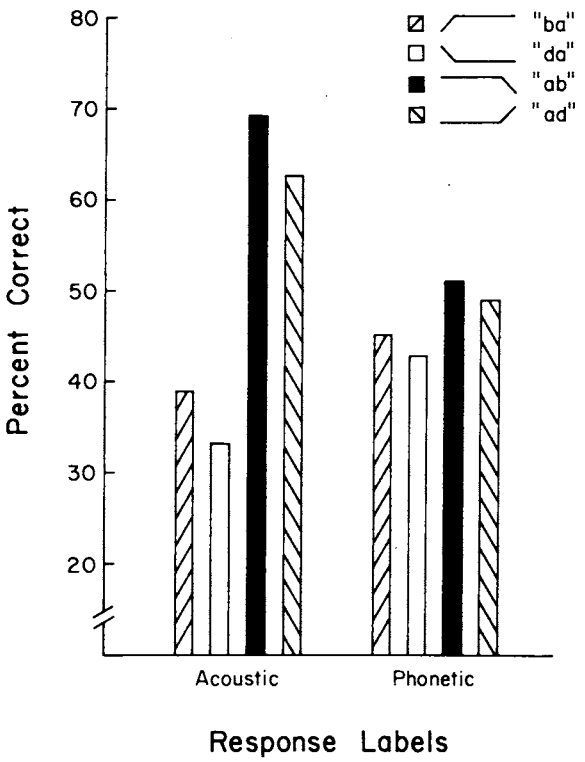

Figure 5. Percent correct identification for individual stimuli in the labeling task shown separately for acoustic and phonetic instruction conditions.

EXPERIMENT 4
SIMILARITY JUDGMENTS

Klatt and Shattuck (1975) and Shattuck and Klatt (1976) found that two rising or two falling frequency glissandos (FMs) were judged to be more similar than mirror-image glissandos. Their stimuli differed from those used in the present studies, however, in that no-steady-state (CF) portion was included. That is, the stimuli consisted of only the FM or transitional portion of the CV syllable. In Experiment 4, we repeated Klatt and Shattuck's similarity judgment task, using our nonspeech acoustic patterns. The principle question of interest here was whether mirror-image stimuli would be perceived to be more similar, as the results of the earlier perceptual learning tasks had suggested, or whether the rise-fall pairs would be judged more similar, as found by Klatt and Shattuck with their glissando-only signals.

Method

Procedure. On each trial, subjects were presented with three acoustic patterns separated by 250 msec of silence. The subjects were instructed to indicate by pressing one of three buttons on the response panel, which stimulus sounded most different from the other two. Each subject heard tones from both the constant-steady-state and constant-average-frequency sets used in Experiment 2. Trials with the two stimulus sets were blocked so that the first three blocks for a given subject contained tones from one stimulus set, while the final three blocks contained tones from the other stimulus set. The order of the stimulus sets was counterbalanced over subjects. Each of 24 possible stimulus triads was presented once in a given block of trials.

In addition to experimental trials in which each of the three presented sounds were different (i.e., dissimilarity trials), we also included identity or "catch" trials in which two of the three stimuli were, in fact, physically the same. The identity trials were included in order to test for differences in the discriminability of the individual stimuli and also to check that subjects were, in fact, attempting to pick out the one stimulus from the triad that was most different from the other two. The odd stimulus on identity

trials occurred equally often in each temporal position in a block of trials. All 36 possible triad configurations for identity trials were used for each stimulus set in an experimental session. Thus, each block of test trials consisted of 24 dissimilarity trials randomly interspersed with 12 identity trials.

Subjects. Eighteen additional subjects were recruited from a paid subject pool at Indiana University to participate in this experiment. None of the subjects had participated in any of the earlier experiments.

Results and Discussion

Only the results of the first half of each session, that is, the initial three blocks of trials for each subject, will be reported here. Subjects' responses on the second half suggested that prior experience with one stimulus set substantially influenced their similarity judgments on the other stimulus set. The task was difficult for subjects, as indicated by the finding that, on 16% of the dissimilarity trials and on 10% of the identity trials, subjects failed to respond at all during the allotted 2.5-sec response interval. The results for the dissimilarity trials are therefore based only on the trials in which subjects provided a response. For the identity trials, however, percent correct was based on all trials unless otherwise noted.

Dissimilarity trials. The main objective of this experiment was to determine which stimulus pattern would be selected as most different from the two other stimulus patterns presented on each trial. Would subjects choose the stimulus that was not the mirror-image of either of the other two tones, the stimulus which had a rising or falling transition in the direction opposite to the transition of the other tones, or the stimulus that had a transition in a temporal position different from those of the other tones? Several factors appeared to influence the subjects' judgments in this task, including the particular stimuli presented on a trial, the order of the stimuli on a trial, and, more importantly, the stimulus set used, whether it was the constant-steady-state set or the constant-average-frequency set.

Figure 6 shows the percent responses for each response type (i.e., mirror-image, rise-fall, or temporal position) pooled over subjects and trials for each stimulus set.

