PERCEPTUAL LEARNING OF MIRROR-IMAGE ACOUSTIC PATTERNS[1]

M.E. Grunke and D.B. Pisoni[2], Psychology Department, Indiana University, Bloomington, Indiana 47401 U.S.A.

The phonetic units of spoken language are often mapped in a one-to-many fashion onto their acoustic representations in speech. As a consequence, a child acquiring language must in some way be able to recognize a variety of different acoustic signals as members of the same phonetic category. For example, the /b/ in the syllable /ba/ is characterized by rapidly rising formant transitions into the following vowel, whereas the same consonant in the syllable /ab/ is characterized by rapidly falling formant transitions.

In view of the lack of one-to-one correspondence between phonemes and their acoustical representations in speech, it would seem advantageous for a child learning language if the various acoustical forms of a particular phoneme were related to each other perceptually. In fact, the acoustical representations of stop consonants in initial and final position, although physically different, are related: stop consonants in final syllable position are roughly the mirror image in time of their counterparts in initial position. But are mirror-image acoustic patterns inherently related perceptually for the listener?

The issue of the perceptual relatedness of mirror image acoustic patterns was addressed recently in a series of experiments by Klatt & Shattuck (1975) & Shattuck and Klatt (1976). They presented brief pure-tone acoustic patterns to adult listeners who had to make a similarity judgment. The acoustic patterns were two-component frequency glissandos with a short-term spectral composition similar to the formant transitions in speech.

The results of these experiments did not support the original hypothesis that mirror-image acoustic patterns are intrinsically similar for a listener. Instead, judgements of perceptual similarity for these patterns were based primarily on

---

1) This research was supported by NIMH research grant MH-24027-04, NIMH Post-doctoral fellowship MH-5823-03 to MEG and a fellowship from the Guggenheim Foundation to DBP.

2) Currently at the Speech Group, Research Laboratory of Electronics, M.I.T. Cambridge, Mass.

the direction of the lower glissando component, the component occurring in the region of the second formant.

We have conducted three experiments that also address the question of whether mirror-image acoustic patterns are intrinsically related. However, we used acoustic patterns that included a steady-state constant frequency (CF) portion as well as a rapid frequency glissando (FM). In addition, we used a perceptual learning paradigm in which listeners had to learn to map four different acoustic patterns into two response categories. We wanted to know which of several mapping arrangements of these patterns would be easiest for listeners to learn.

Stimuli

Three sets of stimuli, with four signals per set, were generated using a complex-tone generating program (Kewley-Port, 1976). Each stimulus component consisted of a 60 msec linear rise or fall (FM) in frequency and a 140 msec constant-frequency (CF) portion. The four signals within a set differed in whether the frequency transition was rising or falling and whether the transition preceded or followed the steady-state portion.

The three stimulus sets differed in the number of component tones, either one, two or three. Frequency values were selected to correspond to values of the first, second, and third formants in the syllables /ba/, /da/, /ab/, and /ad/. For the Single-Tone set, the patterns corresponded to the frequency of the second formant. For the Double-Tone set, component frequencies corresponded to the second and third formants. For the Triple-Tone set, all three formants were represented although the frequency transitions 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 natural speech.

Experimental Procedure and Design

In the perceptual learning task one stimulus from a particular set was presented via headphones on each trial to subjects who responded by pressing one of two response buttons. Correct feedback was provided after each response according to one of three stimulus mapping arrangements. In 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 (R1) whereas stimuli with a falling transition preceding the steady-state ( ) or a rising transition following the steady-state ( ) were assigned to the other response (R2). In the Rise-Fall condition, stimuli with rising transitions either preceding or following the steady-state ( ) were assigned to one response: the two stimuli with falling transitions ( ) were assigned to the other. In the Temporal-Position condition, the stimuli were assigned to responses according to the temporal position of the transitions -- whether the transition preceded ( ) or followed ( ) the steady-state frequency.

In addition to test trials on which responses were collected, study periods were also interspersed to help subjects learn the appropriate stimulus-response mapping. During study periods, several repetitions of each of the test stimuli were presented while feedback was provided.

Results and Discussion

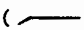
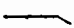
Responses were analyzed in terms of number correct by stimulus. Of the three mapping conditions, the Temporal-Position condition showed the highest performance with 88.4% correct. More importantly, however, the Mirror-Image condition produced more accurate responding, 78% correct, than the Rise-Fall mapping condition with 68.3% correct response.

With respect to stimulus set, the Double-Tone stimuli showed slightly better performance, 81.6%, than the Single-Tone set, with 78.7%. However, the Triple-Tone set which more closely resembled speech showed the poorest performance with only 74.5% correct.

These results have several implications for learning of acoustical patterns resembling speech. First, the easiest stimulus mapping condition to learn was the one based on the temporal position of the transition in the pattern -- a relationship that clearly does not require the subject to analyze out the component frequencies at onset or offset. Subjects can simply use temporal position (i.e., initial vs. final) as the most salient dimension for learning and ignore all other differences. Second, it is also apparent from our results that when the position of the transition becomes an irrelevant attribute to be ignored by the subject, a stimulus arrangement based on a mirror-image relationship is, in fact, easier to learn than one based on only the direction of frequency change of the transitions. Thus, while mirror-image patterns are not the same perceptually, subjects are nevertheless able to recognize and selectively attend to the criterial properties of the stimuli that define their equivalence. In both of these mapping conditions, subjects must "hear-out" the individual components of the patterns and respond to them selectively.

