XIV. SPEECH COMMUNICATION

Academic and Research Staff

Prof. Kenneth N. Stevens
Prof. Morris Halle
Dr. Sheila E. Blumstein*
Dr. Margaret Bullowa
Dr. William E. Cooper
Dr. William L. Henke
Dr. A. W. F. Huggins
Dr. Dennis H. Klatt
Dr. Martha Laferriere†
Dr. Lise Menn
Dr. Paula Menyuk‡
Dr. Colin Painter**
Dr. Joseph S. Perkell
Dr. David B. Pisoni††
Dr. Ernest A. Scatton‡‡
Dr. Stefanie Shattuck-Hufnagel
Dr. Katherine Williams
Dr. Catherine D. Wolf
Barbara J. Moslin

Graduate Students

Marcia A. Bush
Bertrand Delgutte
Gregory M. Doughty
William F. Ganong III
Ursula Goldstein
Stephen K. Holford
Arafim Ordubadi
Victor W. Zue

A. IS THE DETECTION OF PERIODICITY IN ITERATED NOISE CENTRAL?

National Institutes of Health (Grant 5 RO1 NS04332-13)

A. W. F. Huggins

1. Introduction

Let us review two earlier experiments in which sounds were alternated between the ears.

The first experiment involved the intelligibility of alternated speech. In a typical experiment by Cherry and Taylor,[1] and later by Huggins,[2] a continuous speech message was periodically switched alternately to the subject's left and right ears, so that one ear received speech, while the other received silence. The intelligibility was assessed by counting the number of words the subject was able to repeat concurrently — a task called "shadowing." The task was quite easy, and shadowing scores were close to 100%, as long as the rate of alternation was either lower than approximately 1 cps, or when the rate was higher than approximately 10 cps. At intermediate rates around 3-5 cps the task was much harder, and shadowing scores dropped to a minimum of 60-70%, although some subjects were much more severely affected.

What properties of the situation caused the intelligibility to decline? Clearly, the

---

*Assistant Professor, Department of Linguistics, Brown University.

†Assistant Professor of Linguistics, Southeastern Massachusetts University.

‡Professor of Special Education, Boston University.

**Associate Professor of Communicative Disorders, Emerson College.

††Associate Professor, on leave from Department of Psychology, Indiana University.

‡‡Assistant Professor, Department of Slavic Languages and Literatures, University of Virginia.

subject was simply not able to add the waveforms presented to his two ears, and attend to the sum; this would have regenerated the original waveform, and no intelligibility minimum would have occurred. We therefore conclude that, for disparate, nonfusible signals at least, the most elementary stage of auditory processing is performed separately for the two ears. The least that this processing can involve is the neural coding of the chunks of speech waveform together with the intervening silences but, as we shall see, it may involve more.

This argument suggests that alternated speech shows reduced intelligibility at the critical alternation rates because it comprises interrupted speech in each of the two ears. Several experimental results support this interpretation.[3] Briefly, these experiments show that the decline of intelligibility, as the rate of alternation is raised from 1 cps, is due to the progressive shortening of the chunks of speech. The shorter the chunk, the lower the intelligibility. Silent intervals longer than ~150 ms effectively force the perceptual apparatus to treat each speech chunk independently; that is, the duration of the silent intervals has little effect, as long as they are longer than ~150-200 ms.

On the other hand, the recovery of intelligibility, as the alternation rate is further increased, is determined by the duration of the silent intervals. As silent intervals are shortened from 150 ms to 60 ms, the perceptual apparatus becomes able to "bridge the gaps," and intelligibility rises again, even though the speech intervals themselves become progressively less intelligible as they are shortened. Cherry and Taylor's explanation[1] in terms of a breakdown of the ability to switch attention fast enough between the ears was rejected on other grounds by Huggins.[2] Rather, the experiment seems to measure the growth of coherence of the signals in each ear[4] and, concurrently, the resulting stream segregation of the signals in the two ears, as silent intervals are shortened.

The second experiment that we shall review involved alternated clicks.[5] Subjects adjusted the rate of a binaural pulse train until it matched the perceived rate of a dichotically alternated train in which successive pulses went alternately to the left and right ears.

The results were that at slower alternation rates subjects matched the total rate of pulses-into-the-head of the alternated train. At faster rates, they matched the rate in-one-ear of the alternated train.

Phenomenologically, the slower alternated trains sound like a sequence of discrete clicks which just happen to go alternately to the left and right ears. At faster rates, a separate pulse train is heard in each ear, and it is not possible to integrate the two trains into a single train of twice the frequency. Again, we seem to be measuring the growth of perceptual streaming, this time with dichotic clicks. At the faster rates it is as if each pulse "captures" the succeeding pulse in the same ear, and this precludes its

interacting with the intervening pulse in the other ear, although it occurs after half as long a delay. Thus it again appears that the early stages of auditory perceptual processing may be separate for the two ears, at least for nonfusible signals.

In the experiment to be described now we tried to test this hypothesis directly. The phenomenon selected for the experiment was the detection of periodicity in iterated noise segments.[6] The detection of repetitions, in a stimulus as lacking in structure as white noise, implies a memory that will hold a rather primitive representation of the signal; that is, it operates early in the sequence of processing. In this experiment we tried to determine whether periodicity detection is performed separately on the inputs to the two ears, or whether the periodicity detector has access to both signals.

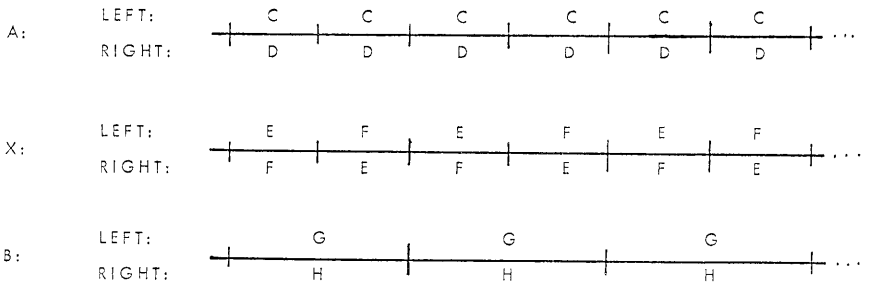

Fig. XIV-1. Format of a single test triplet.

The experimental question to be answered is the following. If two different noise samples, E and F, each lasting 100 ms, are alternated between the ears, as diagrammed in Fig. XIV-1, is the perceived period equal to the repeated chunk in each ear, E followed by F (i.e., 200 ms)? Or is the period equal to the chunk that repeats if the signals in the two ears are mixed, E plus F (i.e., 100 ms)?

2. Procedure

Preliminary tests showed that when the iterated segments were as long as 0.5 s, it took some time to detect the period of iterated noise. Therefore each alternated test stimulus was presented for 5 s. Iterated noise is perceptually quite rich, so in order to concentrate the subject's attention on its periodicity, each test stimulus was preceded by one comparison stimulus and followed by another comparison stimulus. Thus each test stimulus was presented as the middle item of a triplet, in an AXB format, as shown in Fig. XIV-1.

Since a different noise sample was presented to each ear in the alternated test stimuli, each of the comparison stimuli also presented independent noises to the two ears. Thus in the comparison stimuli two different noise segments of the same duration were iterated, one in each ear. The iterated segments in one comparison stimulus were of

the same duration as the alternated segments, and in the other comparison stimulus they were of twice the duration of the alternated segments. Thus if periodicity detection is separate for the two ears, the subject should select the longer period comparison as matching the alternated stimulus. We shall refer to this as monaural periodicity detection.

On the other hand, if the periodicity detector notices that the noise in each ear is immediately iterated in the other ear, the subject should select the comparison with the shorter period. We shall call this "central" detection.

Nine durations were selected, logarithmically spaced between 25 ms and 400 ms. A test tape was made up, containing six practice triplets and 27 test triplets. Nine of the test triplets, one for each period, were catch trials, in which the noise samples in the "alternated" test stimulus were not alternated. Subjects should always match these with the shorter period comparison. The number of stimuli presented to the subjects was doubled simply by turning the tape over at the end, and playing the same stimuli backward. This had the advantage of balancing the design: the order of the triplets, as well as the noise, was reversed and which noise sample was presented to which ear, was also reversed.

