Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*

JANE A. SIEGEL and WILLIAM SIEGEL

University of Western Ontario, London, Ontario N6A 5C2, Canada

Six musicians with relative pitch judged 13 tonal intervals in a magnitude estimation task. Stimuli were spaced in .2-semitone increments over a range of three standard musical categories (*fourth, tritone, fifth*). The judged magnitude of the intervals did not increase regularly with stimulus magnitude. Rather, the psychophysical functions showed three discrete steps corresponding to the musically defined intervals. Although all six subjects had identified in-tune intervals with >95% accuracy, all were very poor at differentiating within a musical category—they could not reliably tell “sharp” from “flat.” After the experiment, they judged 63% of the stimuli to be “in tune,” but in fact only 23% were musically accurate. In a subsequent labeling task, subjects produced identification functions with sharply defined boundaries between each of the three musical categories. Our results parallel those associated with the identification and scaling of speech sounds, and we interpret them as evidence for categorical perception of music.

It has become fashionable in recent years to make a dichotomy between speech and other auditory events, especially music. Speech, it is argued, is handled by a specialized neural processing mechanism localized in the left cerebral hemisphere (Darwin, 1969; Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967), while music is thought to be primarily a right-hemisphere function (e.g., Kimura, 1964; Milner, 1962; Shankweiler, 1966). It has been argued that the speech processor consists of specialized linguistic feature detectors that are “tuned” to the phonemic distinctions of language (Eimas & Corbit, 1973), and as a result, acoustic variations in the auditory stimulus irrelevant to meaning are filtered out. This gives rise to the phenomenon of *categorical perception*, a process whereby continuous acoustic variation is transformed into a discrete set of auditory events, the phonemes of a language system. It has been suggested that categorical perception is unique to speech (Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, Liberman, Harris & Cooper, 1970). Moreover, several intriguing experiments have provided evidence that categorical perception of speech is present at or shortly after birth in the same manner as in adults (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971).

It is now evident, however, that the original view is too simplistic—since there have recently been several convincing demonstrations of categorical perception for nonspeech dimensions (Cutting & Rosner, 1974; Miller,

Wier, Pastore, Kelly, & Dooling, 1976; Pastore, Friedman, Baffuto, & Fink, Note 1). In each case it has been argued that “natural” breaks in the continuum being investigated mediate the categorical perception effect, and it has been suggested (Cutting & Rosner, 1974) that the perceptual categories for speech may have evolved around such natural discontinuities. Such a hypothesis is supported by the fact that categorical processing of some speech continua is present in man at a very early age (e.g., Eimas et al., 1971) and also in sub-human species (Waters & Wilson, 1976). While cross-language studies suggest that experience is also important (Abramson & Lisker, 1970; Miyawaki, Strange, Verbrugge, Liberman, & Jenkins, 1975), the data are consistent with a hypothesis that assigns it a relatively minor role—perhaps one of sharpening or minimizing the effects of existing category boundaries (see Miller et al., 1976). According to this view, therefore, while categorical perception is not unique to speech, it may be restricted to certain “natural” categories.

An alternative hypothesis is that categorical perception may be acquired as a result of specific learning experiences, even when there are no “natural” sensory cues available to the perceiver. Lane (1965) claimed to provide support for this hypothesis in demonstrating the acquisition of categorical perception in the laboratory for several nonspeech dimensions. However, more recent studies have failed to replicate these findings (see Studdert-Kennedy et al., 1970), and this is consistent with the contention that learning alone cannot produce categorical perception. On the other hand, the training periods involved in these experiments have been relatively brief, and perhaps more extensive experience, comparable to that which most of us have had for speech, for example, would demonstrate the existence of acquired perceptual categories more conclusively.

Music provides an interesting opportunity to test this

Thanks are extended to Vickey Woods, Debbie Dorner, and Daphne Tracy for their help in running the experiments. Many individuals provided useful comments, including Peter Denny, Vincent di Lollo, Martin Taylor, and John Walsh. This research was supported by the National Research Council of Canada (Grants A7895 and A8581). Requests for reprints should be sent to Jane A. Siegel, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

