*Psychological Review, 1969, **76**, 300—307*

MUSIC RECOGNITION

DIANA DEUTSCH

University of California, San Diego

Music recognition is discussed, and it is argued that this involves certain specific processes of abstraction. A network is proposed which can perform these abstractions.

It is remarkable that despite considerable speculation and experimentation in the field of visual shape recognition, psychologists have in the last decade taken very little interest in the problem of music recognition. Yet the ability to recognize music is quite distinct from simple pitch perception or discrimination. It requires a mechanism which can abstract the relational properties existing in tonal combinations. Some people have a surprising inability to do this. Wing (1948) describes a boy who had the finest pitch discrimination out of 70, but yet could not master a simple melody. Trotter (1967) reports a similar case.

This paper is concerned with the perception of combinations of tones, both simultaneous and successive. Rhythm is not dealt with. An attempt is first made to specify clearly the types of abstraction involved when we perceive musical combinations. A neural network is also proposed which can perform the specified abstractions and which therefore constitutes a theory of the way we ourselves perform them.

ABSTRACTIONS OCCURRING IN MUSIC RECOGNITION

Simultaneous Interval Recognition and Transposition

The unit of the musical scale as conventionally defined is the semitone. When two tones are played simultaneously we are able to tell how many semitones they are apart independently of the absolute pitch of the component tones. Thus, two tones three semitones apart constitute a minor third, four semitones apart constitute a major third, and seven semitones constitute a fifth. In physical terms, since the musical scale of pitch stands in logarithmic relationship to the frequency scale, intervals appear identical if they stand in the same physical ratios to each other. An octave is produced by the physical ratio 1 : 2, a fifth by the ratio 2 : 3, and so on. Thus equal frequency ratios appear as equal intervals.

Chord Recognition and Transposition

Just as with intervals, we are able to classify simultaneous combinations of more than two tones, independently of their component frequencies. For instance a combination of three tones of which the lowest and middle are four semitones apart, and the middle and the highest three semitones apart is known as the root form of the major triad.

It might be thought that chord recognition follows logically from interval recognition; however, it can be shown that abstraction of all the intervals involved in a chord does not uniquely define the chord. Consider, for instance, the root forms of the major and minor triads. As shown in Figure 1 they are both composed of a major third, a minor third, and a fifth. However, they sound quite different. Thus the problem of how we abstract chords is not solved by the knowledge of how we abstract intervals.

Tune Recognition and Transposition

A tune is a series of successive intervals. When two tones are played successively, we are able to tell how far apart they are from each other on the musical scale, independently of their absolute pitch (in physical terms, we abstract their ratios independently of their absolute frequencies). This process of abstraction is so fundamental that it is easier for us to remember a tune than it is to remember the absolute pitch of the component tones. We transpose tunes so readily that it is extremely difficult for us not to do so.

Tune recognition is more complicated than simultaneous interval abstraction because it involves recognition of the time order in which the two components of the interval occurred. For instance all three intervals in Figure 2 sound different although their component frequencies are identical.

Recognition and Transposition of Sequences of Chords

It follows from the transposability of harmonic sequences that we abstract successive relationships not only between tones played separately, but also between simultaneous combinations of tones. Such successive relationships may however be defined by specifying all the abstracted successive intervals involved. For illustra-

