

Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals^{a)}

Edward M. Burns^{b)} and W. Dixon Ward

Hearing Research Laboratory, University of Minnesota, Minneapolis, Minnesota 55455
(Received 18 May 1977; revised 26 August 1977)

Categorical perception was investigated in a series of experiments on the perception of melodic musical intervals (sequential frequency ratios). When procedures equivalent to those typically used in speech-perception experiments were employed, (i.e., determination of identification and discrimination functions for stimuli separated by equal physical increments), musical intervals were perceived categorically by trained musicians. When a variable-step-size (adaptive) discrimination procedure was used, evidence of categorical perception (in the form of smaller interval-width DL's for ratios at identification category boundaries than for ratios within categories), although present initially, largely disappeared after subjects had reached asymptotic performance. However, equal-step-size discrimination functions obtained after observers had reached asymptotic performance in the adaptive paradigm were not substantially different from those initially obtained. The results of other experiments imply that this dependence of categorical perception on procedure may be related to differences in stimulus uncertainty between the procedures. An experiment on the perception of melodic intervals by musically untrained observers showed no evidence for the existence of "natural" categories for musical intervals.

PACS numbers: 43.66.Lj, 43.66.Hg, 43.66.Fc

INTRODUCTION

A. Musical-interval perception

The musical-interval sense—the ability of trained musicians to make magnitude or category judgments of the frequency ratio between two tones presented either sequentially (melodic musical intervals) or simultaneously (harmonic musical intervals)—is unique in several respects. The precision shown by musicians in adjusting musical intervals (subjective pitch ratios) is much greater than that shown by trained subjects adjusting ratios of other auditory or nonauditory percepts. For example, typical average intrasubject and intersubject standard deviations (of the physical stimulus) for adjustments of "octave above" are 0.8% and 3%, respectively (Ward, 1954). These standard deviations are an order of magnitude smaller than comparable figures obtained from magnitude-production experiments involving pitch extent (Stevens, Volkman, and Newman, 1937), loudness (Garner, 1952), or length of lines and area of squares (Stevens and Guirao, 1963).

On the other hand, since musicians are able to categorize musical intervals without immediate reference to some standard frequency ratio, musical-interval judgments are in this sense absolute judgments. Considered as such, musical intervals are also rather unique in that musicians are able to perfectly identify more than 30 categories of musical intervals, whereas the limit for absolute identification of other unidimensionally based auditory (e.g., Pollack, 1952, 1953) or nonauditory (Miller, 1956) percepts by trained observers is on the order of 5–9 categories.¹

Most modern experiments on the precision of the musical-interval sense have employed the method of adjustment. For example, in an extensive adjustment experiment, Moran and Pratt (1926) obtained mean intrasubject average deviations for adjustments of simultaneous intervals ranging from 13.5 cents for the fourth, to 22

cents for the tritone. These values are in line with the mean intrasubject average deviations for adjustments of simultaneous octaves of 11.1 cents found by Ward (1954) and with the range of the deviations for adjustments of simultaneous intervals reported recently by Rakowski (1976). The variance for adjustments of sequential intervals is of the same order of magnitude (Ward, 1954; Rakowski, 1976). Assuming that DL's for musical-interval size can be estimated from the standard deviation of repeated adjustments (Cardozo, 1965; Wier, Jesteadt, and Green, 1975), the results of the various adjustment experiments imply DL's in the range of 10 to 30 cents.

Apparently only one modern experiment on the precision of the musical-interval sense has employed a paradigm other than method of adjustment. Houtsma (1968) obtained estimates of interval-width DL's for melodic intervals (composed of sinusoids) employing a two-(temporal) interval, two-alternative forced-choice discrimination task with feedback. The average DL (based on the 75% correct points from the psychometric functions) for three subjects at the physical octave was 16 cents. The DL's at other ratios in the vicinity of the octave were not significantly different. A DL of 16 cents is in good agreement with the DL estimated from the standard deviation of repeated adjustments of sequential octaves (about 10 cents) in the same frequency region found by Ward (1954). Preliminary data for DL's at the ratios corresponding to the (just tempered) intervals of the chromatic scale were also obtained for one subject. As in Moran and Pratt's experiment, large differences were found for DL's at different ratios, but the range of DL's (14–25 cents) was in good agreement with their results.

In marked contrast to the extreme accuracy of musical-interval judgments in the experimental situations cited above is the large variability found in measurements of intonation in musical performance. The results of several studies on intonation in performance of western classical music have been summarized by Ward

(1970). They show large variations in the tuning of individual intervals (ranges of almost a semit) in a given performance. Similar variability has been found in measurements of intonation in nonclassical western music (Stauffer, 1954; Fransson, Sundberg, and Tjernlund, 1970; Owens, 1974) and in nonwestern music (Jhairabhoy and Stone, 1963; Callow and Shepherd, 1972; Spector, 1966). Of course this large variability is not in itself surprising since variability in production of tones is involved. The important point, however, is that in the above-cited studies, all listeners, including the performing musicians, agreed that the compositions were performed correctly and the large variability in intonation was not detected. This apparent inability to detect large variations in interval size in certain situations suggests that a phenomenon associated with the perception of speech tokens, "categorical perception," may be involved.

B. Categorical perception

Categorical perception has been of major importance in speech-perception research since it was first named and systematically studied by Liberman *et al.* (1957). As originally formulated, it refers to the concept that some stimuli can only be responded to on an absolute basis, i. e., discrimination is limited by identification. Thus, ideally, categorical perception is defined by two features: (1) well-defined identification functions exist—i. e., stimuli lying along some physical continuum can be systematically and reliably categorized by observers; and (2) discrimination functions for stimuli equally spaced along the same continuum can be correctly predicted from the identification functions, if it is assumed that subjects can discriminate stimuli no better than they can differentially identify them. It has been emphasized (e. g., Studdert-Kennedy *et al.*, 1970) that condition (2) predicts discrimination "troughs" (near chance-level performance) for stimuli which lie well within identification category boundaries, as well as discrimination peaks for stimuli which lie across category boundaries. The implication is that categorical perception involves not merely a sharpening of discrimination across category boundaries, but an actual inability to use other than labeling information as a basis for discrimination.

Categorical perception has been the focus of much controversy since it is seemingly at variance with most classical psychophysical discrimination results. Although grouping of the percepts elicited by a unidimensionally variable physical stimulus into categories is common in all sensory systems, this categorization usually has no effect on discrimination. Subjects can typically discriminate many more stimuli than they can categorize on an absolute basis, and the discrimination functions are smooth and monotonic (Pollack, 1952, 1953; Miller, 1956). This type of perception has been termed continuous perception.

The early research on categorical perception had produced no evidence for categorical perception of non-speech auditory or nonauditory stimuli, in spite of attempts to induce it by training (e. g., Mattingly *et al.*, 1971; Pisoni, 1971; Eimas, 1963; Studdert-Kennedy *et al.*, 1970), and in addition had indicated that even within

speech there was a dichotomy between categorically perceived stimuli (e. g., stop consonants) and continuously perceived stimuli (e. g., vowels; Fry *et al.*, 1962). These results were one of the primary factors which led to the formulation of models of speech perception in which perception is referenced to production (e. g., Studdert-Kennedy *et al.*, 1970). However, recent demonstrations of categorical perception for "speechlike" nonspeech stimuli (Miller *et al.*, 1976; Cutting and Rosner, 1974) and for chords or musical intervals (Locke and Kellar, 1973; Burns and Ward, 1974; Siegel and Sopo, 1975), show that such speech-specific models are not adequate to explain categorical perception.

Ideal (absolute) categorical perception as defined above is rarely achieved experimentally. Discrimination, especially within-category discrimination, is almost always somewhat better than predicted from the identification functions. Furthermore, a number of experiments have shown that variations in the temporal parameters of a discrimination paradigm, or variations in the paradigm itself which presumably affect the memory load of the task, lead to differing degrees of categorical perception (i. e., differences in within-category discrimination performance relative to between-category discrimination performance). This is true for both stop consonants (e. g., Barclay, 1972; Pisoni, 1971; Pisoni and Lazarus, 1974; Pisoni and Tash, 1974) and for vowels (e. g., Fujisaki and Kawashima, 1969; Pisoni, 1971). These results have led to a tempering of the original idea of a dichotomy between categorical and continuous perception, and recent papers allude to a continuum, e. g., "more categorical," "less categorical." To explain these results, a number of information-processing models have been developed (Fujisaki and Kawashima, 1970; Pisoni, 1971; Cutting and Pisoni, 1976) in which within- and between-category discriminations are based on information from different processing stages (e. g., "auditory" information *vis-a-vis* "phonetic" information). The information from the different stages is assumed to be differentially susceptible to memory effects (e. g., "echoic memory" *vis-a-vis* "short-term memory") and this accounts for the differences in within-category discrimination with variations in the parameters.