hypothesis, for it is one of the few situations where a substantial number of people spend much of their lives encoding and decoding nonspeech auditory signals. The process whereby musicians learn to recognize meaningful elements in tonal stimuli is known as "ear training." In our culture, the well-tempered scale divides each octave into 12 logarithmically equal steps, so that the adjacent notes are related to one another by a *semi-tone*, a frequency ratio of 1.0595:1. Only a few musicians, those with *absolute pitch*, learn to identify the frequency of a single tone presented in isolation. Because melodies may be transposed into a different key, it is the *relationship* of the notes that is important in our system. Thus most musicians acquire *relative pitch*, the ability to identify the frequency ratio of two notes on an absolute basis. Each of the standard intervals is assigned a name, such as *unison*, *tritone*, *major third*, or *octave*, and "ear training" consists at least in part of learning to identify the standard intervals. The commonly used *Tonic Solfa* method, developed by Curwen in the 19th Century, teaches musicians to conceptualize the intervals as discrete entities, each with a unique quality, rather than as points along the continuum of pitch (see Helmholtz, 1954).

What is the effect of ear training on the perception of pitch by musicians? Certainly, identification ability improves drastically, as in one study we found that highly trained musicians could label intervals accurately, reliably, and independently of the stimulus context (Siegel & Siegel, 1977). Absolute, context-free identification is more usually associated with speech, and is thought to be one of the phenomena that makes speech "special." Second, musicians report hearing intervals as sounding qualitatively different from one another. A *major third* has a completely different sound from a *minor third* to a musician. It does not sound "larger," but rather is described as having a different character—much like the difference between two consonant phonemes. Finally, there is evidence that musicians may exhibit left-hemisphere processing of tonal stimuli (Bever & Chiarello, 1974), just as they do for speech. Non-musicians, on the other hand, seem to show right-hemisphere dominance for music (Kimura, 1964).

For musicians, then, there appear to be several points of similarity in the processing of speech and music, and this leads us to predict that musicians should show discrete, phoneme-like perception of tonal stimuli. Indeed, several recent experiments have provided some evidence for the categorical perception of notes by musicians with absolute pitch (Harris & J. Siegel, Note 2), and of interval and triad chords by musicians with relative pitch (Locke & Kellar, 1973; Burns & Ward, Note 3, Note 4; W. Siegel & Sopa, Note 5).

Here we attempt to provide converging evidence for the categorical perception of tonal intervals by testing a group of highly trained musicians in a common scaling task, the method of *magnitude estimation*. This paradigm has been used to scale a variety of stimulus attri-

butes, including pitch (cf. Marks, 1974). When non-musicians rate tone frequency in this task, the result is the well-known *mel scale*, where pitch is a monotonically increasing function of tone frequency.

Our research was motivated by an experiment by Vinegrad (1972), carried out at the Haskins Laboratories. As stimuli, Vinegrad used synthetic consonant-vowel syllables ranging from /be/ through /de/ to /ge/, synthetic vowels ranging from /i/ through /I/ to /e/, and tones varying in frequency. The results converged nicely with those typically found in discrimination experiments. For the consonants, there were definite plateaus in the magnitude estimation functions, where substantial physical change gave rise to little perceptual change. A similar, though less pronounced, effect occurred for the vowels, while for tones, the psychophysical function increased smoothly with tone frequency. Vinegrad's data give evidence for categorical perception of speech, and for continuous perception of pitch. His subjects, however, were nonmusicians. Here we asked highly trained musicians with relative pitch to make magnitude estimates of tonal intervals. We predicted that their magnitude estimation functions would be step-like, resembling those obtained by Vinegrad for speech.

METHOD

Apparatus

The experiment was run on-line using a DEC PDP-12 computer. Stimuli were sine-wave tones produced by a Wavetek Model 112 Generator, and were monitored by a Hewlett-Packard 5235B Universal Counter. The control of tone frequency was found to be accurate to within .1 Hz. Tones were delivered to the right ear of Clark/100 headphones at a comfortable listening level. Subjects entered their responses by pushing a set of appropriately labeled buttons, in the simple categorization tasks, or via a video terminal, in the magnitude estimation experiments.

Test of Relative Pitch Ability

Several tests were taken to ensure a subject pool with good relative pitch ability. First, a poster was placed at the Music School of the University of Western Ontario soliciting the participation of musicians with good relative pitch. Those who responded were required to identify a set of seven tonal intervals ranging from a *minor third* (a frequency ratio of 1.892:1) to a *major sixth* (1.6812:1) in random order.¹ Each interval consisted of two .25-sec tones separated by a .1-sec silent period. The first note of each interval was selected randomly from a set of seven musical notes in the range from a (261.6 Hz) to b¹ (493.8 Hz). All stimuli were musically accurate, according to the well-tempered scale. Subjects entered their responses on pushbuttons labeled with the appropriate interval names. The task was self-paced, no feedback was given, and each subject received 392 trials.