Twenty-four subjects were paid for serving; all reported having normal hearing. The subjects were run in groups of up to four, in a soundproof room. The stimuli were presented with stereophonic earphones. The subjects responded by marking an A or a B on the answer sheets, according to which comparison signal was heard to have

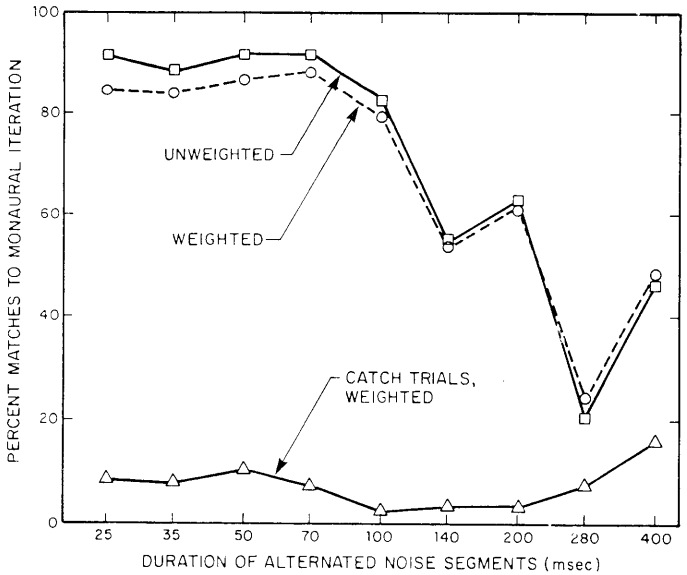

Fig. XIV-2. Proportion of monaural matches as a function of the duration of the alternated noise segments. Pooled results for 24 subjects.

the same period as the test stimulus. They also indicated their degree of confidence as "certain," "think," or "guess."

3. Results

The pooled results are shown in Fig. XIV-2. The duration of the alternated noise segments is on the abscissa, and the ordinate is the percentage of matches to the longer period comparison; that is, of monaural matches. The data are presented both weighted by the degree of confidence, and unweighted. The weights used were 3 for "certain," 2 for "think" and 1 for "guess." The weighted and unweighted results are substantially the same. This was also true of the individual subject's data. All but three or four of the subjects gave results that are very similar to the pooled result.

The results show that the double segment that iterates in each ear overwhelms cross-ear effects in determining periodicity, at least for segments up to 100 ms; that is, up to a period of 200 ms in one ear. D-primes (although not appropriate) are ~3.0, for the four data points at the left of Fig. XIV-2.

At long periods the results become more variable, and the data show two reversals. Consider first the longest period tested, 400 ms. When a segment as long as this is alternated between the ears for 5 s, it is only presented to each ear six times. Thus if the ease with which periodicity can be detected depends on the number of iterations, rather than on the duration of presentation, detection of periodicity in this stimulus may have been substantially harder than the others. Some support for this view comes from the catch trials. Results of the catch trials are indicated in the lower part of Fig. XIV-2 and show that the subjects correctly matched the shorter period comparison. But notice how the function begins to rise for the longest periods. The catch trials became harder, too, at the two longest periods.

The inversion at 140 ms and 200 ms is harder to explain. One possibility is that, by chance, there was an undesired similarity of some sort between the two noise samples that were alternated in the test stimulus. This might lead to a halving of the perceived monaural periodicity. No such similarity was detectable in the test stimuli, at least not from the waveform nor from spectrograms, but it is possible that a more sophisticated analysis such as correlation of the short-term spectra of the alternated noise samples might show something.

Figure XIV-3 shows histograms of the subjects' confidence in their judgments as a function of the period. For each histogram, the matches can be regarded as a seven-point scale, running from a certain match to the short period, at the left, through declining, then increasing confidence to a certain match to the long period, at the right. Matches to the short period are plotted downward, and those to the long period are plotted upward. This figure also supports the idea that the longest period was more difficult:

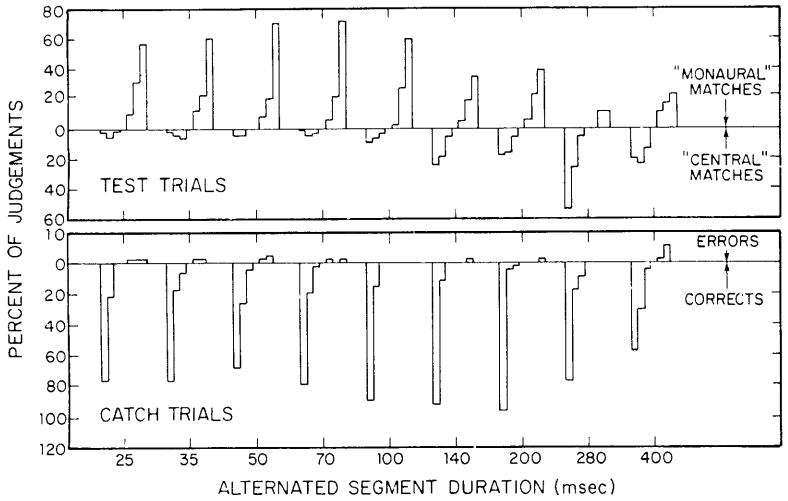

Fig. XIV-3. Histograms showing confidence with which monaural and central matches were made as a function of the duration of the alternated noise segments. Upper: test trials. Lower: catch trials.

the incidence of "guess" confidence was much larger at the longest period.

To circumvent these problems, we have begun an on-line experiment in which the subject himself adjusts the period of the iterated segments in the comparison stimulus. He can switch backward and forward as often as he likes between the alternated standard and the variable comparison, but the noise samples are refreshed every time he does so. That is, the periods are not affected, but the content of the iterated and alternated

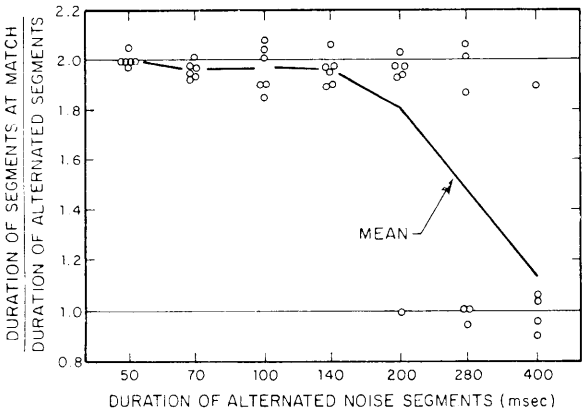

Fig. XIV-4. Results of performance of an adjustment task by a single subject. Each point represents a single match. The ratio of the period of the nonalternating comparison to the period of the alternating standard is plotted against the period of the alternating standard. Monaural matches yield a 2.0 ratio, central matches a 1.0 ratio.

samples changes. Preliminary results for one subject are shown in Fig. XIV-4. Six matches were made for each period between 50 ms and 400 ms. The results are in good agreement with those in the other experiment, and both reversals in the data are gone. Notice also that the matches are bimodal. At the crossover the subject hears both periods, although their relative dominance changes.

In their original description of periodicity in iterated noise, Guttman and Julesz[6] mentioned three different perceptual effects. With periods shorter than approximately 50 ms, the iterated noise had pitch. Between 50 ms and 250 ms, it sounded like "motorboating." Periods longer than 250 ms sounded like "whooshing." The main difference between the whooshing and the motorboating was that in whooshing detail was heard within the iterated segment. The sounds made by an automatic dishwasher have the same property. That is, a pattern of sound is heard to repeat. These results suggest that it is only at the level of this pattern detection that the inputs in the two ears can be compared, when the signals are disparate and nonfusible. This implies that quite extensive processing may be carried out separately and concurrently for the two ears. Earlier results for alternated speech are also compatible with this suggestion.

References

1. E. C. Cherry and W. K. Taylor, "Some Further Experiments upon the Recognition of Speech, with One and with Two Ears," J. Acoust. Soc. Am. 26, 554-559 (1954).
2. A. W. F. Huggins, "Distortion of the Temporal Pattern of Speech: Interruption and Alternation," J. Acoust. Soc. Am. 36, 1055-1064 (1964).
3. A. W. F. Huggins, "Temporally Segmented Speech," Percept. Psychophys. 18, 149-157 (1975).
4. L. van Noorden, "Temporal Coherence in the Perception of Tone Sequences," Thesis from Institute for Perception Research, Eindhoven, The Netherlands, 1975.
5. A. W. F. Huggins, "On Perceptual Integration of Dichotically Alternated Pulse Trains," J. Acoust. Soc. Am. 56, 939-943 (1974).
6. N. Guttman and B. Julesz, "Lower Limits of Auditory Periodicity Detection," J. Acoust. Soc. Am. 35, 610 (L) (1963).