Another explanation of categorical perception (Miller *et al.*, 1976; Pastore, 1976) is based primarily on psychophysical grounds. According to this explanation, categorical perception merely reflects a sharpening of discrimination around a point along a sensory continuum at which there exists a perceptual dichotomy, e. g., a threshold between two distinct percepts. This explanation is primarily based on the results of an experiment (Miller *et al.*, 1976) in which categorical perception was demonstrated for a nonspeech continuum which mimicked the speech continuum of voice onset time. Categorical perception in this case was associated with the judged simultaneity or nonsimultaneity of the onset of two stimuli, a noise and a buzz. The view that categorical perception is associated with such "natural" boundaries has been strengthened by results which demonstrate that both infants (e. g., Eimas, 1975) and chinchillas (Kuhl, 1976) show discrimination peaks at points along speech continua that correspond to phonemic boundaries in adult

speakers. These results, along with the results of another experiment which demonstrated categorical perception for a "speechlike" nonspeech continuum (Cutting and Rosner, 1974) have led to the addition of a "linguistically relevant acoustic feature detector" stage in the latest version of the information-processing model (Cutting and Pisoni, 1976).

It is obvious from the above discussion that there is no consensus concerning the basis for categorical perception. Indeed it is not clear whether categorical perception is a robust phenomenon associated with certain classes of stimuli, or whether it is merely an epiphenomenon associated with the particular procedures used to elicit it. It is also possible that there is more than one type of categorical perception. Since a pilot study indicated that categorization processes are apparent in the perception of musical intervals (Burns and Ward, 1974), it seemed that musical-interval perception might be an ideal vehicle for studying categorical perception. Although the perception of pitch relations is a complex process, the physical stimuli themselves, pure tones, are simple and can be precisely controlled. Speech stimuli, on the other hand, are extremely complex, and variation of one parameter usually leads to variations in the complex interaction between all parameters of the signal. In addition, a pool of potential subjects who possess highly overlearned symbolic labels for the stimuli, perhaps the closest analogy to the highly overlearned labels for speech tokens to be found, already exists.

The series of experiments presented here, therefore, represents an investigation of categorical perception utilizing the musical-interval sense.

I. EXPERIMENT 1

A. Procedure and method

This experiment was essentially a replication of our pilot experiment (Burns and Ward, 1974). The experimental procedures were designed to conform as closely as possible to those commonly used in speech-perception experiments. Identification and discrimination functions were obtained for stimuli equally spaced along a physical continuum, over a range which subtended several perceptual categories. The stimuli were ascending melodic intervals (sequential frequency ratios) composed of two 500-msec sinusoids separated by a 200-msec silent period. Since essentially all musical scales are based on the premise that equal frequency ratios between pairs of tones lead to equal subjective interval sizes, the appropriate continuum in this case is the logarithmic frequency scale defining musical pitch—the cent scale (100 cents equals a frequency ratio of $2^{1/12}$; 1, and corresponds to an equitempered semit).

In the primary identification task ("three-category identification") the subjects were presented frequency ratios from the set composed of the ratios separated by increments of 12.5 cents over the range from 250 to 500 cents. The subjects were required to identify each of the ratios, by pressing an appropriately labeled button on a response panel, as belonging to one of five musical-interval categories based on the chromatic scale (i. e.,

categories separated by semits): major second (M2); minor third (m3); major third (M3); fourth (4); or tritone (TT). The subjects were asked to choose the closest category if the interval sounded "out of tune" or between categories. The ratios were presented in random order, with the restriction that each ratio was presented five times in each trial block; identification functions were based on four blocks of trials. Two such identification functions were obtained for each subject, one initially, and one after all of the discrimination tasks had been completed.

The subjects also performed an additional identification task ("quarter tone identification") in which they were asked to further categorize the out-of-tune ratios. The range of ratios presented was reduced from 250 to 200 cents (ratios used were 300–500 cents), and the subjects were required to identify the ratios as belonging to one of the following categories: minor third (m3); lower quartertone (QT_L); major third (M3); higher quartertone (QT_H); or fourth (4).

In the discrimination tasks the subjects were presented two successive melodic intervals and were asked to judge which interval was wider. No feedback was provided. The two ratios composing a trial were separated by a silent period (ISI) of 1 sec (measured from the offset of the second tone of the first ratio to the onset of the first tone of the second ratio). The two melodic intervals in each trial were adjacent ratios from the set of ratios separated by equal increments over the range from 250 to 550 cents (i. e., an equal-step-size discrimination task). The ratio pairs were chosen randomly with the restriction that each pair was presented ten times in a trial block. Order of presentation of the wider ratio in a pair was also randomized. Discrimination functions were based on three blocks of trials. Three such discrimination functions were obtained, for increment sizes of 25, 37.5, and 50 cents. For both identification and discrimination tasks, rate of presentation was controlled by the observer; a new trial was not presented until an answer for the previous trial had been obtained. For the identification task, response time (measured from the onset of the second tone of the ratio) was also recorded.

The sinusoids comprising the intervals were provided by a programmable oscillator (Krohn-Hite, 4141R) controlled by a small laboratory computer (DEC PDP-8/L). The amplitude envelopes were shaped by an electronic switch (GS 829S) to give rise and decay times of 10 msec. The tones were presented diotically (TDH39 headphones) at a level of 65 dB SPL (± 1 dB).

The frequencies of the first tones of the intervals were randomized over a range of 360 cents around 262 Hz ($f_1 = 262 \text{ Hz} \pm 60n$ cents where $n=0, 1, 2, \text{ or } 3$). This randomization is essential for the discrimination task to prevent subjects from basing their discriminations on pitch differences between the second tones of the intervals. It was also done for the intervals in the identification task for compatibility with the discrimination task, and to minimize order effects due to subjects basing their decisions on pitch differences between the second tones of intervals in different trials.

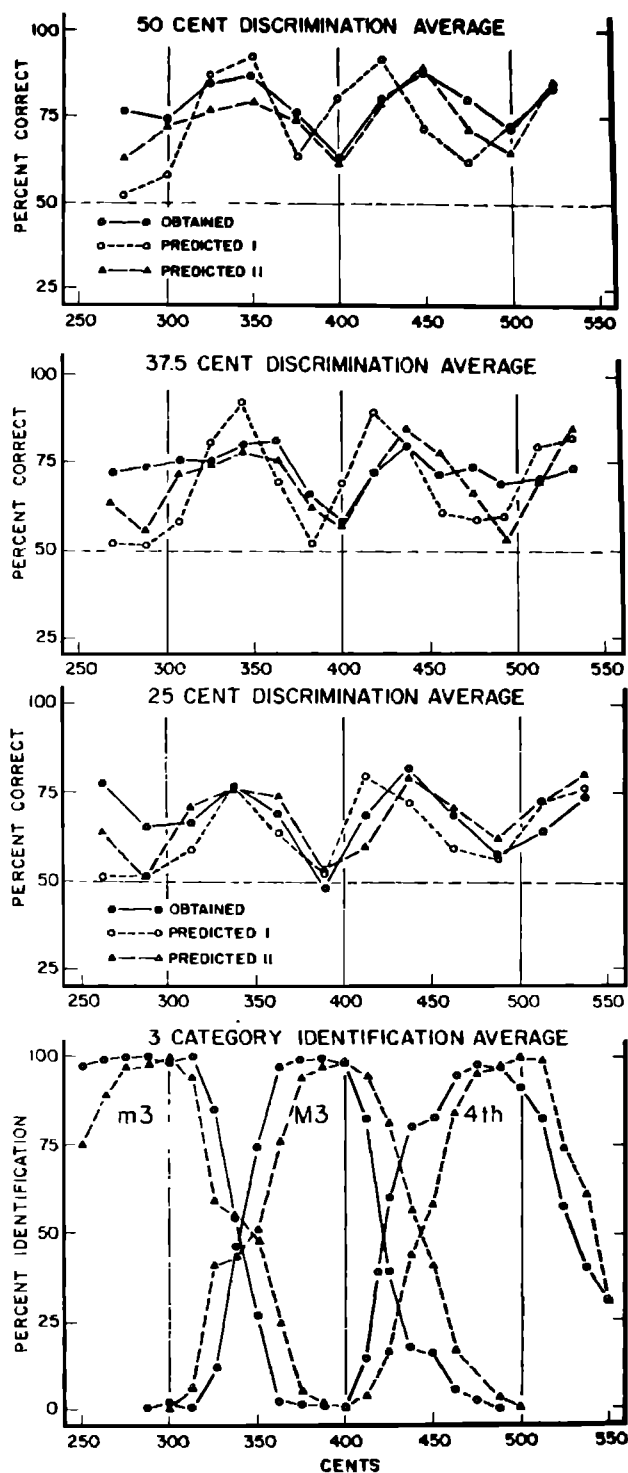


FIG. 1. Average initial (\bullet) and final (\blacktriangle) three-category identification functions, and obtained and predicted discrimination functions for five subjects.