Subject pool. Of the 24 musicians who volunteered for the experiment, six met our criterion of >95% correct in the identification of musically accurate intervals, and this group was chosen for further testing. There were three males and three females, and none reported any hearing disabilities. Some relevant characteristics of the group are reported in Table 1, and it may be seen that all had been heavily involved with music over a number of years.

Magnitude Estimation Task

At the beginning of the experiment, the subjects were

Table 1
Musical Background of Six Subjects with Relative Pitch

Subject	Age	Age Began Formal Musical Training	Number of Instruments*	Hours of Participation in Music per Week**
C.M.	18	5	5	35
O.S.	23	7	6	60+
K.V.	20	6	2	40
A.S.	20	7	6	65
E.P.	20	4	12	51
T.F.	19	9	2	68

*Includes voice

**Includes listening, performing, course work

instructed as follows: "First, you will hear the same interval presented three times. This is the standard, and we would like you to assign it any number over 100. You will then be presented with additional tones for judgment. Listen to each pair carefully, and assign it a number which reflects the distance between the two tones. Make your judgments proportional to your judgment of the standard presented at the beginning. If you think the two tones are twice as far apart as the two tones comprising the standard, assign that pair a number twice as large as the one you gave the standard. If you think the interval is half as large as the standard, assign it a number half as large as the one you gave the standard. This task may be clearer to you if we work through an example using length. Suppose I presented you with a line that was 6 in. long as the standard and you called it 600. Now suppose I presented you with a 12-in. line? A 3-in. line? A 1½-in. line? You can use any number you wish, even numbers with decimals, such as 222.5. So don't feel restricted in terms of what numbers you can use. In each case, just try to use a number which accurately reflects what you hear."

In each of five blocks of trials, the subject heard the standard interval, an accurate *augmented fourth*, or *tritone*, three times. Then he judged 13 intervals, ranging from 20 cents (1/5 semitone) below a *perfect fourth* to 20 cents above a *perfect fifth*, in 20-cent steps. The first note of each interval was selected randomly from seven musical notes within a C major scale, G (196.0 Hz) to f (349.2 Hz). Responses were entered on the video terminal, and were erased at the end of each trial.

The duration of each tone was .5 sec, and the silent period between the two notes of each interval was .1 sec. The task was self-paced, and the next interval was presented 4 sec after the subject responded. There was a short rest period between blocks, and subjects filled out a questionnaire at the end of the session.

Labeling Task

The stimuli were identical to those used in the magnitude estimation task, but here the subjects were required to identify them using standard musical terminology. Again, there were 13 different intervals, ranging from a flat *fourth* to a sharp *fifth*, presented once each in random order in a block of trials. Subjects responded by pressing pushbuttons labeled *perfect fourth*, *tritone*, and *perfect fifth*. The task was self-paced, and the intervals were presented 4 sec after each response. Each subject received 20 blocks of trials in a single session of about 1 h.

RESULTS

Labeling Task

Figure 1 shows identification functions for each of the six subjects who had exceeded our criterion of 95% correct in the relative pitch test. Here we see the percentage of times each response category (*fourth*, *tritone*, *fifth*) was assigned to each of the 13 stimulus intervals. The naming distributions appear to be regular, symmetrical, with little overlap between adjacent categories. They are similar to those produced by musicians with

relative and absolute pitch that we have reported elsewhere, and quite different from the near-random data typically produced by nonmusicians (Siegel & Siegel, 1977).

It is interesting that, despite the fact that all of the subjects were able to identify musically accurate intervals virtually without error, there were still individual differences in the naming distributions shown here. In particular, several of the musicians were somewhat inconsistent in their use of the middle category, the *augmented fourth* or *tritone*. The latter is acknowledged to be one of the most difficult intervals to identify or to produce, probably because it occurs infrequently in Western music. It is much more common in Chinese and Japanese music, and one of our subjects reported that he recognized the *tritone* by "that tinny sound that you sometimes hear in Oriental music."