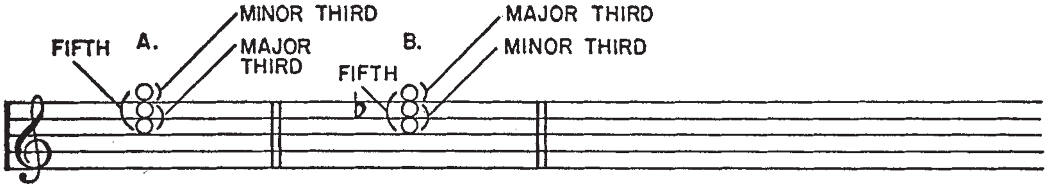


Fig. 1

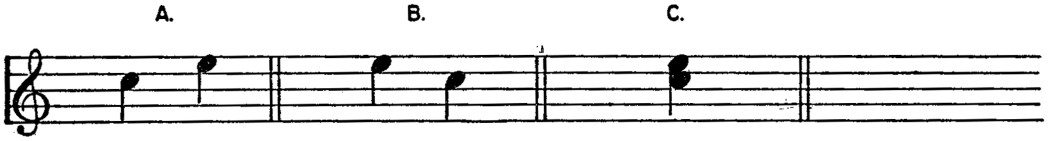
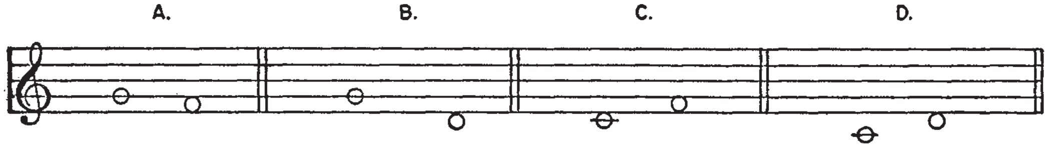



Fig. 2

tion, let us take the simplest case of such a sequence: that of two successive two-note chords.

The component successive intervals involved are:



Thus this succession is uniquely defined as composed of (a) a descending second, (b) a descending fourth, (c) an ascending fourth, and (d) an ascending second. In this way we have abstracted the sequence and so made it transposable.

It might appear here that confusions would be introduced by the definition of chord successions in terms of intervals, as shown above in the case of chords played in isolation. However, there is an essential difference here. More degrees of freedom are present in the sequential situation since the successive intervals may be either ascending or descending. This extra parameter allows the intervals to define the chord sequence absolutely.

Similarity between Tones Standing in the Physical Ratio of a Power of 2:1

It has long been recognized that tones separated by octaves or where frequencies stand in the ratio of a power of 2 : 1 have an essential similarity. In fact, so strong is this similarity that the musical scale is based on it. Tones which have this relationship are given the same name (e.g., C, D, etc.) . Generalization studies both on rat (Blackwell & Schlosberg, 1943) and man (Humphreys, 1939) show that responses to test stimuli increase suddenly at points corresponding to octaves.

Inversion of Chords

By inversion is meant a change in the order in which the components of a chord appear. In Figure 3 is shown the root, the first inversion, and the second inversion of the C major triad.

Inversions of chords are treated by musicians as harmonically equivalent to their root forms, and it is striking that such inversions sound essentially similar, although intervals of different sizes are involved. This similarity between inverted chords is probably related to the similarity between tones separated by octaves. An operation which treats all Cs, Ds, etc., as equivalent could easily be extended to treat any combination of Cs, Ds, etc., as equivalent. However, intervals are not invertable, and neither are tunes. So the suggested extension would have to be specific to combinations of three or more tones.

SUMMARY OF OPERATIONS NEEDED TO PERFORM MUSICAL ABSTRACTIONS

1. Interval abstraction (a) simultaneous, (b) successive ascending, and (c) successive descending.
2. Abstraction of chords consisting of at least three notes. (As explained heretofore, the abstraction of simul-

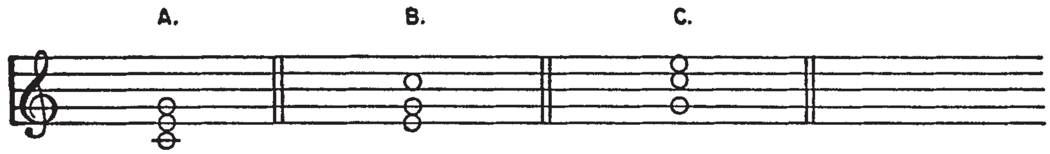


FIG. 3. (a) Root, (b) first inversion, and (c) second inversion of C major triad.

taneous intervals is not a sufficient condition for the abstraction of three-note chords.)