In Experiment 1 we used tonal patterns so that the stimuli would not be heard as speech. The results would be uninteresting if subjects simply heard these patterns as speech and used phonetic labels to mediate acquisition. By this interpretation, mirror-image patterns would be superior to direction-of-transition mapping because the stimuli within a mirror-image pair evoke the same phonetic label -- i.e. "b" or "d" -- and not because their configural properties are intrinsically related. To study the effects of categorization on perceptual learning we carried out another experiment to determine how well subjects could identify these patterns when explicitly provided with labels that emphasized attention to either the acoustic or phonetic properties of the stimuli.

<u>Experimental Procedure and Design: Experiment 2</u>

In Experiment 2 subjects were required to identify the stimuli into one of four categories provided by the experimenter. In the Acoustic-Label condition, subjects were told that the stimuli were tones consisting of a short interval with constant pitch, followed or preceded by a rapid rise or fall in pitch. The acoustic labels were schematic line drawings of the time course of the frequency change of each stimulus (,  , , ). In the Phonetic-Label condition, subjects were told that the stimuli were modified tokens of natural speech and were provided with the labels /ba/, /da/, /ab/, and /ad/.

Results: Experiment 2

For the Single- and Double-Tone stimuli, acoustic labels were matched more accurately than phonetic labels. This effect was reversed, however, for the Triple-Tone stimuli, where phonetic labels were more accurate than acoustic labels. The addition of a component in the region of the first formant markedly increased the accuracy of phonetic categorization while decreasing performance with acoustic labels. Note, however, that the low tone for each Triple-Tone stimulus had either an initial rising or final falling transition which did not parallel the direction of the other transitions. The presence of these conflicting frequency glissandos no doubt produced interference in the acoustic labeling condition.

In Experiment 1 acquisition performance was poorer for Triple-Tone stimuli than for Single- and Double-Tone stimuli, a pattern which was replicated here only in the Acoustic Label conditions. Thus, these results support the interpretation that subjects in Experiment 1 were listening primarily in an auditory, rather than phonetic mode and strongly suggest that phonetic-mediation was not responsible for the outcome observed earlier.

The mirror image patterns used here always shared three properties in common: (1) they had the same steady-state frequencies, (2) the frequency transitions at onset and offset had the same short-term spectral composition, and (3) members of a given mirror-image pair had roughly the same average frequency. These three properties are potential factors that could contribute to the salience of mirror-image pairs and the advantage observed in learning. The first factor could not have played a role in the earlier results since all pairs within a stimulus set, whether mirror-image or not, had identical steady-states. However, the other two properties could have been used as reliable discriminative cues by subjects to facilitate learning.

To determine which of these acoustic attributes, if any, was responsible for the advantage observed in learning mirror-image pairs, we repeated Experiment 2 adding two additional stimulus sets, one with average frequencies and the other with the transitions adjusted to be equal. We also included the original constant steady-state stimuli. Would the mirror-image condition continue to show an advantage in learning under these two new stimulus conditions?

Experimental Procedure and Design: Experiment 3

The experimental procedure was the same as Experiment 1. However, only Double-Tone stimuli were used and the Temporal-Position mapping condition was eliminated. Three stimulus sets were constructed, again with four stimuli per set. Each set contained the initial-rising and final-falling stimuli used earlier. For the stimuli with identical transitions, frequency values for the initial-falling and final-rising stimuli were set lower than in the previous experiment, so that the transitions were identical for all stimuli. These two stimuli were also lowered in frequency for the Average-Frequency set to make all stimuli equal in average frequency.

Results: Experiment 3

The difference in performance between Mirror-Image and Rise-Fall mapping conditions was replicated with the Constant-Steady-State stimuli. The difference between Mirror-Image and Rise-Fall was, however, even greater for the stimuli containing identical transitions. However, for the Average-Frequency stimuli, no significant difference was found between the two mapping conditions. Thus, on the one hand, the advantage of Mirror-Image over Rise-Fall mapping can be made more pronounced by adjusting the frequency transitions to be identical whereas the difference can be attenuated substantially by adjusting all stimuli to have the same average frequency irrespective of the temporal and spectral properties of the patterns.

Conclusions

Mirror-Image acoustic patterns show an advantage in perceptual learning because subjects respond not only to individual components of these patterns but also to properties of the entire pattern in terms of its configural shape. Subjects do not seem to attend selectively to only the gross shape of the spectrum at onset or offset but prefer instead to integrate and deploy salient cues contained in both the transitional and

steady-state portion of the entire patterns. In the case of Mirror-Image patterns, criterial differences between the responses happen also to be correlated with salient and well-defined redundant properties of the patterns such as average pitch which was an irrelevant and uncorrelated dimension when the patterns were arranged in the Rise-Fall mapping condition.

It is apparent from these results with nonspeech signals having properties similar to speech that differences in "mode of processing" can also control perceptual selectivity and influence the perception of individual components of the stimulus pattern as well as the entire pattern itself. This can occur in quite different ways depending on whether the subject's attention is directed to coding the auditory properties or the phonetic qualities of the patterns. These new results on mirror-image patterns have been obtained in a perceptual learning task despite the report that the perceptual similarity of these acoustic patterns cannot be recognized consciously by subjects as shown by earlier experiments.

References

Klatt, D.H. and S.R. Shattuck (1975): "Perception of brief stimuli that resemble rapid formant transitions," In G. Fant and M.A.A. Tatham (eds.): Auditory Analysis and Perception of Speech. New York: Academic Press, 294-301.

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

Shattuck, S.R. and D. Klatt (1976): "The perceptual similarity of mirror-image acoustic patterns in speech," Perception and Psychophysics, 20, 470-474.