Our results provide further support for the view that

Figure 1. Interval identification functions for six musicians with relative pitch.

musicians can acquire absolutely anchored categories for pitch that function in a similar manner to the phonemic categories of speech (cf. Siegel & Siegel, 1977). They contradict the view that the ability to identify acoustic continua on an absolute basis is limited to speech (Liberman et al., 1967). Indeed, the labeling distributions generated by our musicians appear similar to those shown in a number of speech experiments (e.g., Liberman, 1970).

Magnitude Estimation Task

Our purpose was to seek evidence for the view that musicians categorize tonal stimuli *perceptually*, as well as verbally. In this experiment, subjects judged the distance between the two notes of each stimulus interval, with the stimuli being identical to those used in the labeling task. We predicted that, in a scaling experiment, musicians would be unable to differentiate different examples of the same musical category. Their magnitude estimation functions should be step-like.

Figure 2 shows the results of C.M., our "best" musician on several grounds. His naming distributions in the labeling task were more regular than those of any other subject, and in the magnitude estimation task, his variability of responding was the lowest. Since he, like all our musicians, used a constant response increment for each semitone, we were able to convert his responses to the corresponding musical values (in *cents*) for ease of interpretation.

Each panel of Figure 2 shows a single block of trials, and each point plotted is a single, unaveraged judgment. On each of the five trial blocks, C.M. produced a perfect step function with three steps. Although all 13 intervals in a block were acoustically different from one another, C.M. used only three numerical responses throughout the experiment.

C.M.'s data provide strong support for our hypothesis that musicians perceive tonal stimuli categorically in a manner similar to speech. At the same time, the possibility remains that the step functions he produced resulted from a simple response bias, rather than from categorical perception. That is, he may have, indeed, picked up fine within-category acoustic variations, but rounded off his responses to the value corresponding to the nearest accurate interval. We feel that the response-bias hypothesis is somewhat implausible, since subjects were instructed at the beginning of the experiment to make their judgments as fine as possible, and all seemed motivated by musical machismo to demonstrate their finely honed auditory acuity. However, we tested the response-bias theory by asking subjects to judge the proportion of "out-of-tune" intervals at the end of the experiment. C.M. guessed that perhaps one stimulus per block was musically inaccurate. In fact, 10 out of 13 in each block were out-of-tune. We conclude that the step functions produced by C.M. represent a genuine perceptual effect, not a simple response strategy. He apparently heard all examples of a musical category as

Figure 2. Individual judgments of interval size by subject C.M. over five blocks of trials. Magnitude estimates have been transformed to the appropriate musical values (in *cents*).

equivalent, as good examples of a standard musical interval.

Figure 3 shows the mean and standard deviation of C.M.'s judgments of each stimulus over the five trial blocks. There are two interesting phenomena. First, the averaged data are slightly less step-like than the unaveraged individual-block data. Second, the standard deviations show a cycling that is quite different from anything we have seen in magnitude estimation. Typically, judgmental variability increases with stimulus magnitude (cf. Galanter & Messick, 1961), but here the varia-

Figure 3. C.M.'s magnitude estimates averaged over the five trial blocks, with standard deviations, for 13 stimulus intervals.

bility peaks at each category boundary, and is low at the category center. This presumably results from the fact that intermediate stimuli are ambiguous, and may fall into one category or another. Our subjects report that the intermediate stimuli seem to "pop" one way or the other; they are not typically heard as halfway between two musical intervals. A similar phenomenon has been reported in the speech literature, where stimuli varying between, say, /b/ and /p/ are heard as one or the other, not as "in between."

The obscuring of the categorization effect by averaging can be explained by assuming that the probability of placing a stimulus into a higher category increases with stimulus magnitude. Thus it is possible to get evidence for continuous perception from averaged data even when the subject is unaware of intermediate stimuli and never uses intermediate responses. The evils of averaging discrete data have, of course, been pointed out in another context—the one-trial vs. continuous learning debate (e.g., Restle & Greeno, 1970).

C.M.'s data demonstrate one manifestation of perceptual categorization, a tendency to hear musically deviant stimuli as "in tune." On the average, the six subjects that we tested judged 37% of the intervals to be out of tune, whereas 77% were in fact inaccurate. A second manifestation of categorical perception may be seen in Figure 4, which shows individual trial data for O.S., who was the least variable of the remaining subjects, and whose naming distributions were similar to those of C.M. While C.M. had used only three different numbers to rate the stimuli on each block, O.S. used an average of 7.8 different numerical responses. Moreover, after the experiment, he guessed that 65% of the stimuli were out of tune. Yet his data are still clearly step-like. This comes about for two reasons: (a) a tendency to mini-

Figure 4. Individual trial data for subject O.S. in the magnitude estimation task.