The subjects were five music majors from the University of Minnesota who were paid for their services. All listening was done in a double-walled IAC audiometric booth.

B. Results and discussion

All of the subjects were able to identify the ratios as belonging to one of the five chromatic musical-interval

categories, and were able to do so consistently. That is, within-category identifications were at or near 100%, the transitions between adjacent categories were smooth and monotonic, and the transition regions were narrow relative to the width of the subjective categories. The average initial and final identification functions summed over the five subjects are shown in the bottom panel of Fig. 1. The substantial difference between initial and final identification functions presumably reflects long-term variability in subjective category criteria.

The three upper panels of Fig. 1 show, for comparison, the average discrimination functions for increment sizes of 50, 37.5, and 25 cents. These functions possess the form typically associated with categorical perception, i.e., discrimination troughs for stimulus values that lie well within subjective identification categories, and discrimination peaks for stimulus values that lie across category boundaries. Discrimination functions predicted from the identification functions, assuming absolute categorical perception, are also shown in Fig. 1.² The agreement between obtained and predicted discrimination functions is comparable to that found in stop-consonant perception experiments, for data pooled across subjects (see, e.g., Pisoni, 1971; Eimas, 1963; Liberman *et al.*, 1961).

In addition to the long-term intrasubject variability mentioned above, a relatively large intersubject variability is also characteristic of musical-interval judgments. This variability tends to flatten the peak-trough form of discrimination functions averaged over subjects since the location of category boundaries and discrimination peaks varies markedly among subjects.

The (initial) identification functions and the discrimination functions for an individual subject are shown in Fig. 2. In addition to identification functions, the average identification response time for each ratio is also shown in the lower graph. The response-time results are typical of those shown by all of the subjects, i.e., larger response times for ratios which lie at or near category boundaries than for ratios which lie well within categories. Similar results have been found for identifications of synthetic stop-consonant stimuli (Studdert-Kennedy, Liberman, and Stevens, 1963; Pisoni and Tash, 1974) and have been interpreted as additional evidence for categorical perception.

It is evident from the average results in Fig. 1 that the majority of subjects showed results similar to those of subject C4 (Fig. 2): Good agreement between obtained discrimination functions and those predicted assuming ideal categorical perception. Some individual subjects did, however, deviate substantially from this norm. One subject (C5) in particular, showed essentially no correlation between discrimination and identification functions (Burns and Ward, 1977).³

Although the results of the discrimination tasks are in good agreement with comparable data from stop-consonant perception experiments, they do not support the concept of ideal (absolute) categorical perception based on the categories into which the observers would normally place the stimuli (in this case categories based on the

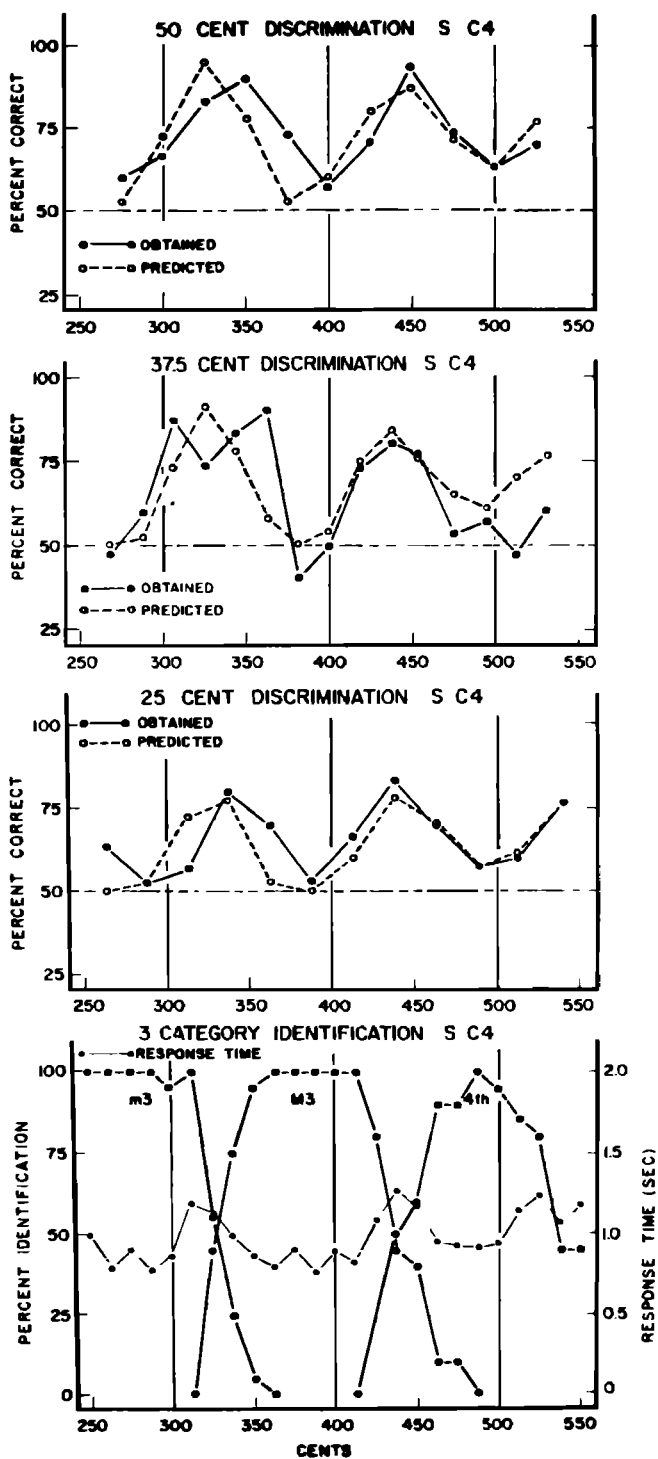


FIG. 2. Three-category identification functions, identification response time, and obtained and predicted discrimination functions for subject C4.

chromatic scale). Within-category discrimination for the wider increment sizes (37.5 and 50 cents) is, in general, better than that predicted on the basis of absolute categorical perception. A careful perusal of the discrimination data from stop-consonant perception experiments, however, shows similar results. If it can be assumed that the ratio of increment size to subjective identification category width is a valid metric for comparing experiments on different sensory continua, then the 25,

37.5, and 50 cent increment sizes in this experiment are comparable to 2, 3, and 4 step sizes in the relevant stop-consonant perception experiments. For larger step sizes (3 or greater), within-category discrimination for stop-consonant perception (e.g., Liberman *et al.*, 1961) is also better than predicted. Liberman *et al.* have attempted to explain these results in terms of the subjects being able to further categorize the stimuli.

Most of the subjects in this experiment were unable to consistently further categorize the ratios into categories based on quarter tones. Several of the subjects, however, e.g., C3, were able to identify the quarter tone between M3 and 4 consistently (Burns and Ward, 1977). Predictions based on quarter-tone identification over this range show fairly good agreement with obtained discrimination (Burns and Ward, 1977). It thus appears that, in some cases at least, better than predicted within-category discrimination can be explained on the basis of more precise identification, rather than on the basis of availability of a different type of information as proposed by the information-processing-type models.

In general, the results of experiment 1 confirm the results of our pilot experiment (Burns and Ward, 1974) and are in basic agreement with the results of Locke and Keller (1973) and Siegel and Sopo (1975); namely, that when equivalent procedures are used, musical intervals show categorical perception to a degree equivalent to that shown by stop consonants, the most "categorically perceived" speech stimuli. These results, along with those of experiments which have demonstrated categorical perception for other nonspeech stimuli (Miller *et al.*, 1976; Cutting and Rosner, 1974) are conclusive evidence against explanations of categorical perception based on unique modes of processing for speech.

III. EXPERIMENT 2

Pisoni (1971) reported an experiment in which the ISI in a 2AFC (same-different) discrimination task was varied from 0 to 2 sec. The results showed that the within-category discrimination of stop consonants (voiced stops and bilabial stops) was very poor and essentially independent of ISI, while the within-category discrimination of both long and short duration vowels was well above chance for short ISI's but decreased significantly with increasing ISI. Between-category discrimination for both stop consonants and vowels was very good and decreased only slightly with increasing ISI. These results provide the strongest support for the information processing type models discussed in the Introduction.