3. Identification together of tones separated by octaves, or standing in the physical ratio of a power of 2: 1. This principle is extended to simultaneous combinations of at least three tones (but not two).

PROPOSALS CONCERNING MUSIC RECOGNITION

Pitts and McCulloch (1947) suggest a mechanism for chord transposition. Basically, this assumes that the auditory cortex consists of layers of neurons, each layer containing a topographical projection of frequency-specific neurons, and arranged so that columns of neurons responding to particular frequencies are formed. In this topographical projection, equal intervals are spaced equal distances apart. Pitts and McCulloch suggest that there are in addition fibers which traverse this columnar mass parallel to each other in a slantwise direction. Thus, three such slanting pathways would determine a three-note chord. This mechanism, however, cannot explain tune, that is, successive interval recognition; neither can it explain inversion of chords, nor the recognition of sequences of chords.

Boomslitter and Creel (1961) suggest a type of telephone theory for music recognition. When two combinations of frequencies stand in the same ratios to each other the pattern generated by these two combinations is the same. However, such an explanation is unlikely to be correct, in view of the neurophysiological evidence. Although at the level of the auditory nerve the summated response of the fibers can follow the stimulus frequency up to 5,000 cycles per second (cps); this upper limit drops drastically as the auditory pathway is ascended, probably to about 800 cps. Further, since the hypothesized pattern would be based on a superposition of the stimulus frequencies, the frequency of following necessary for such a pattern to be resolved would be considerably greater than that necessary to perceive the component frequencies. Another reason against such a pattern theory is that a unique pattern only occurs with simultaneous intervals and not successive intervals, and so it cannot explain tune recognition. It also cannot explain inversion of chords, nor the recognition of sequences of chords.

NEUROPHYSIOLOGICAL STUDIES CONCERNING ABSTRACTION OF SENSORY INFORMATION

It is desirable that a theory of how the nervous system abstracts information should be plausible in the face of related neurophysiological evidence. Single unit studies involving auditory abstractions of the type discussed in this paper have not been performed. However, studies on the response of single units to higher order visual information should be considered here since it is plausible to suppose that higher order sensory processing in vision and audition take place along similar lines.

Recent single unit studies have indicated that the processing of visual information occurs in a number of stages each of which produces a small degree of abstraction. According to Hubel and Wiesel (1962) units with circular receptive fields are joined in groups to higher order units in such a way that units whose receptive fields taken together from straight lines are joined together. These higher order units thus respond specifically to straight lines presented in a particular portion of the receptive field and having a particular orientation. A further degree of abstraction then occurs as these higher order units are joined to further order units in such a way that those sensitive to lines within a certain part of the receptive field and which all have the same orientation are linked together. Thus some degree of abstraction of tilt information is obtained at this level. These stages of abstraction thus seem to be simple linkages of certain types of neurons together. The system to be outlined was devised with the visual studies in mind, and so it too is based on simple linkages of neurons.

A SUGGESTED MECHANISM

A neural mechanism is here described briefly which can perform the abstractions discussed in the first part of the paper. Basically, it consists of two parallel channels each of which has two stages, as shown on Figure 4.

It is well known that impulses generated by tonal stimuli ascend to the primary cortical receiving area for audition. Here there is a topographical distribution of frequency-sensitive neurons.

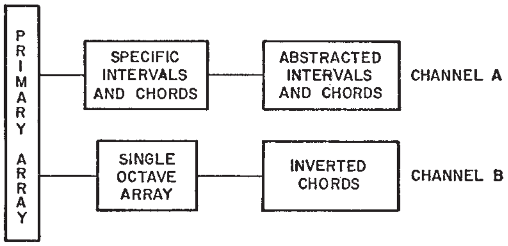


FIG. 4. Flow diagram for music recognition.

In the first stage of transformation on Channel A, primary neurons feed in twos and threes on to second-order neurons. These second-order neurons thus respond to specific intervals and chords. In the second stage of transformation second-order neurons are linked to third-order neurons in such a way that all units activated by seconds feed into one unit, all those activated by thirds into another, those activated by a particular triad into another, and so on. Thus the third-order neurons respond to abstracted intervals and chords (see Figure 5).

