A Scale for the Measurement of the Psychological Magnitude Pitch*

S. S. STEVENS AND J. VOLKMANN, *Harvard University, Cambridge, Massachusetts*
and
E. B. NEWMAN, *Swarthmore College, Swarthmore, Pennsylvania*
(Received August 4, 1936)

A subjective scale for the measurement of pitch was constructed from determinations of the half-value of pitches at various frequencies. This scale differs from both the musical scale and the frequency scale, neither of which is subjective. Five observers fractionated tones of 10 different frequencies at a loudness level of 60 db. From these fractionations a numerical scale was constructed which is proportional to the perceived magnitude of subjective pitch. In numbering the scale the 1000-cycle tone was assigned the pitch of 1000 subjective units (mels). The close agreement of the pitch scale with an integration of the differential thresholds (DL's) shows that, unlike the DL's for loudness, all DL's for pitch are of uniform subjective magnitude. The agreement further implies that pitch and differential sensitivity to pitch are both rectilinear functions of extent on the basilar membrane. The correspondence of the pitch scale and the experimentally determined location of the resonant areas of the basilar membrane suggests that, in cutting a pitch in half, the observer adjusts the tone until it stimulates a position half-way from the original locus to the apical end of the membrane. Measurement of the subjective size of musical intervals (such as octaves) in terms of the pitch scale shows that the intervals become larger as the frequency of the mid-point of the interval increases (except in the two highest audible octaves). This result confirms earlier judgments as to the relative size of octaves in different parts of the frequency range.

TWO different concepts of pitch have commonly been held. To the musician pitch has meant the aspect of tones in terms of which he arranges them on a musical scale. The musical scale divides the range of audible frequencies into octaves, which in turn are divided into tones, semi-tones, etc. Then, musically speaking, two semi-tones in different parts of the audible range are considered as equal intervals of pitch. Perceptually, however, these two semi-tones may represent unequal intervals of pitch.

To the physicist, on the other hand, pitch has generally meant frequency. "Pitch is specified scientifically by the period or frequency of vibration."[1] The error of this conception has been recently demonstrated by experiments which show that it is possible to alter the pitch of a tone without changing its frequency.[2] By increasing the intensity of tones of high frequency we raise their pitch; whereas in the case of low frequencies, an increase in intensity lowers the pitch. This change of pitch may amount to as much as half an octave[3] at certain low frequencies. Clearly, then, pitch is not frequency, nor is it correlated one-to-one with frequency.

These considerations have led to the proposal[4] that the designation of pitch should take into account the loudness of a tone. In particular, it is argued that the pitch of a given tone should be specified in terms of a reference tone at a loudness level of 40 db which is perceived as equal to the given tone in pitch. It has been proposed further that the pitch of a tone at the loudness level of 40 db should be designated by its frequency, or, alternatively, by the number of octaves that it is above a given reference frequency.

This proposal is a partial recognition of the fact that pitch, like loudness, depends upon the perceiving organism, and that the numbers used to designate values of the physical stimulus are not adequate to represent values of such a psychological magnitude. A desirable scale for the pitch of tones at the reference loudness (40-db level) would be one expressed in numbers whose values are directly proportional to the magnitude of the perceived pitch. The present experiment seeks to establish such a psychological scale of pitch.

* Presented at the meeting of the Acoustical Society of America, New York, October 31, 1936.

[1] E. H. Barton, *A Text-book on Sound* (Macmillan, London, 1914), p. 9.

[2] S. S. Stevens, "The Relation of Pitch to Intensity," J. Acous. Soc. Am. **6**, 150–154 (1935).

[3] W. B. Snow, "Change of Pitch with Loudness at Low Frequencies," J. Acous. Soc. Am. **8**, 14–19 (1936).

[4] H. Fletcher, "Loudness, Pitch and the Timbre of Musical Tones and their Relation to the Intensity, the Frequency and the Overtone Structure," J. Acous. Soc. Am. **6**, 59–69 (1934).

The Problem of Sensory Scales

Controversy has for a long time centered around the proposition that it is possible to measure attributes of sensations, or to tell when one sensation is twice or three times as great as another. The truth of this proposition must depend not upon *a priori* conceptions but upon the outcome of experiment. We must first decide what we shall mean by a sensory scale (of pitch, let us say) and then determine by experiment whether or not such a scale can, in fact, be constructed. In other words, we must decide upon the criteria of the scale, and then devise operations for satisfying the criteria.

The criteria for a psychological scale were examined in a previous paper[5] in connection with the development of a scale for loudness. The following points will bear repeated emphasis.