The applicability of such models to musical interval discrimination is not obvious. Clearly, the term "phonetic" must be replaced by a more general term, e.g., "symbolic." Discrimination based on "acoustic information" would presumably have to involve abstract pitch-extent judgments of the type made in mel pitch scaling experiments.

Experiment 2 was run to determine if temporal-parameter-dependent effects are also evident in the perception of musical intervals.

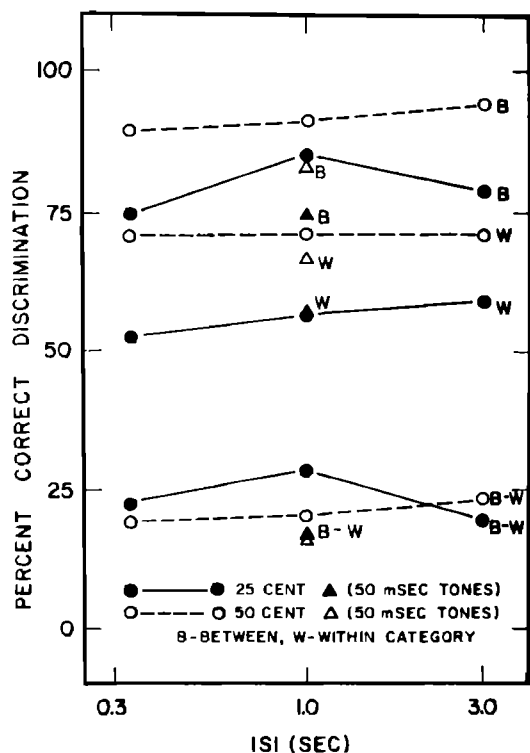


FIG. 3. Average within-category and between-category discrimination for four subjects for three values of ISI.

A. Procedures and methods

The experimental procedures were identical to those used in experiment 1 except that the ISI between interval pairs in the discrimination task was an independent variable. Discrimination functions for step sizes of 25 and 50 cents were obtained for intervals composed of 500-msec tones with ISPs of 330 msec and 3 sec. (Discrimination functions were also obtained for ratios composed of 50-msec sinusoids separated by an ISI of 1 sec. The results were essentially the same as those found for 500-msec sinusoids and are not discussed.) Four subjects from experiment 1 participated in the experiment.

B. Results and discussion

The discrimination functions for the 330-msec and 3-sec ISI's were not appreciably different from those for the 1-sec ISI, although there was a slight tendency for better within-category discrimination for the 3-sec-ISI condition (Burns and Ward, 1977). The results, averaged over the four subjects, are presented in Fig. 3. The points labeled "between category" are the averages of the maximum (peak) values, the points labeled "within category" are the averages of the minimum (trough) values. An estimate of the degree of categorical perception is given by the difference between between-category and within-category discrimination. None of the differences in discrimination as a function of ISI were significant by ANOVA. In particular, there is no decrease in within-category discrimination with increasing ISI as found with vowels (in fact the tendency is in the opposite direction). The fact that this is true even for the 50-cent discrimination function, where within-category discrimination is

well above chance, presents problems for the information processing models. It implies either that the "echoic memory" decays at a different rate for different types of "acoustic information," which makes the concept essentially meaningless, or that the basis for within-category and between-category discrimination of musical intervals is the same, i. e., differential identification. It was shown in experiment 1 that, in some cases at least, the much better obtained than predicted within-category discrimination could be explained on the basis of the subject's ability to identify the intervals more precisely than semits.

The results of experiment 2 provide further evidence that the perception of musical intervals is as "categorical" as that of stop consonants, and qualitatively different from the perception of vowels and most nonspeech stimuli. They also imply that explanations of categorical perception which propose a dichotomous basis for within- and between-category discrimination do not, as they are presently formulated, explain the results of musical-interval perception experiments.

III. EXPERIMENT 3

One implication of the "peak-trough" form of equal-step-size discrimination functions obtained for categorically perceived stimuli is that DL's obtained by the more usual psychophysical techniques, which employ variable step-size discrimination tasks, should also reflect this form. That is, significantly larger DL's should be obtained at stimulus values corresponding to troughs than at stimulus values corresponding to peaks. Assuming musical intervals in the vicinity of an octave are also perceived categorically in an equal-step-size discrimination task, the results of Houtsma (1968), in which no significant differences in DL's for musical-interval discrimination were found for ratios in the vicinity of the physical octave, are clearly at variance with this prediction. Since there is no evidence that the octave is perceived in a qualitatively different manner from other musical intervals (Ward, 1954), the differences between Houtsma's results and our own are apparently related to the differences in procedures. Experiment 3 was designed to determine whether evidence for categorical perception exists for a variable-step-size discrimination task and, in addition, to determine the effect of discrimination training on categorical perception. Previous attempts at improving within-category discrimination for speech stimuli by training have used equal-step-size discrimination tasks and have been largely unsuccessful (e. g., Strange, 1972).

A. Procedure and method

Because of the simplicity of the stimuli in our experiment, it was possible to use one of the efficient adaptive psychophysical techniques, in which the separation between stimuli is varied as a function of subject performance. In addition to providing estimates of DL's from a relatively small number of trials, these techniques also give an indication of the amount of "training" required to reach optimum performance—i. e., the number of trials needed to reach asymptotic performance,

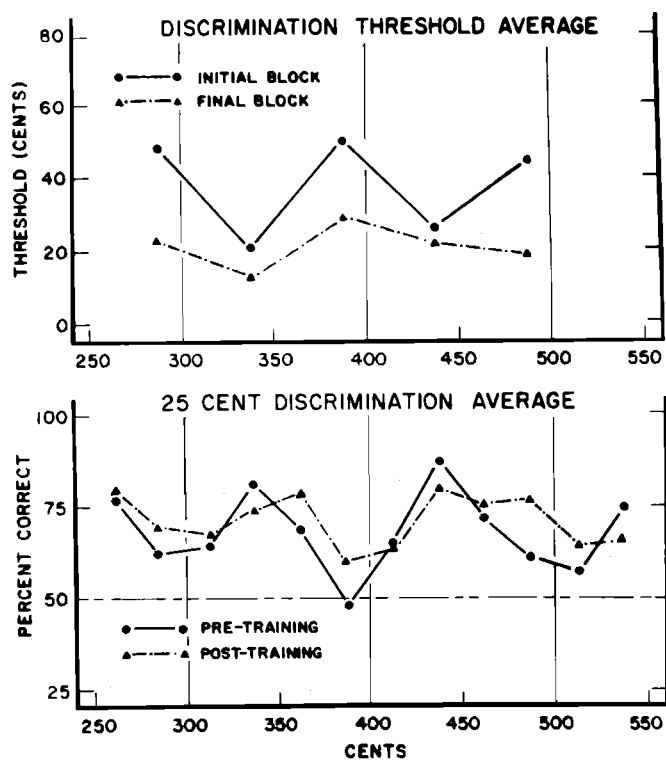


FIG. 4. Average initial and final (asymptotic) estimates of interval discrimination thresholds (DL's), and pretraining and post-training 25-cent discrimination functions for three subjects.

The temporal configuration of the stimuli was exactly the same as in the equal-step-size discrimination task in experiment 1. In this case, however, the two ratios to be discriminated in a given trial, instead of being adjacent ratios selected from the set of ratios separated by a fixed step size in cents over a 300-cent range, were ratios separated by a variable distance in cents around a fixed interval value. That is, intervals to be discriminated were $I + \frac{1}{2}\Delta I$ and $I - \frac{1}{2}\Delta I$, where I is the fixed interval value around which the DL is being estimated and ΔI is the (variable) interval separation. The paradigm was a transformed up-down method (Levitt, 1971) in which interval separation converges on the value at which the subject's performance will be 70.7% correct. Runs were terminated after 16 reversals and means and standard deviations of estimated DL's (i. e., interval separation for 70.7% correct) were calculated ignoring the first two reversals. The subject's task was, as in experiment 1, to indicate which interval was wider. Immediate feedback was provided.

Four of the subjects from experiment 1 participated in this experiment. Ratios at which DL's were estimated were those at which the peak and trough values of each individual subject's 25-cent discrimination functions (from experiment 1) were located.

After the subjects had reached asymptotic performance with this procedure (i. e., when the 70.7% correct threshold estimates of two consecutive runs differed by 5 cents or less), another 25-cent equal-step-size discrimination function was obtained for each subject.