B. IDENTIFICATION AND DISCRIMINATION OF THE RELATIVE ONSET TIME OF TWO-COMPONENT TONES: IMPLICATIONS FOR VOICING PERCEPTION IN STOPS

National Institutes of Health (Grants 1 T32 NS07040-01 and 5 RO1 NS04332-13)

David B. Pisoni

1. Introduction

Within the last few years considerable attention has been devoted to the study of the voicing feature in stop consonants, particularly in terms of the dimension of voice onset time (VOT). The important work of Lisker and Abramson[1, 2] has shown that in a wide diversity of languages the voicing and aspiration differences among stop consonants can be characterized by changes in VOT, which, in turn, reflect differences in the timing of glottal activity relative to supralaryngeal events. According to Lisker and Abramson,[1] it appears that there are three primary modes of voicing in stops: (i) prevoiced stops in which voicing onset precedes the release burst by a value greater than ~20 ms, (ii) short lag voiced stops in which voicing onset is simultaneous or lags briefly behind the release burst, and (iii) long lag voiceless stops in which the voicing onset lags behind the release burst by an amount greater than 20 ms. From acoustic measurements, they found relatively little overlap in the modal values of VOT for the voicing distinctions that occurred in 11 languages that they studied. Moreover, in perceptual experiments with synthetic stimuli they found that subjects identify and discriminate differences in voicing in a categorical manner that reflects the phonological categories of their language.[3, 2] That is, subjects show consistent labeling functions with a sharp crossover point from one phonological category to another and discontinuities in discrimination that are correlated with the changes in the labeling functions. Subjects can discriminate two synthetic stimuli that come from different phonological categories better than two stimuli selected from the same phonological category.[4, 5]

The categorical perception of these synthetic stimuli has been interpreted as evidence for the operation of a special mode of perception, a speech mode, that is unique to the processing of speech signals.[6-8] The argument for the presence of a specialized speech mode is based primarily on three empirical findings. First, nonspeech signals are typically perceived in a continuous mode; discrimination is monotonic with the physical scale. It is well known that subjects can discriminate many more differences than they can label reliably on an absolute basis. Second, until recently, no convincing demonstrations of categorical perception had been obtained with nonspeech signals. Third, it has generally been assumed that the nonmonotonic discrimination functions are entirely the result of labeling processes associated with phonetic categorization. Indeed, the nonspeech control experiments carried out by Liberman et al.[5] and by

Mattingly et al.[9] were designed specifically to determine whether the discontinuities in the speech discrimination functions are due to the acoustic or psychophysical attributes of the signals themselves rather than some speech-related labeling process. Since both of these studies failed to find peaks in the nonspeech discrimination functions at phoneme boundaries, it was concluded that the discrimination functions for the speech stimuli were attributable to phonetic categorization because the stimuli were perceived as speech.

Additional support for the existence of a specialized speech perception mode has come from the results of Eimas and his associates[10] who found that two- and three-month-old infants could discriminate synthetic speech sounds varying in VOT in a manner comparable to that of adults. Infants could discriminate between two speech sounds selected from across an adult phoneme boundary but failed to discriminate two stimuli selected from within an adult phonological category even though the acoustic differences between the pairs of stimuli were constant. The implication of these findings is that infants have access to mechanisms of phonetic categorization at an extremely early age. Furthermore, it has been suggested that these mechanisms are in some way innately determined or develop very rapidly after birth. The important point is that it has been assumed that infants are responding to differences in VOT in a "linguistically relevant" manner which is a consequence of phonetic coding of these signals rather than responses to psychophysical differences prior to phonetic categorization; however, see Stevens and Klatt.[11] If this claim is true, or even partly true, it would provide very strong support for an account of phonological perception based on a set of universal phonetic features that are innately determined. It would also suggest that the environment plays a secondary role in phonological development. Several recent studies, however, have provided some strong evidence for reevaluating this interpretation of the infant data as well as the more general claims associated with a specialized mode of speech perception. These results are based on perceptual experiments with chinchillas,[12] two cross-language experiments with young infants[13, 14] and a study involving more complex nonspeech signals.[15] The common property of these seemingly diverse studies is that they have focused on the voicing distinction in stop consonants, specifically on VOT.

Kuhl and Miller[12] showed that chinchillas could be trained to respond differentially to the consonants /d/ and /t/ in syllables produced by four talkers in three vowel contexts. More important, however, was the finding that the training generalized to a continuum of synthetically produced stimuli varying in VOT. The identification functions for chinchilla were quite similar to human data: the synthetic stimuli were partitioned into two discrete categories with a sharp crossover point. The phoneme boundary for chinchilla occurred at almost precisely the same place as for humans, which suggests a psychophysical rather than a phonetic explanation for the labeling functions. Since chinchillas presumably have no spoken language and consequently have no phonological

coding system, Kuhl and Miller assumed that the labeling behavior in response to synthetic stimuli would be determined exclusively by the acoustic attributes and psychophysical properties of these signals. The results of this study indicate that the boundary between voiced and voiceless labial stops that occurs at approximately +25 ms is probably a "natural" region of high sensitivity along the VOT continuum and, at least in the case of the chinchilla, has little to do with phonetic coding.

Following Eimas' work with infants from English-speaking environments,[10, 16] two cross-language studies were conducted recently using similar methodology and comparable synthetic stimuli differing in VOT. Lasky et al.[13] studied infants 4 to 6 1/2 months old born to Spanish-speaking parents and found evidence for the presence of three categories in their discrimination functions. One boundary corresponded to the voiced-voiceless distinction in the +20-+60 ms region, whereas the other occurred in the pre-voiced region between roughly -20 ms and -60 ms. These results are interesting because Spanish has only one phoneme boundary separating its stops and this boundary does not correspond to either of the two boundaries found in the infant data of Lasky et al. One conclusion to be drawn from these findings is that the environment probably plays only a very minor role in phonological development at this age and that the infants are more likely to be responding to some set of acoustic attributes independently of their phonetic status.

In another related study Streeter[14] found that Kikuyu infants also show evidence of three categories of voicing for labial stops, even though these particular distinctions are not phonemic in the adult language and probably occur very infrequently in the language environment of these infants. The categories and boundaries found in this study were comparable to those of Lasky.

The results of both cross-language investigations of voicing perception are quite similar and indicate that infants can discriminate differences in VOT. Moreover, the pattern of results suggests that infants have the ability to deal with at least three modes of voicing. The basis of these distinctions, however, may be the result of naturally defined regions of high discriminability along the VOT continuum rather than processes of phonetic categorization. Thus the infants may not be responding to these signals linguistically as suggested by the earlier interpretation of Eimas, but instead may be responding to some complex psychophysical relation between the components of the stimulus that occurs at each of these modes of voicing. In anticipation, one such relation is strongly suggested by the results of the present series of nonspeech experiments in terms of changes in sensitivity to differences in temporal order between two components of the stimulus complex. The infants may respond simply to differences between simultaneous and successive events.

In another study, Miller et al.[15] generated a set of nonspeech control signals that were supposed to be analogous to VOT stimuli. The stimuli differed in the duration of

a noise burst preceding a buzz. Identification and discrimination functions were obtained with adults in a manner comparable to those collected in the earlier adult speech perception experiments. For discrimination, the stimuli were presented in an odd-ball paradigm, whereas for labeling the subjects responded with two choices, either "no noise" or "noise" present before the onset of the buzz. The results of this study revealed identification and discrimination functions that were similar to those found with stop consonants differing in VOT. Discrimination was excellent for stimuli selected from between categories and quite poor for stimuli from within a category. The labeling functions were sharp and consistent; the peak in discrimination occurred precisely at the boundary between the two categories.