Figure 5. Means and standard deviations for O.S. in the magnitude estimation task.

mize deviations from the musical prototype, and (b) an inability to differentiate stimuli within a musical category. The latter is illustrated in Block 2, where he calls a flat *tritone* "sharp"; on Block 3, where he calls the sharpest *fourth* "flat"; on Block 4, where he calls the sharpest *fourth* "flat," the flattest *tritone* "sharp" and so forth. This is especially interesting because it provides strong evidence against the response-bias interpretation of our data. O.S. was using many more responses than C.M., and he reported hearing many more out of tune intervals, but he was unable to accurately differentiate sharp from flat. Nor were any of the other subjects that we tested here.

Figure 5 shows O.S.'s data averaged over the five blocks of trials, together with the standard deviation for each stimulus. It may be seen that the average magnitude estimation function has three plateaus corresponding to the standard musical intervals, where substantial changes in the frequency ratio of the stimulus were not encoded accurately. Also, his standard deviations cycled in the same manner as those of C.M., with low variability at the category center and high variability at the category boundary.

The same pattern of results was observed to some degree for the remaining four subjects, whose data are shown (in order of increasing overall variability) in Figure 6. Remember that all of these subjects were able to identify standard musical intervals virtually without error. One interesting phenomenon shown here is their inability to differentiate among members of the same musical category, to tell "sharp" from "flat." Even more striking is a tendency toward *categorical errors*, i.e., hearing an intermediate stimulus as an example of the

Figure 6. Individual trial data, means, and standard deviations for the remaining four subjects, shown in order of overall response variability.

wrong category. A close examination of Figure 6 reveals that most of the large deviations from a step function were in fact categorical errors. Thus, these data, while not as visually pleasing as those of the most reliable subjects, provide further support for the categorical perception hypothesis. Furthermore, while their averaged responses were less step-like than those of the best subjects, their standard deviations showed the same pattern of peaks and troughs as those of C.M. and O.S.

To shed further light on our musician's ability to judge intervals, we carried out more detailed analyses of their responses in the magnitude estimation task. Table 2 presents a confusion matrix in which both stimuli and responses are classified as "flat," "in tune," or "sharp." The cells contain the conditional proportion of each response class given the stimulus class. Only those trials where the subject identified the interval correctly are included here. As a result, sharp-flat confusions are genuine reversals, and do not include any cases where, say, a sharp *tritone* is judged to be a flat *fifth*.

Several interesting conclusions may be drawn from

Table 2. First, the illusion of categorical perception is shown by a tendency to call the stimuli "in tune" roughly 80% of the time, regardless of whether the intervals are flat, sharp, or in tune. Second, there were a number of sharp-flat reversals, especially in the case of the sharp stimuli, which were nearly as likely to be judged "flat" as "sharp." The only evidence for differentiation was for the flat intervals, but even these were correctly identified as flat only 28% of the time.

Finally, we carried out additional analyses of the 102

Table 2
Proportion of Flat, In-Tune, and Sharp Responses Given to Flat, In-Tune, and Sharp Stimuli

Stimuli	Responses		
	Flat	In Tune	Sharp
Flat (N = 121)	.28	.68	.04
In Tune (N = 85)	.08	.86	.06
Sharp (N = 132)	.08	.82	.10

Note—Excluding categorical errors.

"out-of-tune" responses made in the experiment. Of these, only 9% were completely accurate. Twenty-eight percent were categorical errors, and 12% were false alarms, calling a musically accurate interval "out-of-tune." The remaining 51% were partially correct, in that subjects placed the interval into the appropriate category and recognized it as being out of tune. However, of these partially correct judgments, subjects erred in the direction of categorical perception 90% of the time, in that 29% were sharp-flat confusions, and 61% were judgments that were too close to the musical norm.

Comparison of Magnitude Estimation and Identification Data

Further evidence for categorical processing may be obtained by directly comparing the results from the identification and magnitude estimation tasks. The question is whether the discontinuities in the magnitude estimation data correspond to the location of the category boundaries as inferred from each subject's identification performance. That is, are subjects more likely to judge two intervals as perceptually distinct when they lie on opposite sides of a category boundary?