B. Results and discussion

The average DL estimates (labeled "threshold") for the three subjects who had shown categorical perception in experiment 1 are shown in the upper panel of Fig. 4.⁴ The points labeled "initial block" refer to the DL estimate from the initial block of trials at each interval value. The points labeled "final block" are the DL estimates obtained after the subjects had reached asymptotic performance. The lower portion of each figure shows the average 25-cent discrimination function obtained in experiment 1 (labeled "pretraining") and the average 25-cent discrimination function obtained after the subjects had reached asymptotic performance (labeled "post-training").

The initial DL estimates, both for the average data and for individual subjects (Burns and Ward, 1977), showed a good correlation with the 25-cent discrimination curves, much larger DL estimates being found for intervals corresponding to troughs in the 25-cent discrimination function than for values corresponding to peaks,

This correlation between DL estimates and equal-step-size discrimination functions largely disappeared after the subjects had reached asymptotic performance, (i. e., after sufficient "training") in which case the subjects showed more nearly equal DL estimates for all interval values. The fact that, for asymptotic performance, the subjects showed approximately equal DL values across the range of intervals investigated implies that post-training equal-step-size discrimination functions should have been essentially flat. However, the 25-cent discrimination functions obtained after subjects had reached asymptotic performance were not substantially different from those obtained in experiment 1, although there was a slight increase in within-category discrimination performance. [The relatively flatter form of the post-training 25-cent discrimination function in the average data (Fig. 4) was due in large part to greater intersubject differences in peak-trough location that presumably reflect long-term variability in interval category criteria.]

The correlation between the initial DL estimates and the 25-cent discrimination functions, and the fact that many more trial blocks were required to reach asymptotic performance at within-category ratios than at between-category ratios, both indicate that categorical perception is more than an epiphenomenon associated with the equal-stimulus-separation discrimination paradigms usually used. However, the finding that subjects could, in general, be "trained" so that within-category DL's were equal to between-category DL's, but that even after such training the equal-stimulus separation discrimination functions still indicated pronounced categorical perception, shows that the phenomenon is to a large extent dependent on experimental procedures.

Carney, Widin, and Viemeister (1977), have recently obtained similar results for the speech continuum of voice onset time. The form of the psychometric functions obtained after training with a variable-step-size discrimination procedure suggested continuous, rather than categorical perception. In this case the training did appear to transfer to an equal-stimulus-separation discrim-

ination task; discrimination functions obtained using an oddity task were more nearly continuous than categorical.

As discussed previously, differences in the degree to which different experimental paradigms elicit categorical perception have been explained by information processing type models in terms of the temporal parameters and task requirements of the paradigms, and their assumed relationship to the encoding of, and short-term memory for, "auditory" and "phonetic" information. Since both the temporal characteristics and the task requirements of the two discrimination procedures used in this experiment were the same, explanations of this sort are obviously insufficient to explain the differences obtained. Thus the results of this experiment provide additional evidence against these types of models.

IV. EXPERIMENT 4: MUSICAL-INTERVAL DISCRIMINATION BY MUSICALLY UNTRAINED SUBJECTS

As discussed in the Introduction, recent results of experiments in the perception of certain nonspeech continua by adults and in the perception of certain speech continua by infants and chinchillas have been interpreted as reflecting an innate predisposition of the auditory processing system for certain natural categories (or category boundaries) for speech sounds, e.g., "linguistically relevant acoustic feature detectors."⁵

There are a number of theories that postulate natural categories for musical intervals, and indeed the question of natural scales is a long-standing controversy in musicology (see, e.g., Ward, 1970). Most of these theories are based on temporal-periodicity detection for simultaneous ratios and predict natural categories corresponding to small-integer frequency ratios (e.g., Meyer, 1898). These theories have been extended to melodic intervals (Boomsalter and Creel, 1961; Roederer, 1973) by assuming some sort of "neural reverberation." Another theory (Terhardt, 1974) also predicts small-integer ratio categories, based on early learning of the relations between the (usually harmonic) partials of natural sounds (e.g., voiced speech). In either case, the propensity for these "natural" categories would be expected to exist independent of musical training. Since we have shown fairly conclusively that musicians tend to perceive intervals categorically when certain procedures are used, a comparison of the discrimination of intervals by musically untrained subjects with interval discrimination by trained musicians using these procedures should give information on the possible existence of natural categories for musical intervals. Evidence for natural categories would be expected in one of two forms. If the categories were broadly tuned, as would be expected from the early-learning theory of Terhardt, untrained subjects should show discrimination results similar to the results shown by the musicians, that is, larger (initial) DL's for within-category intervals and corresponding troughs in the equal-stimulus-separation discrimination functions. On the other hand, categories based on temporal-periodicity or phase modulation detection might be expected to be finely tuned; that is, (small-integer) "frequency ratio detectors." In this

case subjects might show small DL's in the area of small integer ratios, and corresponding peaks in their equal-stimulus-separation discrimination functions.

A. Procedure and method

Six musically untrained subjects were tested on the adaptive discrimination paradigm used in experiment 3, and the 50-cent equal-step-size discrimination paradigm used in experiment 1. The basic criterion for musically untrained was an inability to recognize or name intervals.

In order to introduce the subjects to the nature of the task they were first given a practice run in the adaptive paradigm. Then an initial DL determination was made at each of five ratios: 300, 350, 400, 450, and 500 cents; these values correspond to within (300, 400, and 500) and between (350 and 450) category points, based on the equal-tempered scale. (The corresponding small-integer-ratio within-category values are 315.6, 386.3, and 498 cents.) The subjects were then run on the 50-cent equal-step-size discrimination paradigm. Finally, an attempt was then made with some of the subjects to reach asymptotic performance in the adaptive paradigm, for several ratios.

B. Results and discussion

The initial interval DL estimates for four subjects are shown in Table I. The data for two of the six subjects were discarded since they were unable to perform the task adequately. Both showed a correlation of greater than 0.5 between direction of frequency randomization and direction of interval judgments. That is, they showed a pronounced tendency to call the second interval wider if the frequency of the first tone of the second interval was higher than the frequency of the first tone of the first interval, and vice versa. Apparently they were unable to ignore changes in pitch height when trying to make interval width discriminations. The other four subjects showed only a slight correlation (less than 0.2), while the musicians showed essentially no correlation.

As seen in Table I, there is a large variability in the initial DL estimates. However, there is no obvious pattern to this variation. This is clearly shown in the average results, where the range of DL's for different ratios is only 7.9 cents. This stands in marked contrast to the initial DL estimates for the musicians in experiment 3, where the average between-category DL's were roughly half the within-category DL's (23.6 cents versus 47.1

TABLE I. Initial estimates of interval DL's (cents): Four musically untrained subjects.

Subject	Interval value (cents)					Average
	300	350	400	450	500	
D1	95.0	69.3	77.9	76.4	62.9	76.3
D2	50.0	52.9	55.0	72.9	97.3	62.0
D3	90.0	91.4	70.7	60.7	74.3	77.4
D4	<u>77.9</u>	<u>85.0</u>	<u>87.1</u>	<u>68.6</u>	<u>89.3</u>	<u>87.6</u>
Average	78.2	74.7	72.7	70.3	76.5	74.5

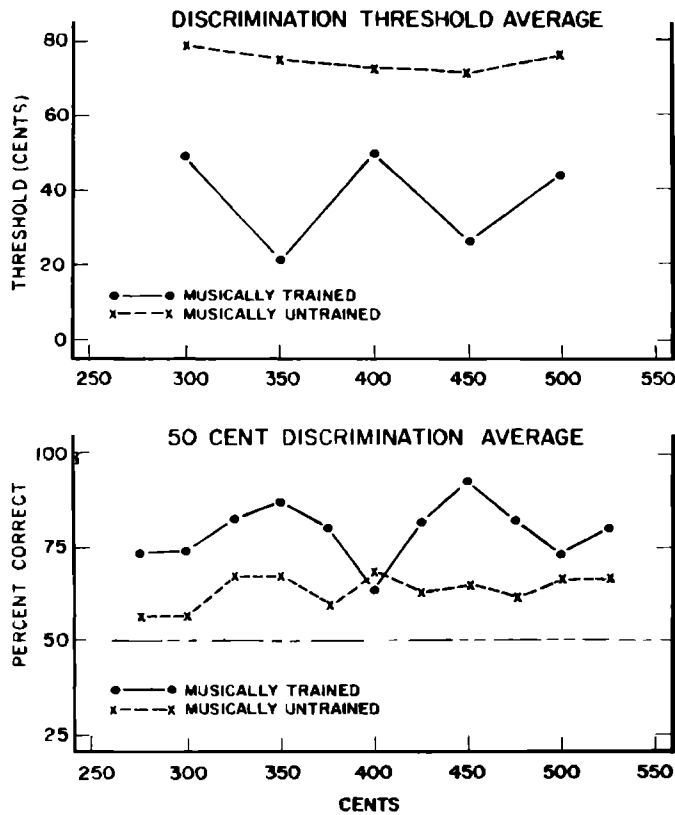


FIG. 5. Average interval width discrimination threshold estimates and 50-cent discrimination functions for four musically trained and four musically untrained subjects.

cents). In addition, the musicians were far superior in discrimination performance (average overall DL of 37.7 cents versus 74.5 cents for the musically untrained subjects). The average results of both the musically untrained subjects, and the musicians (from experiment 3) are plotted in the upper portion of Fig. 5.