Miller and his co-workers[15] offered a psychophysical account of these categorical results in terms of the presence of a perceptual threshold at the boundary between two perceptually distinctive categories. According to their findings, in the case of noise-buzz stimuli there is a certain value of noise-lead time below which subjects can no longer detect the presence of the noise preceding a buzz. At values below this duration the stimuli are members of one category and subjects cannot discriminate differences in duration between stimuli because they are below threshold. At noise durations slightly above this value there are marked changes in sensitivity and response bias as a threshold is crossed and a new perceptual quality emerges from the stimulus complex. Miller et al. suggest that discrimination of differences above this threshold value follow Weber's law and, consequently, constant ratios are needed rather than constant differences in order to maintain the same level of discriminability. The boundary between these categories separates distinct sets of perceptual attributes and results in the partitioning of the stimulus continuum into equivalence classes. These equivalence classes for most purposes are categorical: the relation defining membership in a class is symmetrical, reflexive, and transitive.[17]

The account of categorical perception offered by Miller et al. suggests the presence of naturally determined boundaries at specific regions along the VOT continuum. These boundaries occur at places where a new perceptual attribute emerges in the course of continuous variations in one or more parameters of a complex signal. Based on their suggestions, we generated a set of nonspeech signals that differed in the temporal order of the onsets of two-component tones. The stimuli varied over a range from -50 ms where the lower tone leads the higher tone, through simultaneity, to +50 ms where the lower tone lags behind the higher tone. Our goal in producing these stimuli was to have a set of nonspeech control stimuli that differed on a variable known to play an important role in the perception of voicing, the relative timing between two events. A well-known and important cue to voicing in stops is the onset of the first formant relative to the second, the cutback cue.[18] Thus, in using nonspeech signals such as these, we hoped to learn something about how the timing relations in stop consonants are perceived.

Moreover, we hoped that these results would provide the basis for a more general account of the diverse findings obtained with adults, infants, and chinchillas on VOT stimuli, as well as furnish an account of the results obtained in the nonspeech experiments.

2. Experiment I

In this experiment, subjects were trained to identify stimuli selected from a nonspeech auditory continuum by means of a disjunctive conditioning procedure.[19] The results of this study serve as the baseline for our subsequent experiments.

a. Method

Subjects. Eight paid volunteers served as subjects. They were recruited by means of an advertisement in a student newspaper and were paid at a base rate of $2.00 per hour plus whatever they could earn during the course of the experiment. All were right-handed native speakers of English.

Stimuli. The stimuli were 11 two-tone sequences that were generated digitally with a computer program that permits the user to specify the amplitude and frequency of two sinusoids at successive moments in time. Schematic representations of three of the signals are shown in Fig. XIV-5. The lower tone was set at 500 Hz, the higher tone at 1500 Hz. The amplitude of the 1500-Hz tone was 12 dB lower than the 500-Hz tone. The experimental variable under consideration was the onset time of the lower tone relative

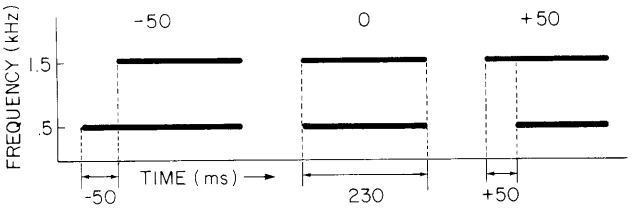

Fig. XIV-5. Representation of three stimuli differing in relative onset time: leading (-50 ms), simultaneous (0 ms), and lagging (+50 ms).

to the higher tone. For the -50 ms stimulus, the lower tone leads the higher by 50 ms, for the 0 ms condition both tones are simultaneous, whereas for the +50 ms condition the lower tone lags the higher tone by 50 ms. All remaining intermediate values, which differed in 10 ms steps from -50 ms through +50 ms , were also generated. Both tones were terminated together. In all cases, the duration of the 1500-Hz tone was held constant at 230 ms and only the duration of the 500-Hz tone was varied to produce these stimuli. The eleven stimuli were recorded on audio tape and later digitized by an A-D converter and stored on disk memory under the control of a small laboratory computer.

I am particularly indebted to Dr. Dennis Klatt for his help with the program used to generate these stimuli.

Procedure. All experimental events involving the presentation of stimuli, collection of responses, and feedback were controlled by a PDP-11 computer. The digitized waveforms of the test signals were reconverted to analog form by a 12-bit D-A converter and presented to subjects binaurally through Telephonics (TDH-39) matched and calibrated headphones. The stimuli were presented at a comfortable listening level of ~80 dB which was maintained consistently throughout all experiments reported here.

The present experiment covered two 1-hour sessions which were conducted on separate days. All subjects were run in small groups. The order of presentation of the test sequences is given in Table XIV-1. On Day 1 subjects received identification training sequences; on Day 2 they were tested for identification and ABX discrimination.

Table XIV-1. Order of presentation of training and test sequences for Experiment I.

Day	Session	Sequence Description	Feedback	No. Trials
1	Training	Initial Shaping Sequence (-50, +50)	Yes	160
1	Training	Identification Training (-50, +50)	Yes	160
1	Training	Identification Training (-50, -30; +30, +50)	Yes	160
2	Training	Warm-up Sequence (-50, +50)	Yes	80
2	Labeling	Identification Sequence (all 11 stimuli)	No	165
2	Discrimination	ABX Discrimination (9 two-step comparisons)	Yes	252

In the initial training sessions, subjects were presented with the end-point stimuli, -50 and +50, in random order and were told to learn which of two buttons was associated with each sound. Immediate feedback for the correct response was provided, although no explicit coding or labeling instructions were given. Subjects were free to adopt their own strategies. After 320 trials, two additional intermediate stimuli (-30 and +30) were included as training stimuli. Immediate feedback was maintained throughout the training conditions.

For identification, subjects were presented with all 11 stimuli in random order and told to respond to this condition as they had to the presentation of end-point stimuli. No

feedback was provided in this condition. In ABX discrimination all 9 two-step pairs along the continuum were arranged in the four ABX permutations and presented to subjects with feedback for the correct response. Subjects were told to determine whether the third sound was most like the first sound or the second sound. Timing and sequencing in the experiment were self-paced to the slowest subject in a given session.

b. Results and Discussion

All eight subjects learned to respond to the end-point stimuli with a probability greater than .90 during the training sessions. The results of the identification and ABX discrimination tests are shown in Fig. XIV-6. The labeling functions are shown by filled

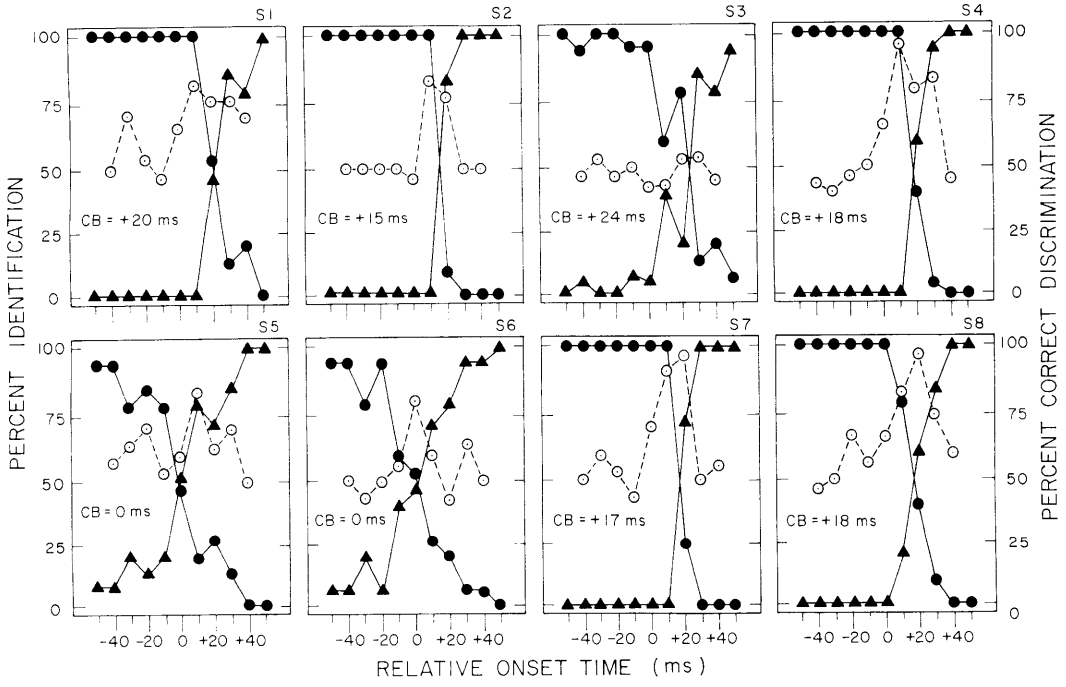

Fig. XIV-6. Labeling functions (left ordinate) and ABX discrimination functions (right ordinate) for 8 individual subjects in Experiment I.