The second-order neurons on Channel A which respond to combinations of two tones fall into three categories. Those belonging to the first category are excited immediately by inputs from both primary neurons. Their output thresholds are set so that excitation has to arrive from both pathways to produce output. Such neurons would be sensitive to simultaneous intervals. Neurons belonging to the second and third categories are excited by one primary neuron and inhibited by the other. When the inhibitory pathway ceases to transmit there is a postinhibitory rebound. If this occurs at the same time as excitation from the first pathway appears these two sources of excitation will sum to produce an output. Thus such a neuron will produce an output only when a successive interval occurs. These second- and third-category neurons are reciprocally arranged so that each is excited from the primary neuron through which the other is inhibited. Thus one will respond only to the ascending successive interval, and the other only to the descending successive interval (Figure 2). It is further assumed that these pathways are activated only by tonal onset, and so the postinhibitory rebound occurs even when the initially inhibitory tone continues to sound.

In the first stage of transformation on Channel B, the primary frequency-specific neurons are linked in such a way that all neurons separated by octaves are joined to the same second-order neuron (see Figure 6). In the second stage of transformation these second-order neurons are joined in groups of three to third-order neurons. However there is no linking of combinations of two here. Thus inversion of three-note chords is produced, but intervals and tunes are not inverted.

This network would thus produce abstraction of simultaneous and successive intervals, and simultaneous chords; and of sequences of simultaneous chords, and also inversion of chords.

AUDITORY NEURAL PATHWAYS

Studies on neural pathways in the auditory system have very largely been concerned only with responses to the simplest type of auditory stimulation that is, pure tones and clicks. The main stimulus parameters investigated have been the characteristic frequency of a neuron and its response area. (These are obtained by finding, for each of a number of stimulus frequencies, the lowest intensity which will evoke spike discharges.) In contrast to studies on the visual system, there has been very little attention paid to the response of neurons to combinations of tones (Bishop, 1967).

The auditory cortex investigated in this fashion appears to have a surprisingly large number of unresponsive units. As Goldberg and Lavine (1968) point out: "The apparent unresponsiveness of neurons in the auditory cortex remains the most perplexing problem in auditory physiology [p. 331]." And they continue: "It is, of course, possible that many neurons in the auditory cortex do not respond in a simple manner to any auditory stimulus [p. 331]." This would hardly be surprising, since animals in their natural environment are much more concerned with auditory pattern recognition than with pure tones.

Observations on the deficit of function following various cortical ablations also lead us to conclude that the auditory cortex is not concerned with processing the most simple types of auditory information. Cats without auditory cortex are able to discriminate the onset of a

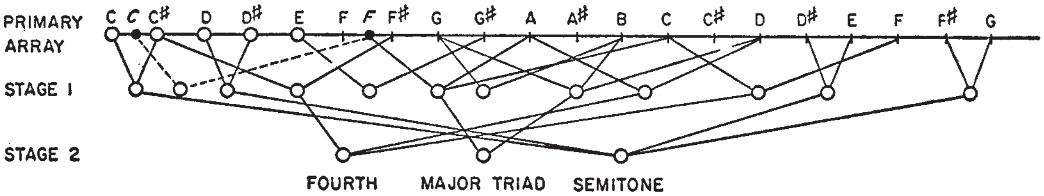


FIG. 5. Two stages of abstraction of interval and chord information. (Although the neurons in the primary array are identified for purposes of clarity in musical notation it is obviously assumed that intervening neurons also exist, and are linked up in the same way. These are represented by *C* and *F* in the diagram..)