By referring to the identification data of each subject, interval pairs were classified according to whether they were both in the same musical category or on opposite sides of a category boundary. All possible one-step comparisons (intervals that were 20 cents apart) and two-step comparisons (intervals that were 40 cents apart) were considered, and category boundaries were inferred from the cross-over point of each subject's identification function. Figure 7 compares, for each subject, the judged perceptual distance between intervals within a category and on opposite sides of a category boundary. The figure shows that intervals belonging to different categories are judged to be further apart than those belonging to the same category, even when the physical distance between them is held constant. Four of the six subjects show this effect for the one-step comparisons, and all six subjects for the two-step comparisons. Moreover, those subjects who show the effect most

strongly are those with the cleanest identification functions (see Figure 1).

Finally, we examined the reliability with which the subjects differentiated the two types of intervals across the five blocks of trials, using a series of *t* tests for matched pairs. Over 90% of the statistically significant comparisons ($p < .05$) involved intervals on opposite sides of a category boundary. Stimuli in different categories, therefore, are not only judged as more distinct, they are also discriminated more reliably from one another.

In general, then, one can conclude that highly trained musicians with excellent relative pitch typically have learned just enough about tonal intervals to differentiate the musically relevant cues. Their ability to differentiate different examples of the same musical interval is extremely limited. Most were unaware of how poorly they were doing during the course of the experiment, and considerable tact and compassion were required when we debriefed the subjects.

DISCUSSION

One can draw the following conclusions from the present experiment:

(1) Musicians with good relative pitch can label tonal intervals accurately on an absolute basis. Their naming distributions are symmetrical, with well-defined category boundaries and little overlap between adjacent categories. As such, their performance resembles that usually found with speech (see Siegel & Siegel, 1977, for a replication and extension to absolute pitch).

(2) Corresponding to the musically defined interval categories are *perceptual* categories. In the magnitude estimation task, musicians had a strong tendency to rate out-of-tune stimuli as in tune. Their attempts to make fine, within-category judgments were highly inaccurate and unreliable. That is, they could not tell "sharp" from "flat." This finding, while perhaps counter intuitive, replicates our preliminary magnitude estimation results with possessors of relative and absolute pitch (Harris & J. Siegel, Note 2; W. Siegel & Sopo, Note 5). Moreover, it converges with the findings of Burns and Ward (Note 3, Note 4), who reported that within-category discrimination of intervals was at close to chance level in a forced-choice task.

(3) All six subjects showed evidence for categorization, but the effect was strongest for those with the best relative pitch, as assessed in the labeling task. Subjects who performed relatively poorly in the labeling task made more categorical errors in magnitude estimation. This obscured the step-like character of the data and led to artifactual, intermediate responses when the results were averaged.

(4) A postexperimental questionnaire supported the conclusion that the magnitude estimation data represented the musicians' subjective experience, not some arbitrary response bias, in that subjects greatly

Figure 7. Average perceptual distance for within-category and between-category comparisons.

underestimated the number of out-of-tune stimuli in the experiment.

Categorical perception has been previously demonstrated for nonspeech stimuli, but it has been argued in these cases that it is based on "natural" cues—that is, that it is innately determined. The data presented here, however, suggest that training and experience alone may be sufficient for the development of categorical perception. While the semitone is the basic unit for Western music, other cultures have widely different scale forms. Moreover, nonmusicians have great difficulty in even identifying tonal intervals on an absolute basis, and real proficiency seems to appear only with relatively extensive musical training (Siegel & Siegel, 1977). In fact, there were substantial individual differences in identification ability, even in the present experiment, in which the six musicians tested were selected from a much larger sample on the basis of their good relative pitch.

While we expected that our musicians would show some evidence of perceptual categorization, we were somewhat surprised to find that their level of within-category discrimination was so poor. We had expected that at least some of our subjects would be able to "fine-tune," and that their performance with intervals might parallel results that are typically obtained with vowels in the speech literature. Vowel continua may be reliably partitioned into categories, but unlike consonants, discrimination within a category is much better than would be predicted from identification data (Pisoni, 1973). As a result, their perception has been described as "continuous," rather than "categorical" (see Studdert-Kennedy et al., 1970). In fact, however, none of our musicians were able to make reliable within-category discriminations, and those who produced more evidence for "continuous" perception were those with the least reliable identification functions. That is, they exhibited poor between-category discrimination rather than good within-category discrimination.