circles and triangles connected by solid lines. For five of the eight subjects (S1, S2, S4, S7, S8), the labeling functions are extremely sharp and consistent and show a very small region of ambiguity between the two response categories. For the remaining three subjects (S3, S5, S6), the labeling functions are less consistent although with additional training these functions would probably have leveled out. For the most part, however, the labeling data for these nonspeech signals are quite good, given the modest number of training trials (560) during the two-day experiment. We should also note that the crossover points for six of the eight subjects do not occur precisely at the

0 onset time value but are displaced toward the category containing lagging stimuli. This asymmetry might be due either to the relatively greater masking of high frequencies by low frequencies or to some limitation on temporal order processing. In order to test the masking interpretation, we ran a pilot study in which the amplitude of the 1500-Hz tone was varied over a 24-dB range from -12 dB through +12 dB relative to the amplitude of the 500-Hz tone. If the asymmetry in the labeling function is caused by masking of the higher tone by the lower tone, we would expect increases in the amplitude of the higher tone to produce systematic shifts in the locus of the category boundary toward progressively shorter onset-time values. No such shift was observed in the pilot experiment, which suggests that the temporal order account is the more likely cause of the asymmetry in the placement of the boundary. The results of the subsequent experiments also support this conclusion.

The observed two-step ABX discrimination functions are shown by open circles and broken lines and are plotted over corresponding labeling functions for comparison. Most subjects show evidence of categorical discrimination: there is a peak in discrimination function at the category boundary and there are troughs within both categories. Subject S2 is the most extreme example in the group, showing very nearly the idealized form of categorical perception.[8]

The labeling data and the discrimination functions indicate that categorical perception can be obtained quite easily with these nonspeech signals. To test the strength of these results against the categorical perception model, the ABX predictions from the labeling probabilities were compared with the observed discrimination functions.[4] A chi-square test was used to test the goodness of fit between the expected discrimination functions and the observed functions.[20] The observed and predicted discrimination scores, as well as the individual chi-square values for each subject are given in Table XIV-2.

The fit of observed and prediction functions is quite good in cases such as S2 and S6. In other cases, however, the fits are poor and the chi-square values reach a very conservative level of significance (i.e., S1, S4). In the case of S4 the discrimination function is the right shape and level but is just shifted slightly from the discrimination functions predicted from the labeling probabilities.

In general, however, the data from the present experiment show categorical perception effects that are at least as comparable as those obtained with speech sounds, particularly stop consonants. Thus the results of this study serve as another demonstration of categorical perception with nonspeech signals and suggest that this form of perception is not unique to speech stimuli.[21] But what is the basis for the present categorical perception results? Are these results due to some labeling process brought about by the training procedures as Lane has argued[19] or is there a simpler psychophysical explanation? In order to rule out the labeling explanation, it is necessary to obtain ABX

Table XIV-2. Observed and predicted ABX discrimination scores and chi-square values for goodness of fit.

Subject		Stimulus comparison (ms)								
		-50/-30	-40/-20	-30/-10	-20/0	-10/+10	0/+20	+10/+30	+20/+40	+30/+50	SUM
1	Observed (O)	.50	.71	.54	.46	.68	.82	.79	.79	.71	
	Predicted (P)	.50	.50	.50	.50	.50	.61	.88	.56	.51	
	O-P	0	.21	.04	.04	.18	.21	.09	.23	.20	
	Chi-square	0	5.04	0	0	3.64	5.32	1.96	5.88	4.76	26.6*
2	Observed	.50	.50	.50	.50	.46	.86	.79	.50	.50	
	Predicted	.50	.50	.50	.50	.50	.88	1.00	.51	.50	
	O-P	0	0	0	0	-.04	.02	.21	.01	0	
	Chi-square	0	0	0	0	0	0	0	0	0	0
3	Observed	.46	.54	.46	.50	.43	.43	.54	.54	.46	
	Predicted	.50	.50	.50	.50	.55	.51	.61	.68	.50	
	O-P	-.04	.04	-.04	0	.12	.08	.07	.14	-.04	
	Chi-square	0	0	0	0	1.68	.56	.56	2.52	0	5.4
4	Observed	.43	.39	.46	.50	.68	.96	.79	.82	.43	
	Predicted	.50	.50	.50	.50	.50	.68	.94	.58	.50	
	O-P	-.07	-.11	-.04	0	.18	.28	-.15	.24	-.07	
	Chi-square	.56	1.12	0	0	3.64	10.36	10.36	6.72	.56	33.3*
5	Observed	.57	.64	.71	.54	.61	.82	.64	.71	.50	
	Predicted	.51	.50	.50	.58	.68	.52	.50	.54	.51	
	O-P	.06	.14	.21	-.04	-.07	.30	.14	.17	.01	
	Chi-square	.02	.08	.18	.01	.02	.36	.08	.13	0	24.6
6	Observed	.50	.43	.50	.57	.82	.61	.43	.68	.50	
	Predicted	.51	.50	.52	.58	.56	.56	.52	.51	.50	
	O-P	-.01	-.07	-.02	-.01	.26	.05	-.09	.17	0	
	Chi-square	0	.56	0	0	7.84	.28	.84	.28	0	9.8
7	Observed	.50	.61	.54	.43	.71	.93	.96	.50	.57	
	Predicted	.50	.50	.50	.50	.50	.77	1.00	.54	.50	
	O-P	0	.11	.04	-.07	.21	.16	-.04	-.04	.07	
	Chi-square	0	1.12	0	.56	5.04	3.92	0	0	.56	11.2
8	Observed	.46	.50	.68	.57	.68	.79	.96	.75	.61	
	Predicted	.50	.50	.50	.50	.52	.68	.72	.58	.51	
	O-P	-.04	0	.18	.07	.16	.11	.24	.17	.10	
	Chi-square	0	0	3.58	.56	2.80	1.40	8.12	3.36	1.12	20.9

*$p < .001$
df = 8

discrimination functions before any training experience. If peaks in discrimination still remain in the absence of any labeling experience, we will have reason to suspect some psychophysical basis to the improved discrimination. The next experiment was carried out to test this hypothesis.

3. Experiment II

a. Method

Subjects. Twelve volunteers served as subjects. They were recruited in the same way as the subjects for the previous experiment and met the same selection requirements.

Stimuli. The eleven stimuli of Experiment I were also used in Experiment II.

Procedure. The procedures for the ABX discrimination tests were identical to those used in the previous experiment. The experiment covered two 1-hour sessions held on separate days. On each day the subjects received 360 ABX trials with immediate feedback provided for the correct response. At the end of the experiment each of the 9 two-step stimulus comparisons was responded to 80 times by each subject.

b. Results and Discussion

The ABX discrimination functions for all 12 subjects are shown in Fig. XIV-7. Except for S1 whose performance is close to chance, all of the other subjects show one of two patterns of discrimination performance. Four of the subjects show evidence of a single peak in the discrimination function at approximately +20 ms, whereas the rest of the subjects show discrimination functions with two peaks. For this group one peak occurs at approximately +20 ms, whereas a second peak occurs at approximately -20 ms. Broken vertical lines have been drawn through the discrimination functions at values of -20 ms and +20 ms to facilitate these comparisons.

The ABX discrimination functions reveal the presence of two distinct regions of high discriminability, one at roughly +20 ms that is comparable to that found in the previous experiment and another at approximately -20 ms. A reexamination of Fig. XIV-6 shows some evidence of a smaller peak in the -20 ms range for several subjects in Experiment I, although the major peak occurs at +20 ms and is correlated with changes in the labeling function.