(1) There are, in general, two types of scales, commonly referred to as *intensive* and *numerical* scales. Scales which measure intensive magnitude enable us to place the things measured in a rank-order, i.e., arrange them according to increasing magnitude. Such a scale does not, however, enable us to say how many times one magnitude is greater than another, but only that it is greater. Numerical scales, on the other hand, have numbers which express the numerical relations between things measured. Thus, when two magnitudes are measured by a numerical scale, the scale numbers can be manipulated in accordance with arithmetical laws in order to determine additional relationships such as the sum of two magnitudes, the relative separation of two pairs of magnitudes, etc.

(2) These manipulations of the numbers on numerical scales have significance only if the manipulations correspond to a set of concrete operations which can be performed on the things measured. Otherwise, the validity of the outcome of the manipulations cannot be tested empirically. The concrete operations will, in general, be different for different types of measurement. Thus, the procedure for verifying that 2 meters can be added to 2 meters to make 4 meters is very unlike that for showing that 2 henries of inductance can be added to 2 henries to give 4 henries. Similarly, in the case of sensation, we may reasonably expect to find that numerical scales are based on operations peculiar to it alone.

(3) Now, in the case of psychological measurements, most scales have been scales of intensive magnitude. What we want is a numerical scale—one whose numbers represent some aspect of the response of a living organism to a certain class of stimuli.

(4) The numbers of the numerical scale should be applied to the attribute of sensation (which is, of course, an aspect of an organism's response) in such a way that when they are manipulated according to the rules of arithmetic, one obtains a result which can be verified observationally. To the manipulations and to the result there should correspond a set of concrete operations.

(5) Although we could conceivably choose any one of several sets of operations as defining the scale,[6] that set will ultimately prove most satisfactory which leads to scale numbers bearing the most reasonable relation to the experience of the observer. A reasonable scale is one on which the number N stands for a sensation which does in fact appear to be half as great as that represented by the number $2N$, etc.

Procedure

In view of the success encountered in the construction of a numerical scale of loudness[7, 8] we employed an analogous procedure for pitch. The problem is to determine the relation between a numerical scale of perceptual pitch and the scale of frequency used to measure the stimulus. The observer was required, in our experiment, to adjust the frequency of a second tone until it sounded just half as high in pitch as a standard tone. In order to guard against the effect of intensity on pitch, tones were presented at a constant loudness level of 60 db. The fractionaation of several tones, scattered throughout the audible range, thus provides a basis for assigning numbers to the tones such that they constitute a numerical scale of perceived pitch.

[5] S. S. Stevens, "A Scale for the Measurement of a Psychological Magnitude: Loudness, "Psych. Rev., **43**, 405–416 (1936).

[6] C. H. Graham, "Psychophysics and Behavior," J. Gen. Psych. **10**, 299–310 (1934).

[7] H. Fletcher, "Newer Concepts of the Pitch, Loudness and Timbre of Musical Tones," J. Frank. Inst. **220**, 405–429 (1935).

[8] B. G. Churcher, "A Loudness Scale for Industrial Noise Measurement," J. Acous. Soc. Am. **6**, 216–226 (1935).

This scale can be no more reliable than the performance of the observers. The question arises, in the first place, whether it is possible to make such a judgment about pitch. Several observers made the statement *a priori* that pitch is not something of which they could take half; yet these same observers found upon trying that the judgment is quite possible. They were not explicit as to how they made the judgment, except to say that when the second tone was set to a certain frequency, it "felt like half," whereas at other frequencies it "felt" too high or too low. Observers differed greatly in the ease with which they acquired the necessary attitude.

The following written instructions were given:

"You will be presented with two tones which differ in pitch. The pitch of one tone may be varied by turning a crank; you are to adjust this tone until its pitch is just *half* of the pitch of the fixed tone. During the course of the adjustment, take care to produce values of the variable pitch which are plainly higher than the desired half-value and other values which are plainly lower. If the fixed and the variable tones differ widely in loudness, report this fact."

Apparatus

The tones were generated by two beat-frequency oscillators. The tone from one oscillator was fixed in frequency at 125, 200, 300, 400, 700, 1000, 2000, 5000, 8000 or 12,000 cycles. The other tone could be varied continuously by the observer, who turned a small crank geared to the tuning condenser of the second oscillator (General Radio type 713-A). The observer faced the source of sound, a dynamic and a crystal speaker connected in parallel and mounted in a baffle. The experiment was conducted in a highly absorbent sound-room.