In speech, the level of discrimination within a category appears to be influenced by a number of variables. For example, while the perception of vowels is typically more continuous than the perception of consonants, categorical processing may be enhanced by manipulating such variables as the memory load of the task (Pisoni, 1973, 1975). Within-category discrimination of consonants, on the other hand, seems resistant to the effects of this variable (Pisoni, 1973). These data can be encompassed by a model of speech perception in which judgments are made on the basis of both phonetic and acoustic information (Fujisaki & Kawashima, 1970; Studdert-Kennedy, 1973). Differences in the processing of consonants and vowels are explained by the fact that, while phonetic information is available for both kinds of stimuli, acoustic information is more readily available for vowels. Acoustic information decays rapidly from short-term memory, however, and so the perception of vowels may be made to appear more or less categorical, depending upon the memory load in the task.

Applying a similar model to music perception, one can ask if the perception of tonal intervals by musicians is more like vowels or consonants. In the present experiment, our musicians exhibited more consonant-like data, in that their discrimination of within-category differences was uniformly poor. According to the model, however, this result is not entirely unexpected, since our magnitude estimation task tests subjects' discrimination ability under conditions of high memory load in which one would expect a rapidly decaying acoustic trace to be relatively unavailable (see W. Siegel, 1972; Siegel & Siegel, 1972). Similar data, however, have been obtained by Burns and Ward (Note 4) using a forced-choice discrimination procedure with a minimal retention interval. It appears therefore that categorical processing of intervals is quite robust and more akin to categorization effects observed with consonants than with vowels.

A question which remains is why musicians' ability to differentiate stimuli from the same category is so poor. Musicians pride themselves in having a "fine ear," and a number of tests of musical aptitude, including the well-known Seashore Measures of Musical Talents, include tests of pitch discrimination (see Lehman, 1968). Why is it, then, that musicians are so poor at telling "sharp" from "flat"? A possible clue may come from the speech literature, which led us in the first place to predict categorical perception in the musical domain. The discovery of categorical perception of speech was made on the basis of spectrographic studies of natural speech where it was found that the acoustic correlate of a given phoneme is typically highly variable. A variety of quite different acoustic events are all heard as the same sound, a /d/, for example (see Liberman et al., 1967). Thus, categorical perception appears to be an effective means of transforming a messy acoustic signal into a small set of meaningful units.

Is it possible that the acoustic correlate of music is as variable as speech? Common sense suggests the opposite, that good musical performance is characterized at the very least by an accurate rendition of the notes in the written score. In fact, acoustic measurements of performances by well-known artists indicate a high degree of variability, similar to that found in speech. It is only because of the illusion of categorical perception that we are largely unaware of the gross pitch deviations that are the norm in musical performance. Winckel (1967) quotes von Hornbostel's statement that Europeans sing "terribly inaccurately," as well as Abraham's finding of "enormous deviation of intonation" even in very good singers. He quotes L. Euler as stating in 1764 that "the ear usually hears what it wants to hear, even if that does not correspond to the acoustically given interval (Winckel, 1967, p. 128)." Seashore and his colleagues made extensive acoustic measurements of the well-known artists of his time, and his findings supported Euler's view quite well. He concluded that: "It is shockingly evident that the musical ear which hears the tones indicated by the conventional notes is extremely

generous and operates in the interpretative mood. Compare this principle for the various singers, and you will see that the hearing of pitch is largely a matter of conceptual hearing in terms of conventional intervals" (Seashore, 1967, p. 269).

The statements of Euler and Seashore are supported nicely by our experimental findings, although we would be inclined to look further inside the head for the locus of the effect. In music, categorical perception allows one to recognize the melody, even when the notes are out of tune, and to be blissfully unaware of the poor intonation that is characteristic of good musical performance.