It is clear from the results of Experiment II that the peaks in discrimination do not arise from the training procedures employed in Experiment I and the associated labels. Rather, it appears that natural catetories are present at places along the stimulus continuum that are marked by narrow regions of high sensitivity. Based on these results, it is possible to describe 3 categories within the -50 ms through +50 ms range. Going

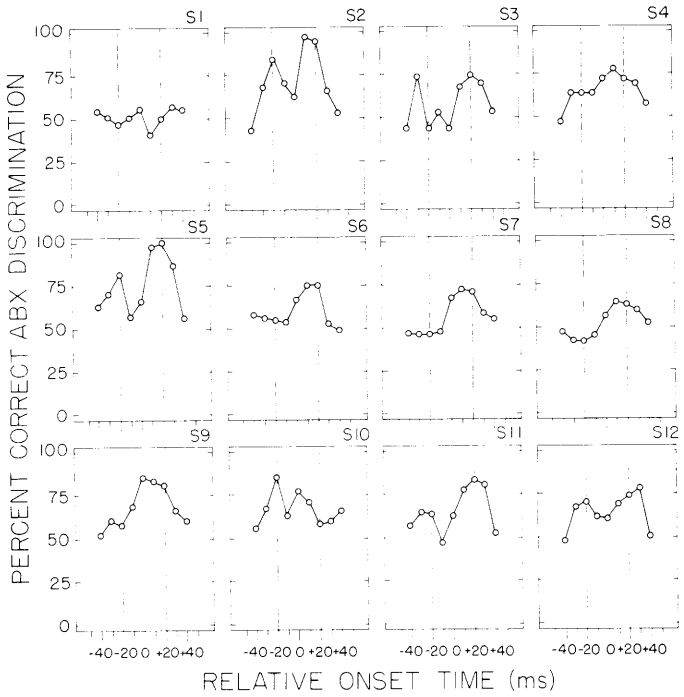

Fig. XIV-7. ABX discrimination functions for 12 individual subjects in Experiment II. Broken lines drawn through -20 ms and +20 ms illustrate the three regions of onset time.

from left to right, the first category contains stimuli with the lower tone leading by 20 ms or more; the second category contains stimuli in which both tones occur more or less simultaneously within the -20 ms to +20 ms region, whereas the third category contains stimuli in which the lower tone lags behind the higher tone by 20 ms or more. Within the context of this experiment, the three regions correspond, respectively, to leading, simultaneous, and lagging temporal events.

The presence of peaks in the ABX discrimination functions for these nonspeech stimuli is in sharp contrast to the results obtained previously by Liberman et al.,[5] and by Mattingly et al.,[9] who found marked differences in discrimination between speech and nonspeech signals. In these experiments, nonspeech control stimuli were created that nominally contained the same acoustic properties of speech but, nevertheless, did not sound like speech. For example, in the Liberman study the synthetic spectrograms of the /do/-/to/ stimuli were inverted before being converted to sound on the pattern playback. In the Mattingly study, the second-formant transitions (i.e., chirps) were isolated from the rest of the stimulus pattern, since it was assumed that these acoustic cues carry the essential information for place of articulation. When these nonspeech stimuli were presented to subjects in a discrimination task the discrimination functions that were obtained failed to show peaks and troughs that corresponded to those found with the parallel set of speech stimuli from which they were derived. The discrimination

functions were flat and very nearly close to chance in most cases, especially in the earlier study by Liberman and his co-workers.

The failure to find peaks and troughs in the discrimination functions of the nonspeech control stimuli may have been due to lack of experience with these stimuli and the absence of feedback during the discrimination task. With complex multidimensional signals it may be difficult for subjects to attend to the relevant attributes that distinguish these stimuli. For example, if the subject is not specifically attending to the initial portion of the stimulus but focuses instead on other properties, his discrimination performance may be no better than chance. Indeed, the Liberman et al. results indicate precisely this. Moreover, without feedback in tasks such as this the subject may focus on one aspect or set of attributes in a given trial and a different aspect of the stimulus in the next trial. As a result, the subject may respond to the same stimulus quite differently at different times during the course of the experiment. The results of Experiment II strongly indicate that nonspeech signals can be responded to consistently and reliably from trial to trial when the subject is provided with information about the relevant stimulus parameters that control his response.

The argument for the presence of 3 natural categories and our interpretation of the previous nonspeech control experiments would be strengthened if it could be demonstrated that subjects can classify these same stimuli into 3 distinct categories whose boundaries occur at precisely these regions on the continuum. We addressed this question in the next experiment.

4. Experiment III

In this experiment we used the same training procedures as in the first experiment except that subjects were now required to use three response categories instead of two. Our aim was to determine whether subjects would partition the stimulus continuum consistently into three distinct categories and whether the boundaries would lie at the same points of high discriminability identified in the previous experiments.

a. Method

Subjects. Eight additional subjects were recruited for Experiment III. They were obtained from the same source and met the same requirements as the subjects in the previous experiments.

Stimuli. The same basic set of 11 tonal stimuli was used in the present experiment.

Procedure. The experiment took place on two separate days. The first day was devoted to shaping and identification training with 3 stimuli; on the second day the labeling tests were conducted. Subjects were not given any explicit labels to use in the task and, as in the previous experiments, were free to adopt their own coding

strategies. The procedure used in this experiment was very similar to that used in Experiment I. Subjects were presented with three training stimuli, -50, 0 and +50 ms and were told to learn to respond differentially to these signals by pressing one of three buttons located on a response box. The order of presentation of the test sequences is given in Table XIV-3. Immediate feedback was provided for the correct response in each case.

Table XIV-3. Order of presentation of training and test sequences for Experiment III.

Day	Session	Sequence Description	Feedback	No. Trials
1	Training	Initial Shaping Sequence (-50, 0, +50)	Yes	180
1	Training	Identification Training (-50, 0, +50)	Yes	300
2	Training	Warm-up Sequence (-50, 0, +50)	Yes	90
2	Labeling	Identification Sequence (all 11 stimuli)	No	165

b. Results and Discussion

The identification functions for the eight subjects are shown in Fig. XIV-8. As shown here, all subjects partitioned the stimulus continuum into 3 well-defined categories. As anticipated, the boundaries between categories occur at approximately -20 ms and +20 ms. While there is some noise in the data as compared with the results of Experiment I, it is clear that subjects could reliably and consistently use the three responses and associate them with 3 distinct sets of attributes along the continuum. There is very little confusion or overlap between the three response categories, although the results are not as consistent as those obtained with the stop consonants.

The identification data from this experiment would probably have been more consistent if several additional members of each category were used during training, as in the first experiment, and if the range of stimuli were expanded slightly. Because of time constraints we used only one exemplar of each category during training. Further experiments are needed to resolve these questions.

In this experiment we did not explicitly provide subjects with an appropriate set of labels to use in encoding these sounds, although it is likely that they invented labels of their own. We assumed that by training subjects on representative members of a

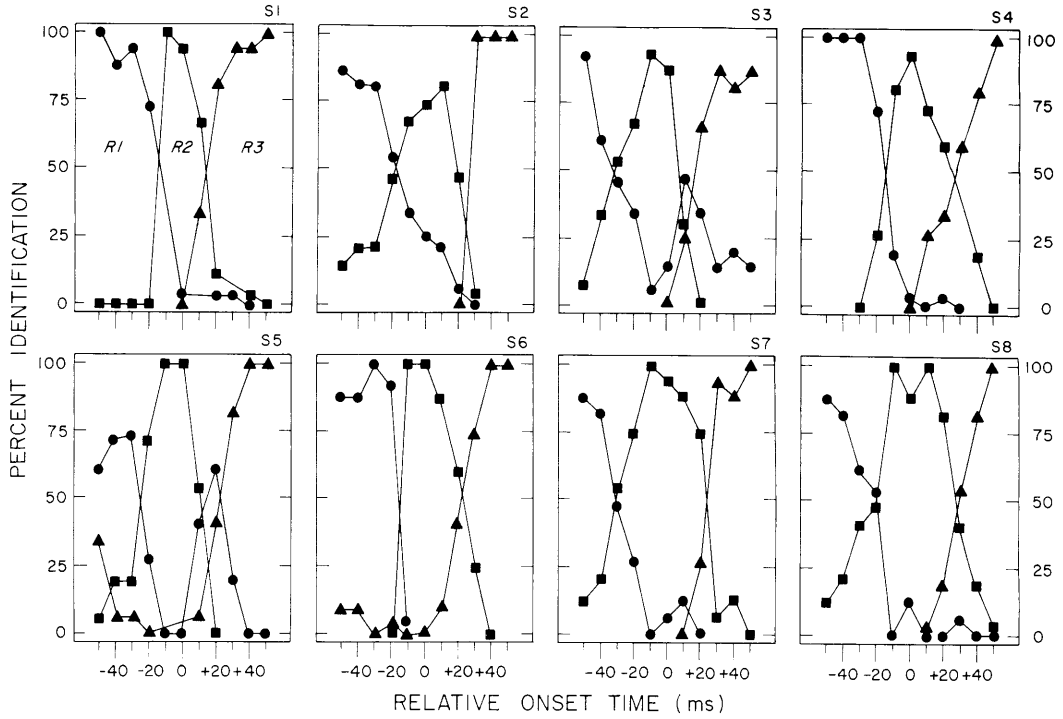

Fig. XIV-8. Labeling functions for individual subjects in Experiment III after training on -50 ms, 0 ms, and +50 ms stimuli as representative of each of the three categories.

category we could reveal some aspects of the underlying categorization process, and therefore gain some insight into the basis for defining category membership. The results of these experiments have revealed the presence of 3 natural categories that can be defined by the presence of certain distinct perceptual attributes at onset. These categories are separated by regions of high discriminability corresponding more or less to what might be called a perceptual threshold. We suggested earlier that the three categories observed along this continuum could be characterized by the subject's ability to discriminate differences in temporal order among the components of a stimulus complex. Thus the middle category corresponds to stimuli that have components appearing to be more or less simultaneous at onset, whereas both of the other two categories contain stimuli that differ in terms of two distinct events at onset separated by a very brief period of time.