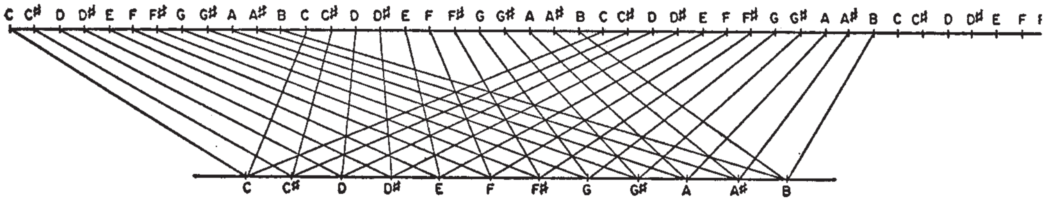


FIG. 6. Telescoping of frequency-specific neurons into a single octave array. (As in Figure 5, it is obviously assumed that neurons responding to intervening frequencies not included in the diagram are also linked in this fashion.)

sound (Kryter & Ades, 1943; Meyer & Woolsey, 1956; Stoughton & Neff, 1950); changes in sound intensity (Oesterreich & Neff, 1960, Raab & Ades, 1946; Rosenzweig, 1946); and changes in tonal frequency (Butler, Diamond, & Neff, 1957; Goldberg & Neff, 1961; Meyer & Woolsey, 1956). Such discriminations are still possible after bilateral cortical ablation including at least Al, A2, Ep, and the insular-temporal region. However, extirpation of either Al, A2 and Ep (Diamond & Neff, 1957), or the insular-temporal region (Goldberg, Diamond, & Neff, 1957) leads to a profound loss in ability to discriminate tonal patterns (i.e., the same tones arranged in a different order). Also Dewson (1964) reports that cats with bilateral ablation of the ventral insular-temporal cortex are unable to discriminate between the speech sounds “u” and “i.” These findings suggest that a proportion of auditory cortical neurons are involved in the processing of tonal combinations. Yet in attempting to study the behavior of such neurons we are faced with an overwhelmingly large number of such combinations from which to choose. While it is apparent that the visual system must process lines and angles, the choice of tonal combinations to use as stimuli in an analogous investigation depends on a theory of auditory pattern perception. It is hoped that the system proposed here can serve as such a framework.

This suggestion rests, of course, on the assumption that the animal, studied neurophysiologically, processes auditory information in a fashion similar to man. Ideally, a behavioral analysis should first be carried out to determine if this is indeed so, and, if not, how the animal does abstract auditory information. Yet the illuminating studies of Hubel and Wiesel (1962) on the visual system of the cat were carried out without such a behavioral analysis; and it seems not unreasonable to suppose, as a first guess, that auditory mechanisms for man and the cat are similar.

Recently, investigators have started to search the auditory cortex for neurons responding in a more complex fashion to auditory stimuli. Oonishi and Katsuki (1965) describe several types of auditory cortical neuron in the cat. One type, which had been reported previously, and is also found in the medial geniculate body, has a short latency and a response area with one sharp peak. Two medium latency neurons were also described. One has a high threshold and irregular response area, and the other has a response area with several peaks; both of these were facilitated by two-tone stimulation with specific frequencies. A long-latency type of neuron with flat response area was also described. On the basis of latency studies, Oonishi and Katsuki suggest that the sharp peak-type neurons project to the multipeak type, which in turn project to the flat type; a functional organization which, as the authors point out, is similar to that found by Hubel and Wiesel (1962) in the visual cortex. Although we cannot draw more than a superficial analogy between the behavior of the neurons described here and the response characteristics predicted on the proposed system, the multipeak- and irregular-type neurons could be involved in the first stage of transformation on both Channels A and B, and the flat-type neurons in the second stage. (It should be pointed out that the frequency separation between the peaks in the multi-peak type is larger than would be expected on the proposed system, though not incompatible with it.) However, this suggestion is only tentative, since the theory can only be investigated properly with the use of specific tonal combinations.