Analogous results were obtained in the 50-cent equal-step-size discrimination task. The average results of the four subjects, along with the average results of the musicians from experiment 1, are shown in the lower portion of Fig. 5. The discrimination function for the musically untrained subjects is essentially flat and, except at 400 cents, the discrimination performance of musically untrained subjects is always poorer than that of the musicians, who showed the typical peak-trough results associated with categorical perception.

An attempt was made to reach asymptotic performance in the adaptive paradigm at several interval values for two of the musically untrained subjects. The results were characterized by two features: (1) There was large between-trial-block variability; instead of smoothly approaching an asymptotic value, as in the case of the musicians in experiment 3, the results of consecutive trial blocks showed large variations. (2) The increase in performance (decrease in estimated DL) over trial blocks was much smaller than that obtained for the musicians in experiment 4; minimum DL values obtained were on the order of 50–60 cents, compared to an average asymptotic value of 20.9 cents for the musicians.

The results of both discrimination procedures indicate that subjects who are not able to identify musical intervals use a different and considerably less efficient strategy for making frequency ratio discriminations than do musicians who are able to identify intervals. The results of neither paradigm, however, showed any evidence for natural categories. For musically untrained subjects, the equal-step-size discrimination function was essentially flat and there was no significant difference among initial DL's obtained at different ratios using the variable-stimulus-separation procedure. These results imply that "natural" perceptual boundaries along a sensory continuum are not a prerequisite for robust categorical perception, as has been proposed by Pastore (1976), although such boundaries obviously facilitate discrimination. The categorical perception of musical intervals is apparently based purely on learned categories.

V. PRELIMINARY EXPERIMENTS ON THE EFFECTS OF STIMULUS UNCERTAINTY

The fact that categorical perception is highly dependent on procedure and, as shown in experiment 3, can be essentially eliminated with moderate training when certain procedures are used, implies that the concept is of little use as an index of absolute perceptual ability. This does not imply, however, that the concept is without utility. If it can be shown that the experimental procedures which tend to elicit categorical perception more closely approximate conditions found in "real life" perception, e.g., perception of running speech, or music in performance, then studies of categorical perception should provide insight into the higher levels of auditory processing, involving, e.g., attention and memory. This is especially true for cases, of which melodic musical-interval perception appears to be one, where categorical perception is not concomitant with natural boundaries imposed by characteristics of the peripheral auditory processing system. It is of interest, therefore, to try to determine the relevant differences between the procedures that elicited different degrees of categorical perception for melodic intervals in experiments 1 and 3.

A likely factor is the difference in stimulus uncertainty between the procedures. In the equal-step-size discrimination task used in experiment 1, stimuli in any given trial could be in one of five possible perceptual categories, and since stimuli were chosen at random, there was no trial-to-trial continuity. Uncertainty in this task is obviously much higher than in the adaptive procedure used in experiment 3 where stimuli in any given trial were in the region of one, or at most two, perceptual categories, and the stimuli in consecutive trials came in a predictable manner based on performance. Watson, Kelly, and Wroton (1976) give a detailed discussion of the effects of stimulus uncertainty on discrimination. For simple discrimination, such as pure-tone frequency discrimination, stimulus uncertainty has relatively little effect. If this frequency discrimination is embedded in a complex stimulus, however, stimulus uncertainty associated with the stimulus itself, or with the procedure, has an enormous effect on discrimination ability (DL's differ by up to a factor of 40). Since discrimination of frequency ratios is a much more complex task than pure-

tone frequency discrimination, it seems reasonable to assume that stimulus uncertainty will also have a large effect in this case.

There would seem to be two alternative explanations of the effect of stimulus uncertainty on frequency ratio discrimination. It is possible that the increased uncertainty in the equal-step-size discrimination task forces subjects to rely on absolute identification rather than on abstract judgments of pitch extent, which may be the basis for discriminations in the adaptive paradigm. Alternatively, it may be that the basis of discrimination is the same in both procedures, namely absolute identification, but the reduced uncertainty in the adaptive paradigm allows the subject to classify the ratios into finer categories, and hence to discriminate more precisely.

The latter would seem to be the more likely alternative for several reasons. First, in postexperiment discussions, the subjects claimed that their strategy in making discriminations was the same in experiment 3 as in experiment 1, and was on the basis of identification of the intervals (e.g., "flat major third" versus "good major third," etc.) Second, the results of experiments 1 and 2 implied that within-category discriminations were on the basis of more precise identification, rather than on other information. Finally, the fact that discrimination performance of musicians in experiment 4 was always better than that of nonmusicians (except at 400 cents), who presumably base their judgments on pitch extent information, also implies that the musicians are using labeling information.

Two experiments (5 and 6), in which slightly different procedures were used, were undertaken to investigate the role of stimulus uncertainty on interval perception. In experiment 5, the effect of reducing interval uncertainty on the subjects' ability to categorize the intervals more precisely than into semits was investigated. In experiment 6, the effect of increasing stimulus uncertainty, by increasing the range of intervals presented, on the degree of categorical discrimination was investigated.

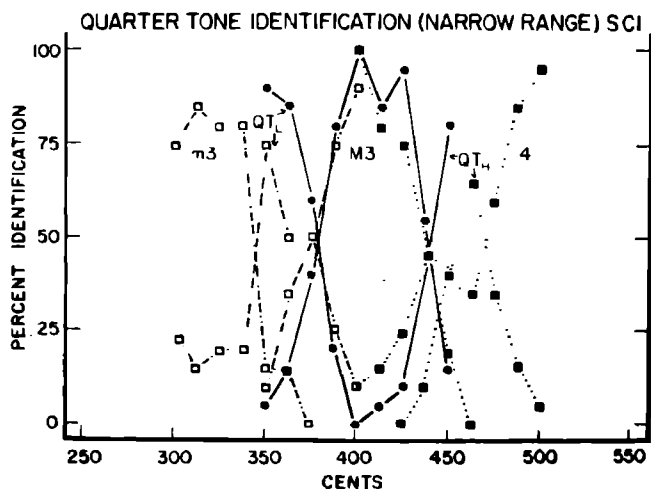


FIG. 6. Narrow range (100-cent) quarter-tone identification functions for subject C1.

A. Experiment 5

1. Procedure and methods

Two subjects each completed three variations of a quarter-tone identification task, similar to that performed in experiment 1, but over a reduced range. In each variation, the range of intervals presented was limited to 100 cents and was centered at one of the interval values at which DL estimates had been obtained in experiment 3. The responses were limited to three categories. The range and respective response categories for the three variations were: (1) 300–400 cents; m3, QT_L, M3; (2) 350–450 cents; QT_L, M3, QT_H; and (3) 400–500 cents; M3, QT_H, 4. This procedure represented an attempt to approximate the conditions faced by the subject in identifying the intervals in the adaptive procedure. However, the intervals in that procedure are presumably even less uncertain since they come in a predictable manner (depending on performance) rather than in random order as in experiment 5.

2. Results and discussion

The results of the quarter-tone identification task for one of the subjects (C1) are shown in Fig. 6. A comparison with her previous attempts to categorize intervals more precisely than semits (see Burns and Ward, 1977) showed that restricting the range of possible intervals, which presumably reduced the uncertainty, did indeed result in more consistent identification functions. The results of the other subject showed a similar improvement. It is not possible to ascertain from these limited results whether identification is precise enough to account for the results of the adaptive paradigm. Interval width DL's at 350, 400, and 450 cents for subject C1 predicted from the data in Fig. 6 (20, 36, and 20 cents) agree fairly well with the asymptotic values obtained in experiment 3. Unfortunately, the subjects who showed the smallest DL estimates in experiment 3, C3 and C4, were not available for experiment 5.