For the constant-steady-state set, the largest proportion of responses, 39%, were in the direction of mirror-image similarity (i.e., the two nonselected tones were, in fact, mirror images of one another). Thirty-five percent of the responses were based on temporal position, and 27%, on rise-fall. An examination of individual subject performance, as contrasted with the pooled group data, indicated even more strongly the predominance of mirror-image responses in the data. Of the nine subjects run in this experimental condition, eight gave mirror-image responses most frequently and only one gave temporal position responses most frequently.

For the constant-average-frequency set, in which overall average frequency (i.e., pitch) was experimentally neutralized as a property that varied systematically across stimuli, the pattern of results changed in the same manner as the results obtained in Experiment 2, using the perceptual learning task. Rise-fall judgments were now slightly more prevalent (37%) than mirror-image responses (35%), which in turn were more prevalent than temporal position responses (28%). For the nine subjects who listened to the constant-average-frequency set, rise-fall was the predominant response for four of the subjects, mirror image was for four other subjects, and temporal position was for only one subject.

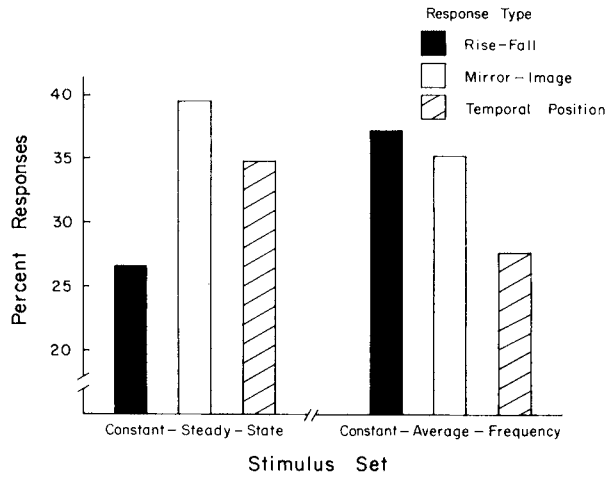

Figure 6. Distribution of responses obtained in the similarity judgment task for the two stimulus sets used in Experiment 4.

Identity trials. Responses on identity trials were scored as correct or incorrect depending on whether the selected stimulus was, in fact, different from the other two presented on the trial. Overall percent correct on identity trials was 77.01%. Excluding omission responses, percent correct responses, given that some response was made, was 85.45%.

The identity trials were sorted into three categories according to whether the different, or "odd-out," stimulus: (1) was the mirror image of the two identical stimuli, (2) had transitions in the same direction as the other two stimuli, or (3) had transitions in the same temporal position as the other stimuli in a triad. Based on responses from the dissimilarity trials, several predictions can be made about the distribution of responses on identity trials. For example, with the constant-steady-state stimuli, it would be expected that identity trials requiring a discrimination between mirror-image stimuli would be more difficult than trials requiring a discrimination based on a temporal position or rise-fall discrimination. This outcome was anticipated because mirror-image stimuli were judged to be the least dissimilar on the dissimilarity trials examined above.

The observed pattern on identity trials for both stimulus sets was entirely consistent with the data from dissimilarity trials: With the constant-steady-

state stimuli, mirror-image trials showed poorer discrimination accuracy (76.85%) than either the rise-fall trials (82.41%) or the temporal position trials (81.48%). With the constant-average-frequency stimuli, accuracy was slightly poorer with rise-fall trials (70.37%) than with mirror-image trials (72.22%); it was least poor with temporal position trials (78.70%).

To summarize the results from Experiment 4, responses on identity trials and dissimilarity trials converged to support the same conclusions: With pure-tone patterns that contain both a frequency-transition (FM) and a steady-state (CF) segment, mirror-image pairs are judged to be more similar and are more difficult to discriminate from one another than rise-fall pairs. This is true providing perceived pitch is also available as a discriminative cue. That is, when the average overall frequency of the components is selectively neutralized so that perceived pitch cannot be used as a distinctive cue, rise-fall pairs tend to be judged as slightly more similar than mirror-image pairs, a result that is entirely consistent with findings from previous perceptual learning experiments that used the same stimulus patterns. Thus, the results from both experimental paradigms converge to produce the same conclusion, namely, that mirror-image acoustic patterns share properties that define their similarity at a more abstract level of analysis.

SUMMARY AND CONCLUSIONS

Mirror-image acoustic patterns of the kind used in these experiments show an advantage in perceptual learning; subjects respond not only to the individual components of these patterns, but also to properties of the entire pattern in terms of its configural shape and perceived internal organization. Subjects do not seem to attend selectively to only the gross shape of the spectrum at onset or offset; instead, they prefer to rely on salient acoustic cues contained in both the transitional (FM) and the steady-state (CF) portions of the overall patterns. Subjects apparently can use these dimensions effectively in learning to assign nonspeech stimuli to arbitrary response categories in a perceptual learning task.