TABLE I. *Showing the geometric means of fractionations by each of five observers, and average errors for 2 observers.*

Standard Freq.	Frequency at Half-Pitch						Average Error (%)	
	Observer					Geometric Mean	Observer	
	E	*V*	*N*	*M*	*D*		*E*	*V*
125	96	84	97	87	96	90	3.4	7.5
200	131	131	123	101	107	121	7.4	6.7
300	183	205	195	118	133	171	13.0	7.8
400	237	251	237	173	253	233	13.5	9.0
700	391	407	474	256	403	384	12.3	9.8
1,000	590	640	622	391	485	558	11.0	13.9
2,000	998	979	1006	508	662	851	14.8	13.7
5,000	1930	1800	2260	1230	1590	1767	12.0	13.5
8,000	2360	2170	3110	1650	2110	2239	19.1	13.9
12,000	2970	2400	3720	2320	2980	2788	19.9	11.5

The two tones were presented alternately by a relay which switched from one to the other at two-second intervals. Care was taken to keep both tones at a loudness level of 60 db. Voltages corresponding to the 60-db level were determined beforehand in terms of the thresholds of the three of us, and then as the observer changed the frequency of the second tone, the experimenter adjusted its voltage, if necessary, to keep it at the 60-db loudness level.

Results

Five observers (two of the authors and three others) were used in the experiment. At each experimental session each observer made one fractionation of each of the 10 standard tones. Five sets of fractionations were made by 3 observers and 10 sets by 2 observers. These 10 sets were made in order to get an indication of the reliability with which an observer is able to set one pitch to half of the value of another pitch. The averaged results for each observer are presented in Table I. In averaging it appeared reasonable to take the geometric mean of the individual adjustments rather than the arithmetic mean, in order to compensate for the fact that the differential sensitivity of the ear (measured in cycles per second) is less for high than for low frequencies. In taking the final

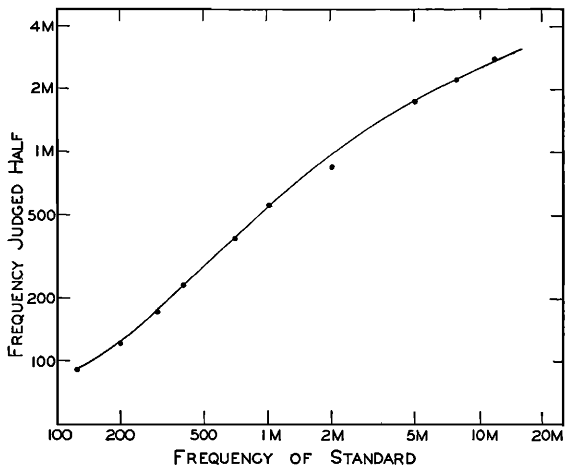

FIG. 1. The ordinate represents the frequency of the tone whose pitch was judged half as high as the pitch of the standard tone (abscissa). The smooth curve was fitted visually to the geometric mean of the adjustments of five observers. The point at 2000 cycles deviates from the general trend, due to the tendency of observer M to select values an octave below the other observers at this frequency.

average for the 5 observers their individual averages were weighted according to the number of judgments made by each.

These observers show good agreement, considering the supposed difficulty of fractionating pitch. Only one of them (M) departs significantly from the average of the group. He is a trained musician, and reported an inability to disregard octaves and other musical intervals when setting the second tone at half the pitch of the first. The other observers were apparently not confused by the recognition of these musical relationships. Observer M, however, tended to come more into line with the other observers as the experiment proceeded. His earlier settings were also upset by his conscious effort to imagine what "zero pitch" might be. It appears that a somewhat non-analytical attitude is necessary on the part of the observer if he is to maintain a consistent criterion for his judgment.

In order to construct a pitch scale the averaged results were plotted as in Fig. 1, and a smooth curve was fitted visually to the points. Then, from this smooth curve we proceeded to construct the *pitch function* in Fig. 2. We assigned the number 1000 to the pitch of a 1000-cycle tone and determined from Fig. 1 the frequency of the tone which sounds half as high and which should have, therefore, the number 500 assigned to it. Similarly for the pitch number 250, etc. The result is a function (Fig. 2) which has, within the limitations of our particular procedure, the numerical significance which we set out to give it; namely, the numbers on the pitch scale are related to each other as the subjective magnitudes of the pitches. A pitch of 1000 units (mels[9]) is subjectively twice as high as a pitch of 500 units. "Twice as high" is, of course, defined by the operations of this experiment.