B. Experiment 6

1. Procedure and method

Four subjects completed two equal-stimulus-separation discrimination tasks. One was a 50-cent discrimination task over the range 250–550 cents, exactly equivalent to the 50-cent discrimination task in experiment 1. The second was a 50-cent discrimination task over the range 150–950 cents, which is labeled "wide-range" discrimination. The latter range encompasses ten possible chromatic semit categories (minor second to minor seventh) compared with the five possible categories in the range 250–550 cents. Equipment limitations made it impossible to obtain identification functions over the same range.

2. Results and discussion

The average results of the 50-cent discrimination tasks for the three subjects are shown in Fig. 7. In addition to the 50-cent discrimination function for the range 250–550 cents (labeled "post-training") and the 250–550 cent

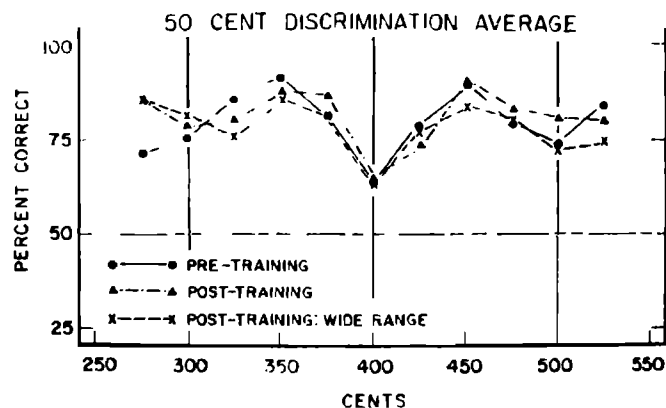


FIG. 7. Average pretraining, post-training, and wide-range post-training 50-cent discrimination functions for three subjects.

portion of the 150–950 cent discrimination function (labeled “post-training: wide range”), the results of the 50-cent discrimination function from experiment 1 (labeled “pretraining”) are also shown.

As reflected in Fig. 7, there is no evidence that the increase of range, and presumed corresponding increase in stimulus uncertainty, had any significant effect on the form of the obtained discrimination functions. Specifically, there is no evidence that the degree of categorical perception (as determined by peak-trough differences) was increased with an increase of stimulus uncertainty. The main effect of the wider range was a slight decrease in overall discrimination performance.

C. Discussion

In experiment 5, the reduction of the interval range, and the corresponding reduction in uncertainty, led to more precise quarter-tone identification functions than had been obtained in experiment 1. It is hypothesized that an analogous increase in identification precision in the variable-stimulus-separation paradigm relative to the equal-stimulus-separation paradigm may account for the better discrimination performance and lack of categorical perception (based on semit categories) in that procedure. However, a corresponding increase in range (doubling of possible semit categories) in experiment 6 seemed to have no effect on the degree of categorical perception over the range investigated (250–550 cents). This result is not as contradictory as it would appear. Watson, Kelly, and Wroton (1976) have shown that when stimulus uncertainty is relatively high, a further increase has little effect. It may also be that the extensive experience of trained musicians in identifying semit categories over the range of an octave has led to a degree of over-learning, such that little is gained in reduction of uncertainty, and hence in precision of identification of semit categories, by limiting the range to five categories. It was not possible to directly investigate the identification of intervals over the wider range because of equipment limitations.

Although the results of experiment 5 provide some evidence that the differences between the equal-stimulus

separation and adaptive procedures in eliciting categorical perception may be due to the relative precision with which the subjects can identify the intervals in the respective paradigm, the evidence is at best preliminary. It is clear, however, that for a complete understanding of categorical perception the effects of stimulus uncertainty will have to be thoroughly investigated. For example, a more efficacious method of ascertaining the role of identification in the adaptive paradigm would be to obtain labeling data simultaneously with discrimination data.

VI. SUMMARY AND CONCLUSIONS

The results of these studies show quite conclusively that, when equivalent procedures are used, the perception of melodic musical intervals is equivalent to the perception of stop consonants. Specifically, when equal step-size-discrimination tasks are used, musical intervals are perceived categorically, whereas when variable-step-size discrimination tasks are employed, categorical perception can be eliminated with moderate training.

Although musical intervals and stop consonants elicit similar results, none of the current explanations of categorical perception adequately explain the perception of musical intervals: (1) Speech-specific models obviously are not applicable; (2) there is no evidence that musical-interval discriminations are based on information from more than one processing stage, as proposed by information-processing-type models; and (3) there is no evidence for the existence of natural perceptual boundaries for melodic musical intervals, although such boundaries clearly exist for certain speech continua.

It is tentatively concluded that categorical perception in the case of musical intervals is related to the degree of stimulus uncertainty associated with the procedures used, although obviously considerably more research is necessary in this regard.

These results have some obvious implications for musical perception. First, since stimulus uncertainty in “real world” perception is, in general, high, it might be expected that categorical perception of musical pitch would be the normal situation. This conclusion is supported by the results of the various investigations of intonation in performance. Second, the lack of evidence for the existence of natural categories for musical intervals implies that individuals in a given culture learn the scales of their culture from experience, not because of any innate propensity of the auditory system for specific intervals.

ACKNOWLEDGMENT

The authors thank Marnie Cushing and John Bradley for preparing the figures.

²This article is based on a doctoral dissertation submitted by the first author to the Graduate School, University of Minnesota, 1977. Portions of these results were reported at the 86th, 90th, and 92nd Meetings of the Acoustical Society of America: Los Angeles, CA, Nov. 1973; San Francisco, CA,

Nov. 1975; San Diego, CA, Nov. 1976.

^{b)}Present address: Research Laboratory of Electronics, 36-763, Massachusetts Institute of Technology, Cambridge, MA 02139.

¹Experiments to determine the maximum number of perfectly identifiable interval categories have apparently not been attempted. A number of experiments (e.g. Plomp, Wagenaar, and Mimpen, 1973; Killam, Lorton, and Schubert, 1976) have shown that some subjects can perfectly identify the 12 ascending and 12 descending intervals of the chromatic scale. In an informal experiment in our laboratory, one of our subjects was able to perfectly identify ascending and descending intervals from unison to major tenth (33 categories). A greater range was not attempted. It is a moot point, however, whether or not ascending and descending versions of the same melodic interval, or intervals separated by an octave, represent separate categories.

Experiments on possessors of absolute pitch, the obvious exceptions to Miller's "7 ± 2" rule, show that the best subjects can perfectly identify about 75 categories (Ward, 1963). Their identifications are based on musical scale categories; however, the relationship, if any, between absolute pitch and interval perception is not clear.

²For a 2AFC discrimination task with categories A and B and stimulus values 1 and 2, the predicted discrimination (proportion correct) assuming ideal categorical perception is given by $P_d = 0.5[P_1(A)P_2(A) + P_1(B)P_2(B)] + 1.0[P_1(A)P_2(B) + P_1(B)P_2(A)]$. $P_1(A)$ is the probability that stimulus 1 is identified as A, $P_2(A)$ is the probability that stimulus 2 is identified as A, etc.

³See AIP document No. PAPS JASMA-63-456-24 for 24 pages of figures showing the individual subject data for the experiments reported in this paper. Order by PAPS number and journal reference from American Institute of Physics, Physics Auxiliary Publication Service, 335 East 45th Street, New York, N.Y. 10017. The price is \$1.50 for each microfiche or \$5.00 for photocopies of up to 30 pages with \$0.15 for each additional page over 30 pages. This material also appears in *Current Physics Microfilm*, the monthly microfilm edition of the complete set of journals published by AIP, on the frames immediately following this journal article.

⁴Since subject C5 did not show categorical perception in experiment 1, and since our primary emphasis in this experiment is in comparing the effects of different procedures on categorical perception, her results are not included in the average data. A comprehensive discussion of individual subject data is presented in the original thesis.³

⁵Since many phonetic features are determined by steady-state (vowels) or time-varying (consonants) relationships among formant regions, there is an obvious analogy between formant ratio perception in speech and frequency ratio perception in music, and indeed a number of studies have been based on this analogy. On reflection, however, it is clear that this is a tenuous comparison at best. There is little physical or perceptual similarity between frequency (pure-tone) ratios and formant-region relationships. In the case of varying frequency ratios, the primary perceptual changes are a change in the pitch relations of the tones and a change in the fundamental (residue) pitch, both of which are determined by the frequency relations of the tones. Timbre is essentially constant. In the case of varying formant regions the pitch is constant (since the relations among the harmonics are constant) and the percept is one of timbre change.

Barclay, R. J. (1972). "Noncategorical Perception of a Voiced Stop: A Replication," *Percept. Psychophys.* **11**, 411-412.