Whitfield and Evans (1965) have studied the response of neurons in the auditory cortex of the cat to changing frequency parameters. They found that a proportion of such units which did not respond to steady tones did respond to frequency-modulated stimuli. Some units responded only to an increasing frequency, others only to a decreasing frequency (and yet others had more complex characteristics). Suga (1965) found neurons in the auditory cortex of bats which also responded specifically to a change in tonal frequency either up or down. A change in the opposite direction did not excite this category of neuron. It is tempting to draw an analogy between the behavior of these neurons and those processing successive intervals suggested above. Yet here again, the relevant stimulus parameters were not explicitly employed, so the analogy can only be tentative.

In discussing the processing of successive intervals, the system requires a unit to be excited by one frequency and inhibited by another. This has been found at several levels of the auditory pathway including the cortex (Evans & Whitfield, 1964).

The neurophysiological evidence thus far certainly suggests the presence of units with complex properties. So far the evidence is quite consistent with the hypothesized mechanism. However, such a theoretical structure must still be put to a more direct test by the use of stimulus combinations more pertinent to its assumptions.

DISCUSSION

It must not be considered that this scheme presupposes an innate disposition to group tones rigidly according to the quanta employed in traditional music. For purposes of clarity the diagrams here have been drawn showing mainly the tones and intervals traditionally used. However, it is assumed that units responding to other tones and combinations also exist. These are represented in Figure 5 by units *C* and *F* and their linkages. The number of tonal quanta and combinations involved is a matter to be experimentally determined.

The question also arises as to whether our grouping of intervals and triads into categories such as thirds, fourths, etc., is due to an innate linkage or whether it occurs through experience. This is similar to the question of whether visual shape recognition is innate or learned. The mechanism here proposed does not depend on any one interpretation. For if such abstractions are performed, whether this is brought about through learning or through an innate process, the neural network underlying their performance must at all events be present at the time of their performance. Thus the scheme outlined could either be suggested to be independent of experience, or the relevant neurons could be connected by experience.

There is little doubt that experience with music makes us more aware of the categorizations outlined above. In a recent experiment, Allen (1967) used two groups of subjects: students with musical training and those without. He asked both groups to rate pairs of tones on a similarity criterion, and found that the musical group showed a large increase in similarity ratings at octaves, while the nonmusical subjects showed less of an increase, though they did also show such a trend. Since even the rat shows an “octave effect” in generalization studies (Blackwell & Schlosberg, 1943) it appears that the nonmusical group here were tending to ignore this categorization in their similarity ratings.

One note in each chord of a harmonic sequence generally stands out, and these notes define the tune. This requires a selective attention mechanism which picks out of a sequence tones with specific characteristics. (The selective attention mechanism proposed by Deutsch & Deutsch, 1963, would perform such a function.) The factors determining which tones are selectively attended to are fairly complex, and only tangentially related to the scheme proposed here. Broadbent (1958) concludes from verbal studies that higher and louder frequencies capture attention better than lower and softer frequencies. Following earlier experiments by Stumpf (1890) and Valentine (1913), Farnsworth (1938) required subjects to make judgments comparing the pitch of isolated intervals with the pitch of single tones. He concluded that the highest tones appear predominant for persons trained in melody hunting, but for the musically naive and also for many basses the lowest tones stand out instead.

The existence of the scheme here outlined cannot at present be directly shown. However it seems probable by examining human musical capabilities that a system like this is among the least complex that might be expected to exist. At present it cannot claim to have more than a heuristic value. It could help the neurophysiologist who is looking for Hubel-Wiesel-type units in the auditory system. It might guide him in choosing among the potentially infinite set of tonal combinations. At the behavioral level the scheme might help to rationalize the psychology and aesthetics of music by providing larger and more fundamental building blocks out of which all music must be built.

REFERENCES

ALLEN, D. Octave discriminability of musical and nonmusical subjects. *Psychonomic Science*, 1967, **7**, 421-422.

BISHOP, P. O. CNS: Afferent mechanisms and perception. *Annual Review of Physiology*, 1967, **29**, 427-484.

BLACKWELL, H. R., & SCHLOSBERG, H. Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, 1943, **33**, 407-419.

BOOMSLITTER, P., & CREEL, W. The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, 1961, **5,** No. 2, 2-30.