REFERENCE NOTES

1. Pastore, R. E., Friedman, C., Baffuto, K. J., & Fink, J. A. *Categorical perception of both simple auditory and visual stimuli*. Paper presented at the Acoustical Society of America, Washington, D.C., 1976.
2. Harris, G., & Siegel, J. A. *Categorical perception and absolute pitch*. Paper presented at the Acoustical Society of America, Austin, Texas, 1975.
3. Burns, E., & Ward, W. D. *Categorical perception of musical intervals*. Paper presented at the Acoustical Society of America, Los Angeles, California, 1973.
4. Burns, E., & Ward, W. D. *Further studies in musical interval perception*. Paper presented at the Acoustical Society of America, San Francisco, California, 1975.
5. Siegel, W., & Sopo, R. *Tonal intervals are perceived categorically by musicians with relative pitch*. Paper presented at the Acoustical Society of America, Austin, Texas, 1975.

REFERENCES

- ABRAMSON, A. S., & LISKER, L. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*. Prague: Academia, 1970. Pp. 569-573.
- BEVER, T., & CHIARELLO, R. J. Cerebral dominance in musicians and nonmusicians. *Science*, 1974, **185**, 537-539.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, **16**, 564-570.
- DARWIN, C. J. Laterality effects in the recall of steady-state and transient speech sounds. *Journal of the Acoustical Society of America*, 1969, **46**, 114.
- EIMAS, P. D., & CORBIT, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, **4**, 99-109.
- EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P., & VIGORITO, J. Speech perception in infants. *Science*, 1971, **171**, 303-306.
- FUJISAKI, H., & KAWASHIMA, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute*, 1970, **29**, 207-214.
- GALANTER, E., & MESSICK, S. The relation between category and magnitude scales of loudness. *Psychological Review*, 1961, **68**, 363-372.
- HELMHOLTZ, H. *On the sensations of tone*. New York: Dover, 1954.
- KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, **15**, 166-171.
- KIMURA, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 355-358.
- LANE, H. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, **72**, 275-309.
- LEHMAN, P. R. *Tests and measurements in music*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.
- LIBERMAN, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Perception and its disorders, Proceedings of A.R.N.M.D.* Baltimore: Williams and Wilkins, 1970.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. S., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.
- LOCKE, S., & KELLAR, L. Categorical perception in a non-linguistic mode. *Cortex*, 1973, **9**, 355-368.
- MARKS, L. E. *Sensory processes: The new psychophysics*. New York: Academic Press, 1974.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. A. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, **60**, 410-417.
- MILNER, B. Laterality effects in audition. In E. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins University Press, 1962.
- MIYAWAKI, K., STRANGE, W., VERBRUGGE, R., LIBERMAN, A. M., JENKINS, J. J., & FUJIMURA, O. An effect of linguistic experience: The discrimination of (n) and (l) by native speakers of Japanese and English. *Perception & Psychophysics*, 1975, **18**, 331-340.
- PISONI, D. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, **13**, 253-260.
- PISONI, D. Auditory short-term memory and vowel perception. *Memory & Cognition*, 1975, **3**, 7-18.
- RESTLE, F., & GREENO, J. G. *Introduction to mathematical psychology*. Reading, Mass: Addison-Wesley, 1970.
- SEASHORE, C. E. *Psychology of music*. New York: Dover, 1967.
- SHANKWEILER, D. P. Effects of temporal lobe damage on perception of dichotically presented melodies. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 115-119.
- SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to the left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 59-63.
- SIEGEL, J. A., & SIEGEL, W. Absolute judgment and paired-associate learning: Kissing cousins or identical twins? *Psychological Review*, 1972, **79**, 300-316.
- SIEGEL, J. A., & SIEGEL, W. Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, 1977, **21**, 143-152.
- SIEGEL, W. Memory effects in the method of absolute judgment. *Journal of Experimental Psychology*, 1972, **94**, 121-131.
- STUDDERT-KENNEDY, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. XII). The Hague: Mouton, 1973.
- STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K., & COOPER, F. S. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, **77**, 234-249.
- VINEGRAD, M. D. A direct magnitude scaling method to investigate categorical vs. continuous modes of speech perception. *Language and Speech*, 1972, **15**, 114-121.
- WATERS, R. S., & WILSON, W. A. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, 1976, **19**, 285-289.
- WINCKEL, F. *Music, sound and sensation: A modern exposition*. New York: Dover, 1967.

NOTE

1. In fact, the subjects were tested first in the magnitude estimation task, second in the relative pitch test, and third in the labeling task. We ran them first in the magnitude estimation experiment so that they would not be biased to label, rather than judge, the stimulus intervals in that experiment.

(Received for publication December 20, 1976;
accepted for publication January 18, 1977.)