In the case of mirror-image patterns, the criterial differences between the response categories also happen to be correlated with a salient and well-defined redundant property of the patterns, that is, their overall pitch. This property of the patterns was an irrelevant and uncorrelated dimension when the stimuli were arranged in the rise-fall condition. The present findings parallel those of Divenyi and Hirsh (1978), who studied the identification of complex auditory patterns. Their results, using experienced listeners, showed that subjects perceive three-tone melodic sequences as a single pattern rather than as a succession of separate auditory events displaced in time. The subjects in the present experiments were inexperienced listeners with these sound patterns, but, nevertheless, they responded to the stimulus patterns in holistic fashion, as shown by their reliance on the average frequency or perceived pitch of the whole signal.

It is also apparent from our results with these nonspeech signals, which have properties similar to those found in speech, that differences in "mode of processing" can also control perceptual selectivity quite substantially and can subsequently influence the perception of individual components of the stimulus pattern as well as the entire pattern itself (see, e.g., Schwab, 1981). This can occur in quite different ways with the same stimulus patterns, depending primarily on whether the subject's attention in the task is directed toward coding either the auditory properties of the signals or the phonetic qualities of the overall patterns. In the former case, the process is more analytic, involving the processing or "hearing out" of the individual components of the stimulus, whereas in the latter case, the process is more nearly holistic, insofar as the individual components may be combined to form a well-defined and highly familiar phonological category. With regard to the perception of speech, our results imply that listeners probably do not isolate and then subsequently process only the distinctive speech cues in the stimulus as suggested a number of years ago by Mattingly (1972). Rather, it seems very likely that listeners respond to these so-called "speech cues" as simply part of the overall context or the configuration of a spectrally complex time-varying auditory pattern. Indeed, the recent findings reported by Remez, Rubin, Pisoni, and Carrell (1981) suggest that listeners are able to perceive the linguistic message in sinusoidal patterns that do not contain the traditional acoustic cues held to underlie segmental phonetic perception.

In summary, the combined results of our experiments indicate that acoustic patterns similar to pre- and postvocalic variants of a particular stop consonant in CV and VC syllables have more salient correlated properties in common with each other than with similar acoustic patterns that have transitional movements in the same direction. Our results on the perception of nonspeech acoustic patterns, obtained in a perceptual learning task, contrast with earlier reports, by Klatt and Shattuck, which suggest that the perceptual similarity of mirror-image acoustic patterns cannot be recognized by adult subjects. Further research on this problem is currently planned in our laboratory with infants, young children, and nonhuman primates to determine developmental trends and to delimit cross-specific differences in processing complex patterns that resemble speech signals.

REFERENCE NOTE

1. Kewley-Port, D. *A complex-tone generating program* (Research on Speech Perception Progress Report No. 3). Bloomington: Department of Psychology, Indiana University, 1976.

REFERENCES

ADES, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, **16**, 61-66.

DIVENYI, P. L., & HIRSH, I. J. Some figural properties of auditory patterns. *Journal of the Acoustical Society of America*, 1978, **64**, 1369-1385.

KLATT, D. H., & SHATTUCK, S. R. Perception of brief stimuli that resemble rapid formant transitions. In G. Fant & M.A.A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975.

MATTINGLY, I. G. Speech cues and sign stimuli. *American Scientist*, 1972, **60**, 327-337.

PISONI, D. B., & TASH, J. Auditory property detectors and processing place features in stop consonants. *Perception & Psychophysics*, 1975, **18**, 401-408.

REMEZ, R. E., RUBIN, P. E., PISONI, D. B., & CARRELL, T. D. Speech perception without traditional speech cues. *Science*, 1981, **212**, 947-950.

SCHWAB, E. C. *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral thesis, State University of New York at Buffalo, August 1981.

SHATTUCK, S. R., & KLATT, D. H. The perceptual similarity of mirror-image acoustic patterns in speech. *Perception & Psychophysics*, 1976, **20**, 470-474.

WOLF, C. G. Perceptual invariance for stop consonants in different positions. *Perception & Psychophysics*, 1978, **24**, 315-326.

NOTE

1. In computing the average frequencies of each composite stimulus, we included the contributions of both the steady-state and transition portions of each pattern, which were weighted in proportion to their temporal duration. The frequency of the transition portion was defined as the frequency at the midpoint of the transitional segment, since the transitions contained linear changes in frequency from onset or offset to the steady-state values. The mean frequency for each sinusoidal component was computed according to the following equation:

$$F_M = [D_{SS}/(D_{SS} + D_T)]F_{SS} + [D_T/(D_{SS} + D_T)][(F_{SS} + F_{EP})/2],$$

where D and F are the duration and frequency of the steady-state constant frequency segment. D is the duration of the transition, and F is the frequency of the transition endpoint, either the initial or final frequency of the sinusoid. The overall mean computed frequency was defined as the simple average of the F of the separate sinusoids.

(Manuscript received August 5, 1980;
revision accepted for publication November 1, 1981.)