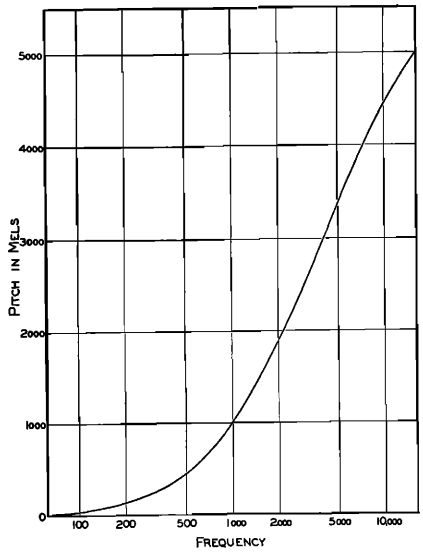

Fig. 2. The pitch function, showing the relation of perceived pitch (in mels) to the frequency of the stimulus. The values in mels are derived from the curve of Fig. 1; the 1000-cycle tone is arbitrarily assigned the value of 1000 mels.

Relation of the Pitch Function to Other Data

Having established a pitch scale, we can use it to measure certain psychological magnitudes, and by comparing it with physiological data, we can obtain a suggestion as to the probable basis of the judgment of pitch magnitudes.

Relation of pitch to DL's

Psychologists have long debated the proposition that all just-detectable increments in frequency (difference limens or DL's) are subjectively equal in magnitude. Now, if each DL increases the pitch of a tone by a constant amount, the summated DL's should give a pitch scale which agrees with the scale in Fig. 2. In order to test this hypothesis we integrated graphically the DL's determined by Shower and Biddulph[10] and obtained the values shown in Fig. 3. The integration was made of the data for the loudness level of 60 db. Several assumptions are possible as to the best limits of integration. For the lower limit we chose 60 cycles, because at lower frequencies Shower and Biddulph found an anomalous decrease in the size of the relative DL's. The number of DL's below 60 cycles is so small, however (about 10 or 20) that their effect on the total integration is negligible for our purpose. At the upper end we integrated as far as available data would permit (about 12,000 cycles). This coincides with the highest tone we actually used in determining the pitch function in Fig. 2.

[9] The name *mel* was chosen as a name for the subjective pitch unit. It was taken from the root of the word *melody*.

[10] E. G. Shower and R. Biddulph, "Differential Pitch Sensitivity of the Ear," J. Acous. Soc. Am. **3**, 275–287 (1931).

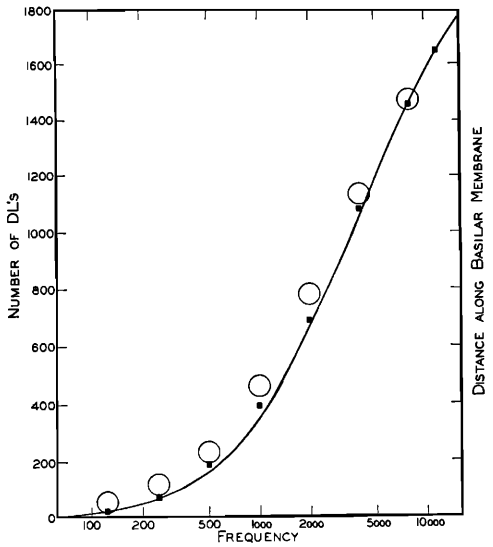

FIG. 3. The relation of the pitch function (solid curve) to the integrated DL's (solid squares) and to the experimental location of resonant positions on the basilar membrane (circles). The ordinate scale (on left) shows the number of DL's for pitch as a function of frequency when the integration is made at the 60 db loudness level. The pitch scale in mels can be obtained by multiplying the ordinate values by the factor 2.83. The relative locations of resonant areas on the basilar membrane are obtained by laying the linear extent of the membrane along the ordinate scale. Thus the ordinate on the right represents the relative linear extent of the membrane both in man and guinea pig.

The close agreement of the integrated DL's and the pitch function (Fig. 3) shows clearly that, within the limits of accuracy of present measurements, all DL's for pitch, at a constant loudness level, are of equal subjective magnitude. This conclusion is unlike that reached in the case of loudness[11] where DL's increase rapidly in size with increasing intensity of the stimulus. The fact that the size of DL's for pitch is constant, whereas the size of DL's for loudness varies, testifies to the fact that pitch and loudness are based on quite dissimilar physiological mechanisms. Perhaps the two judgments of loudness, magnitude and DL, involve different types of physiological mechanisms, whereas the judgments of magnitude and DL for pitch are based upon the same type of physiological mechanism. Both magnitudes and DL's for pitch depend upon discrimination of the *location* of stimulation on the basilar membrane. Loudness appears to depend upon the total *number* of fibers stimulated in the auditory nerve. The mechanism underlying the judgment of DL's for loudness is as yet undetermined.