BROADBENT, D. E. *Perception and communication*. London : Pergamon Press, 1958.

BUTLER, R. A., DIAMOND, I. T., & NEFF, W. D. Role of auditory cortex in discrimination of changes in frequency. *Journal of Neurophysiology*, 1957, **20**, 108-120.

DEUTSCH, J. A., & DEUTSCH, D. Attention : Some theoretical considerations. *Psychological Review*, 1963, **70**, 80-90.

DEWSON, J. H. Speech sound discrimination by cats. *Science*, 1964, 144, 555-556.

DIAMOND, I. T., & NEFF, W. D. Ablation of temporal cortex and discrimination of auditory patterns. *Journal of Neurophysiology*, 1957, **20**, 300-315.

EVANS, E. F., & WHITFIELD, I. C. Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. *Journal of Physiology*, 1964, **171**, 476-493.

FARNSWORTH, P. R. The pitch of a combination of tones. *American Journal of Psychology*, 1938, **51**, 536-539.

GOLDBERG, J. M., DIAMOND, I. T., & NEFF, W. D. Auditory discrimination after ablation of temporal and insular cortex in the cat. *Federation Proceedings*, 1957, **16**, 47.

GOLDBERG, J. M., & LAVINE, R. A. Nervous system: Afferent mechanisms. *Annual Review of Physiology*, 1968, **30**, 319-358.

GOLDBERG, J. M., & NEFF, W. D. Frequency discrimination after bilateral ablation of cortical auditory areas. *Journal of Neurophysiology*, 1961, **24**, 119-128.

HUBEL, D. H., & WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**, 106-154.

HUMPHREYS, L. G. Generalization as a function of method of reinforcement. *Journal of Experimental Psychology*, 1939, **25**, 361-372.

KRYTER, K. D., & ADES, W. H. Studies on the function of the higher acoustic nervous centers in the cat. *American Journal of Psychology*, 1943, **56**, 501-536.

MEYER, D. R., & WOOLSEY, C. N. Effects of localized cortical destruction upon auditory discriminative conditioning in the cat. *Journal of Neurophysiology*, 1956, **19**, 500-512.

OESTERREICH, R., & NEFF, W. D. Higher auditory centers and the DL for sound intensities. *Federation Proceedings*, 1960, **19**, 301.

OONISHI, S., & KATSUKI, Y. Functional organization and integrative mechanism of the auditory cortex of the cat. *Japanese Journal of Physiology*, 1965, **15**, 342-365.

PITTS, W., & McCULLOCH, W. S. How we know universals. The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, 1947, **9**, 127-147.

RAAB, D. W., & ADES, H. W. Cortical and mid-brain mediation of a conditioned discrimination of acoustic intensities. *American Journal of Psychology*, 1946, **59**, 59-83.

ROSENZWEIG, M. Discrimination of auditory intensities in the cat. *American Journal of Psychology*, 1946, **59**, 127-136.

STOUGHTON, G. S., & NEFF, W. D. Function of the auditory cortex : The effects of one-stage vs. two-stage ablation. *American Psychologist*, 1950, **5**, 474.

STUMPF, C. *Tonpsychologie*, 1890, 384.

SUGA, N. Functional properties of auditory neurons in the cortex of echo-locating bats. *Journal of Physiology*, 1965, **181**, 671-700.

TROTTER, J. R. The psychophysics of melodic interval: Definitions, techniques, theory and problems. *Australian Journal of Psychology*, 1967, **19**, 13-25.

VALENTINE, C. W. The aesthetic appreciation of musical intervals among school children and adults. *British Journal of Psychology*, 1913, **6**, 190-216.

WHITFIELD, I. C., & EVANS, E. F. Responses of auditory cortical neurons to stimuli of changing frequency. *Journal of Neurophysiology*, 1965, **28**, 655-672.

WING, H. Tests of musical ability and appreciation. *British Journal of Psychology*, 1948, No. 27 (Monogr. Suppl.).

(Received May 27, 1968)