In order to provide additional support for this account, we carried out another experiment in which subjects were required to determine whether they could perceive one or two distinct events at stimulus onset. The results of this study should provide information bearing on the potential range of attributes that define the perceptual qualities resulting from continuous variations in the relative onset of the two-tone components.

5. Experiment IV

a. Method

Subjects. Eight additional volunteers were recruited as subjects. None had participated in the previous experiments nor had any of them taken part in a previous psychophysical experiment. Thus they were experimentally naive observers.

Stimuli. The same 11 tonal stimuli were used.

Procedure. The experiment was conducted in a single 1-hour experimental session. Each of the eleven stimuli was presented singly, in random order. There were 40 replications of each stimulus, which gave a total of 440 trials. Subjects were told to listen to each sound carefully and then to determine whether they could hear one or two events at stimulus onset. They were told that in some trials the stimuli contained only one event at onset, whereas in other trials the stimuli contained 2 events at onset. Subjects were provided with a response box and told to press the button labeled "1" for one event at onset or the button "2" for two events at onset. No feedback was provided at any time during the experiment. There was a short break after the first 220 trials.

b. Results and Discussion

The results for each of the eight subjects are shown in Fig. XIV-9 where the percent judgments of two events are displayed as a function of the stimulus value. All subjects showed similar U-shaped functions with a fairly sharp crossover point between categories. There is a region in the center of the continuum, bounded by -20 ms and +20 ms, which is judged by every subject to contain stimuli whose components are predominantly simultaneous. On the other hand, there are two distinct regions at either end of the continuum in which subjects can reliably judge the presence of two distinct temporal events at stimulus onset, one leading and one lagging. Thus the results of this experiment, as well as the findings of the other experiments, indicate the presence of 3 natural categories that may be distinguished by the relative discriminability of the temporal order of the component events. This experiment indicates that such judgments are relatively easy to make and are consistent from subject to subject. This suggests the presence of a fairly robust perceptual effect for processing timing information which may also extend to the perception of voicing distinctions in stops. We shall return to this again when we attempt to develop an account of this effect in more detail.

6. General Discussion

The results of the present series of experiments are consistent with the findings of Hirsh,[22] Hirsh and Sherrick,[23] and, more recently, of Stevens and Klatt[11] who found that 20 ms is about the minimal difference in onset time needed to determine the temporal order of two distinct events. Stimuli with onset times greater than ~20 ms are

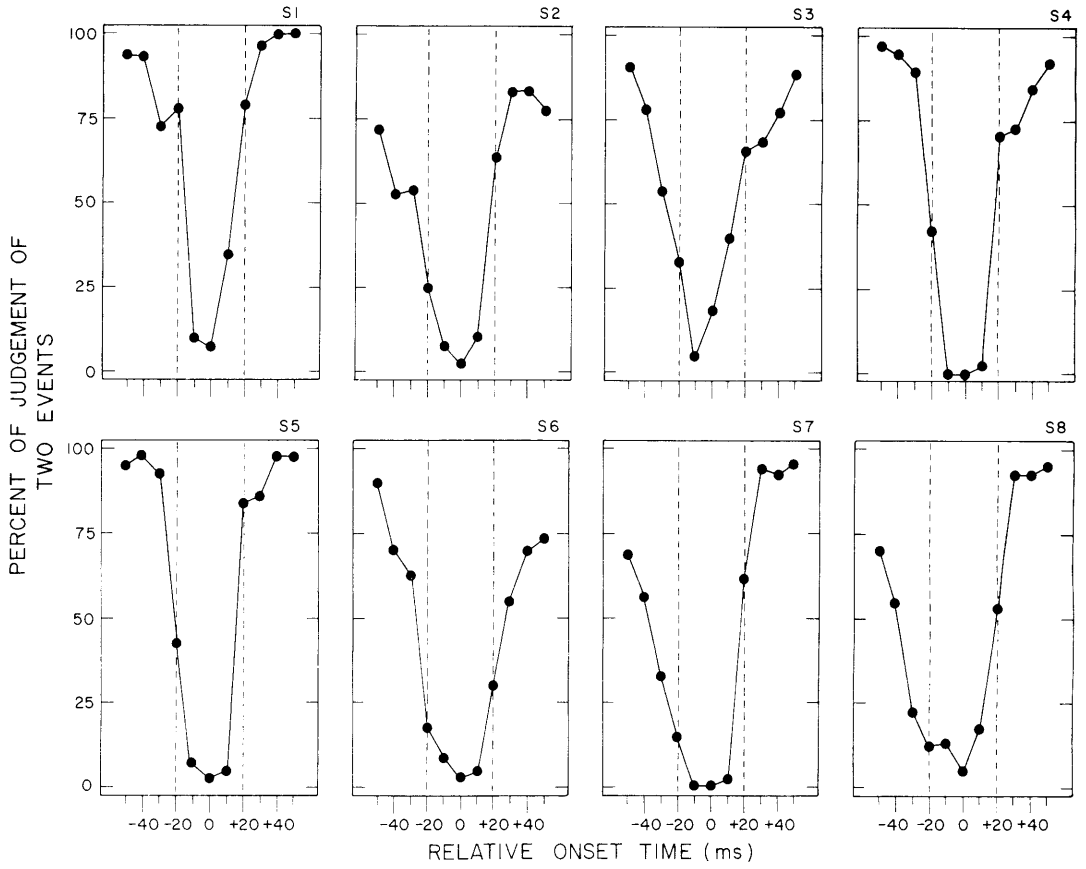

Fig. XIV-9. Percent judgment of two events for individual subjects as a function of relative onset time of the two components. Broken lines drawn through -20 ms and +20 ms permit comparisons.

perceived as successive events; stimuli with onset times less than ~20 ms are perceived as simultaneous events.

Based on the results of these four experiments with nonspeech stimuli differing in relative onset time, we would like to offer a general account of the labeling and discrimination data that can handle the four seemingly diverse sets of findings that have been reported. To review briefly, these 4 sets of findings are the perceptual results obtained for (i) infants, (ii) adults, and (iii) chinchillas with synthetic speech sounds differing in VOT, and (iv) the recent findings obtained for adults with nonspeech control stimuli differing in noise-lead time. Although specific accounts have been proposed to handle these findings individually, in our view a more general account of voicing perception is preferable.

We suggest that the four sets of findings simply reflect differences in the ability to perceive temporal order. In the case of the voicing dimension, the time of occurrence of an event (i.e., onset of voicing) must be judged in relation to the temporal attributes of other events (i.e., release from closure). The fact that these events, as well as

others involved in VOT, are ordered in time implies that highly distinctive and discriminable changes will be produced at various regions along the temporal continuum. Although continuous variations in the temporal relations may nominally be present in these stimuli, at least according to the experimental operational criteria, the only perceptual change to which the listener is sensitive appears to be the direction rather than the magnitude of difference between events. Thus the precision in perceiving specific temporal differences in tasks such as these is poor, whereas perceiving discrete attributes is excellent. This, of course, is the implication of the categorical perception experiments. Phonological systems have exploited this principle during the evolution of language. As Stevens and Klatt have remarked,[11] the inventory of phonetic features used in natural languages is not a continuous variable but consists of the presence or absence of sets of attributes or cues. This also seems to be the case with nonspeech stimuli having temporal properties similar to speech.