Boomsalter, P., and Creel, W. (1961). "The Long Pattern Hypothesis in Harmony and Hearing," *J. Mus. Theory* **5**, 2-31.

Burns, E. M., and Ward, W. D. (1974). "Categorical perception of musical intervals," *J. Acoust. Soc. Am.* **55**, 456(A).

Burns, E. M., and Ward, W. D. (1973). "Categorical Perception—Phenomenon or Epiphenomenon: Evidence from Experiments in the Perception of Musical Intervals. Supplementary Material; Individual Subject Data," AIP Document No. PAPS JASMA-63-456-24 (see footnote 3 of the present paper).

Callow, G., and Shepherd, E. (1972). "Intonation in the Performance of North Indian Classical Music," Paper presented at the 17th annual meeting of the Society for Ethnomusicology (Nov. 30-Dec. 3, 1972, Toronto, Canada) (unpublished).

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

Cardozo, B. (1965). "Adjusting the Method of Adjustment," *J. Acoust. Soc. Am.* **37**, 786-792.

Cutting, J., and Pisoni, D. (1976). "An Information-Processing Approach to Speech Perception," Paper presented at Implications of Basic Speech and Language Research for the School and Clinic (Belmont, Elkridge, Maryland, 23-26 May 1976) (unpublished).

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

Eimas, P. D. (1963). "The Relation Between Identification and Discrimination Along Speech and Nonspeech Continua," *Lang. Speech* **6**, 206-217.

Eimas, P. D. (1975). "Speech Perception in Early Infancy," in *Infant Perception*, edited by L. B. Cohen and P. Salapetek (Academic, New York,).

Fransson, F., Sundberg, J., and Tjernlund, P. (1970). "Statistical Computer Measurements of the Tone-Scale in Played Music," *Speech Transmission Laboratory Quarterly Progress and Status Report 2-3* (Royal Institute of Technology, Stockholm, Sweden).

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). "The Identification and Discrimination of Synthetic Vowels," *Lang. Speech* **5**, 171-189.

Fujisaki, H., and Kawashima, T. (1969). "On the Modes and Mechanisms of Speech Perception," *Ann. Rep. Eng. Res. Inst.* **28** (University of Tokyo), 67-73.

Fujisaki, H., and Kawashima, T. (1970). "Some Experiments on Speech Perception and a Model for the Perceptual Mechanism," *Ann. Rep. Eng. Res. Inst.* **29** (University of Tokyo) 207-214.

Garner, W. R. (1952). "Statistical Aspects of Loudness Judgments," *J. Acoust. Soc. Am.* **24**, 153-157.

Houtsma, A. J. M. (1968). "Auditory Discrimination of Frequency Ratios," *J. Acoust. Soc. Am.* **44**, S383(A).

Jhairazbhoy, N., and Stone, A. (1963). "Intonation in Present Day North Indian Classical Music," *Bulletin of the School of Oriental and African Studies* **26** (University of London), 118-132.

Killam, R., Lorton, P., and Schubert, E. (1976). "Interval Recognition: Identification of Harmonic and Melodic Intervals," *J. Mus. Theory* **20**, 212-234.

Kuhl, P. (1976). "Speech perception by the chinchilla: Categorical perception of synthetic alveolar plosive consonants," *J. Acoust. Soc. Am.* **60**, S581(A).

Levitt, H. (1971). "Transformed up-down methods in psychoaoustics," *J. Acoust. Soc. Am.* **49**, 467-477.

Liberman, A. M., Harris, K. S., Eimas, P., Liesker, L., and Bastian, J. (1961). "An Effect of Learning on Speech Perception: The Discrimination of Durations of Silence With and Without Phonemic Significance," *Lang. Speech* **4**, 175-195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. S. (1957). "The Discrimination of Speech Sounds Within and Across Phonem Boundaries," *J. Exp. Psychol.* **61**, 379-388.

Locke, S., and Kellar, (1973). "Categorical Perception in a Non-Linguistic Mode," *Cortex* **9**, 355-369.

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

- Meyer, M. (1998). "Zur Theorie Der Differenztone und der Gehörsempfindungen überhaupt," *Beitr. Akust. Musikwiss.* **2**, 25-65.
- Miller, G. A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychol. Rev.* **63**, 81-96.
- Miller, J., Wier, C., Pastore, R., Kelly, W., and Dooling, R. (1976). "Discrimination and labeling of noise-buzz sequences with varying noise-level times: An example of categorical perception," *J. Acoust. Soc. Am.* **60**, 410.
- Moran, H., and Pratt, C. C. (1926). "Variability of Judgments of Musical Intervals," *J. Exp. Psychol.* **9**, 492-500.
- Owens, T. (1974). "Applying the Melograph to 'Parkers Mood'," *Selected Reports in Ethnomusicology* **2**, (Institute for Ethnomusicology, University of California at Los Angeles) 166-175.
- Pastore, R. (1976). "Categorical Perception: A Critical Re-evaluation," in *Hearing and Davis: Essays Honoring Hallowell Davis*, edited by S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, and S. R. Silverman (Washington University, St. Louis, Missouri), pp. 253-263.
- Pisoni, D. B. (1971). "On the Nature of Categorical Perception of Speech Sounds," Ph.D. dissertation (University of Michigan, Ann Arbor) (unpublished).
- Pisoni, D. B., and Lazarus, J. (1974). "Categorical and non-categorical modes of speech perception along the voicing continuum," *J. Acoust. Soc. Am.* **55**, 328-333.
- Pisoni, D. B., and Tash, J. (1974). "Reaction Times to Comparisons within and Across Phonetic Categories," *Percept. Psychophys.* **15**, 285-290.
- Plomp, R., Wagenaar, W., and Mimpen, A. (1973). "Musical Interval Recognition with Simultaneous Tones," *Acustica* **29**, 101-109.
- Pollack, I. (1952). "The Information in Auditory Displays," *J. Acoust. Soc. Am.* **24**, 745-749.
- Pollack, I. (1953). "The Information in Auditory Displays II," *J. Acoust. Soc. Am.* **25**, 765-769.
- Rakowski, A. (1976). "Tuning of Isolated Musical Intervals," *J. Acoust. Soc. Am.* **59**, S50(A).
- Roederer, J. (1973). *Introduction to the Physics and Psychophysics of Music* (Springer-Verlag, New York, Heidelberg, Berlin).
- Siegel, W., and Sopo, R. (1975). "Categorical Perception of Tonal Intervals," *Research Bulletin* 325 (Department of Psychology, University of Western Ontario, London, Canada) (unpublished).
- Spector, J. (1966). "Classical Ud Music in Egypt with Special Reference to Maqamat," *Ethnomusicology* **14**, 243-257.
- Stauffer, D. (1954). *Intonation Deficiencies of Wind Instruments in Ensemble* (Catholic University of America, Washington, D. C.) (unpublished).
- Stevens, S. S., and Guirao, M. (1963). "Subjective Scaling of Length and Area and the Matching of Length to Loudness and Brightness," *J. Exp. Psychol.* **66**, 177-186.
- Stevens, S. S., Volkman, J., and Newman, E. G. (1937). "A Scale for the Measurement of the Psychological Magnitude Pitch," *J. Acoust. Soc. Am.* **8**, 185-190.
- Strange, W. (1972). "The Effects of Training on the Perception of Synthetic Speech Sounds: Voice Onset Time," Ph.D. dissertation (University of Minnesota) (unpublished).
- Studdert-Kennedy, M., Liberman, A. M., and Stevens, K. (1963). "Reaction Time of Synthetic Stop Consonants and Vowels at Phonetic Centers and at Phonetic Boundaries," *J. Acoust. Soc. Am.* **35**, 1900.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K., and Cooper, F. S. (1970). "The Motor Theory of Speech Perception: A Reply to Lane's Critical Review," *Psychol. Rev.* **77**, 234-249.
- Terhardt, E. (1974). "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.* **55**, 1061-1069.
- Ward, W. D. (1954). "Subjective Musical Pitch," *J. Acoust. Soc. Am.* **26**, 369-380.
- Ward, W. D. (1963). "Absolute Pitch, Part II," *Sound* **2**, 33-41.
- Ward, W. D. (1970). "Musical Perception," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic New York).
- Watson, C., Kelly, W., and Wroton, H. (1976). "Factors in the discrimination of tonal patterns II. Selective attention and learning under various levels of stimulus uncertainty," *J. Acoust. Soc. Am.* **60**, 1176-1186.
- Wier, C., Jesteadt, W., and Green, D. M. (1975). "Comparison of method of adjustment and forced-choice procedures in frequency discriminations," *J. Acoust. Soc. Am.* **57**, S27(A).