The account of voicing perception proposed here does not minimize the importance of the F1 transition cue[11, 24] or of the duration of aspiration noise preceding voicing onset,[15] as well as the numerous other cues to the voiced-voiceless distinction.[1] We argue that these cues are simply special cases of the more general process underlying voicing distinctions, i.e., whether the events at onset are perceived as simultaneous or successive.

The range of values found in the present experiments between -20 ms and +20 ms probably represents the lower limits on the region of perceived simultaneity. We assume that experience in the environment probably serves to tune and align the voicing boundaries in different languages and, accordingly, there will be some slight modification of the precise values associated with different regions along a temporal continuum such as VOT. It is also possible, as in the case of English voicing contrasts, that if appropriate experience is not forthcoming with the particular distinction, its discriminability will be substantially reduced. The exact mechanisms underlying these processes, as well as their developmental course, is under extensive investigation.[16, 25]

In summary, the results of these four experiments suggest a general explanation for the perception of voicing contrasts in initial position in terms of the relative discriminability of the temporal order between two or more events. These findings may be thought of as still another example of how languages have exploited the general properties of sensory systems to respresent phonetic distinctions. As Stevens has suggested,[26] all phonetic features of language probably have their roots in acoustic attributes with well-defined properties. We suggest that one of these properties corresponds to simultaneity at stimulus onset as reflected in voicing.

This work was supported in part by National Institute of Mental Health Grant MH-24027 to Indiana University. I wish to thank Jerry Forshee and Judy Hupp for their assistance in running subjects and processing data at Indiana University. I am also

grateful to Professor Kenneth N. Stevens and Dr. Dennis H. Klatt of M. I. T. for their interest and advice at various stages of this project.

References

1. L. Lisker and A. S. Abramson, "A Cross Language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20, 384-422 (1964).

2. L. Lisker and A. S. Abramson, "The Voicing Dimension: Some Experiments in Comparative Phonetics," in Proc. of the Sixth International Congress of Phonetic Sciences, Prague, 1967 (Academia, Prague, 1970), pp. 563-567.

3. A. S. Abramson and L. Lisker, "Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis," Proc. 5th International Congress of Acoustics, Liège, September 1965.

4. A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," J. Exp. Psychol. 54, 358-368 (1957).

5. A. M. Liberman, K. S. Harris, J. A. Kinney, and H. L. Lane, "The Discrimination of Relative Onset Time of the Components of Certain Speech and Nonspeech Patterns," J. Exp. Psychol. 61, 379-388 (1961).

6. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychol. Rev. 74, 431-461 (1967).

7. A. M. Liberman, "Some Characteristics of Perception in the Speech Mode," in D. A. Hamburg (Ed.), Perception and Its Disorders (Williams and Wilkins Co., Baltimore, 1970), pp. 238-254.

8. M. Studdert-Kennedy, A. M. Liberman, K. Harris, and F. S. Cooper, "The Motor Theory of Speech Perception: A Reply to Lane's Critical Review," Psychol. Rev. 77, 234-249 (1970).

9. I. G. Mattingly, A. M. Liberman, A. K. Syrdal, and T. Halwes, "Discrimination in Speech and Non-Speech Modes," Cognitive Psychol. 2, 131-157 (1971).

10. P. D. Eimas, E. R. Siqueland, P. Jusczyk, and J. Vigorito, "Speech Perception in Infants," Science 171, 303-306 (1971).

11. K. N. Stevens and D. H. Klatt, "The Role of Formant Transitions in the Voiced-Voiceless Distinction for Stops," J. Acoust. Soc. Am. 55, 653-659 (1964).

12. P. K. Kuhl and J. D. Miller, "Speech Perception by the Chinchilla: Voiced-Voiceless Distinction in Alveolar Plosive Consonants," Science 190, 69-72 (1975).

13. R. E. Lasky, A. Syrdal-Lasky, and R. E. Klein, "VOT Discrimination by Four to Six and a Half Month Old Infants from Spanish Environments," J. Exp. Child Psychol. 20, 213-225 (1975).

14. L. A. Streeter, "Language Perception of 2-Month-Old Infants Shows Effects of Both Innate Mechanisms and Experience," Nature 259, 39-41 (1976).

15. J. D. Miller, C. C. Wier, R. Pastore, W. J. Kelly, and R. J. Dooling, "Discrimination and Labeling of Noise-Buzz Sequences with Varying Noise-Lead Times: An Example of Categorical Perception" (to appear in J. Acoust. Soc. Am.).

16. P. D. Eimas, "Auditory and Phonetic Coding of the Cues for Speech: Discrimination of the r-l Distinction by Young Infants," Percept. Psychophys. 18, 341-347 (1975).

17. J. S. Bruner, J. J. Goodnow, and G. A. Austin, A Study of Thinking (John Wiley and Sons, Inc., New York, 1956).

18. A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position," Lang. Speech 1, 153-167 (1958).

19. H. L. Lane, "The Motor Theory of Speech Perception: A Critical Review," Psychol. Rev. 72, 275-309 (1965).

20. D. B. Pisoni, "On the Nature of Categorical Perception of Speech Sounds," Ph.D. Thesis, University of Michigan, August 1971; also SR-27 (Supplement), "Status Report on Speech Research," Haskins Laboratories, New Haven, Conn., 1971, pp. 1-101.

21. J. E. Cutting and B. S. Rosner, "Categories and Boundaries in Speech and Music," Percept. Psychophys. 16, 564-570 (1974).

22. I. J. Hirsh, "Auditory Perception of Temporal Order," J. Acoust. Soc. Am. 31, 759-767 (1959).

23. I. J. Hirsh and C. E. Sherrick, "Perceived Order in Different Sense Modalities," J. Exp. Psychol. 62, 423-432 (1961).

24. L. Lisker, "Is It VOT or a First-Formant Transition Detector?" J. Acoust. Soc. Am. 57, 1547-1551 (1975).

25. P. D. Eimas, "Developmental Aspects of Speech Perception," in R. Held, H. Leibowitz, and H. L. Teuber (Eds.), Handbook of Sensory Physiology: Perception (Springer-Verlag, New York, 1976).

26. K. N. Stevens, "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in E. E. David, Jr., and P. B. Denes (Eds.), Human Communication: A Unified View (McGraw-Hill Book Company, New York, 1972), pp. 51-66.

JSEP

C. COMPUTER-AIDED SIGNAL-PROCESSING CONTROL BY HIGHER LEVEL DIALOGUE-ORIENTED REPRESENTATIONS

Joint Services Electronics Program (Contract DAAB07-75-C-1346)

William L. Henke

The objective of this work is to find techniques of system representations that promote a rapid interactive graphical dialogue type of specification of time signal-processing configurations and parameter adjustments. Such systems find application in signal analysis and feature extraction domains where the signal source characteristics and/or the environment propagation characteristics are partially unknown or dynamically changing, and in situations where signal feature correlates are being sought for independently observed phenomena, for example, speaker voice correlates of emotional stress or some pathology. In such situations effective strategies are found best by on-line computer-aided human-directed and evaluated trials of potentially effective signal-processing, display, and decision techniques, and parameter values. Recent advances in digital hardware have made implementation of the relatively complex processing schemes feasible, and so the rapid selection/specification/adjustment/-design of such processes becomes a necessity for effective application of such "hardware" capabilities.

JSEP

JSEP

The fundamental requirements of such an approach are that appropriate processing "primitives", such as filters, Fourier transformers, sector space averagers, display integrators, and so forth, be defined and represented in such a way that they can be instantly "connected" into a compatibly designed network or higher level configuration representation. Graphically oriented representations have been found to be the most effective, and techniques have been developed to implement processing algorithms from such graphical descriptions.

Recent work has focused on making the process representation adaptable to matching the talents of different levels of users. "Operators" of lesser expertise need to be able to select and then adjust parameters of processing modules "designed" or packaged by "analysts" of greater expertise. A "macro-block" representation has been designed to facilitate such a hierarchy of levels, and techniques of "compilation" have been developed that allow the implementation of processing configurations from hierarchical representations. This representation has been implemented and evaluated as a result of being used by persons with widely diverse technical backgrounds, and the representation has proved to be easily mastered by most nonexperts.

JSEP