Relation of the pitch scale to basilar mechanics

Experiments designed to determine the location of the region of the basilar membrane which is resonant to different frequencies have been conducted by means of the electrical recording of potentials generated in the ears of guinea pigs.[12, 13] Surgical damage to the basilar membrane at a particular location results in a decreased sensitivity to tones of a certain frequency. The "map" of the basilar membrane obtained by correlating locations of lesions with abnormalities in the corresponding audiograms is almost precisely the same "map" as that resulting from an integration of DL's for pitch. The latter is based on the assumption that for each DL the two tones being discriminated must stimulate parts of the basilar membrane separated by a constant distance.

Now, if we let the ordinate of Fig. 3 represent the linear extent of the basilar membrane, and plot the resonant positions as a function of frequency, we obtain the function suggested by the circles in Fig. 3. The size of these circles represents approximately the probable error of the measurements made in the work with guinea pigs. The obvious correspondence between the locations of the resonant regions determined by experiment and by integration of DL's, and their correspondence in turn to the pitch function suggests an interesting hypothesis. Apparently, when an observer is asked to set a second tone to half the pitch of a given tone, he changes its frequency until it stimulates a position on the basilar membrane midway between the position stimulated by the given tone and the apical end of the membrane. He is, of course, not aware of these locations as such, but the underlying physiological process which makes comparison of pitches possible seems to be characterized chiefly by spatial differentiation. Although there are subsequent central nervous processes, the form of certain discriminatory functions is evidently imposed by the receptor mechanism.

[11] S. S. Stevens and H. Davis, "Psychophysiological Acoustics: Pitch and Loudness," J. Acous. Soc. Am. **8**, 1–13 (1936).

[12] S. S. Stevens, H. Davis and M. H. Lurie, "The Localization of Pitch Perception on the Basilar Membrane," J. Gen. Psych. **13**, 297–315 (1935).

[13] E. Culler, Ann. Otol. Rhin. Laryn. **44**, 808–814 (1935).

Relation of the pitch scale to musical intervals

An interesting application of the pitch scale is the measurement of the size of musical intervals. We can measure the subjective size of the various octaves by determining from Fig. 2 how much the pitch changes in going from one octave to the next. In a similar way we can measure the size of other musical intervals. In general, the subjective size of any musical interval is approximately proportional to the slope of the pitch function (Fig. 2) at the frequency which falls in the middle of the interval.

In order to illustrate these relationships, the subjective size of successive octaves and fifths, as measured in mels, is plotted in Fig. 4. The plot for other intervals would be similar in form but different in ordinate value.

Quite definitely, musical intervals become subjectively larger as frequency increases up to the fourth octave above middle C (up to 4096 cycles). In other words, throughout the useful musical range, intervals increase in subjective magnitude with increasing frequency of the stimulus. This conclusion is not entirely novel. The eminent psychologist-musician, Stumpf,[14] decided that in spite of the great difficulty of making these subjective comparisons, the upper octaves are perceptually larger than the lower octaves. Thus the pitch scale enables us to confirm Stumpf's judgment.

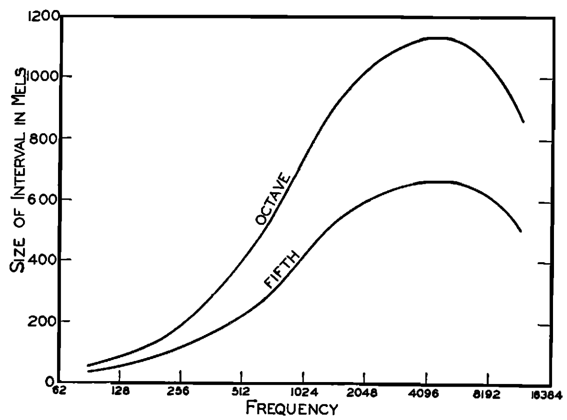

FIG. 4. The size of musical intervals in terms of mels. The upper curve shows the perceived size of octaves as a function of the geometric mean of the limiting frequencies. The lower curve shows in a corresponding manner the perceived size of fifths.

Discussion

The question arises concerning the possibility of verifying the pitch function by other procedures such as bisecting tonal intervals, i.e., setting a third tone to a value half-way, in pitch, between two other tones. Such verification is theoretically possible—in fact it is theoretically required if the pitch scale is valid. The ability of any two methods of fractionation or bisection to confirm each other is limited, however, by the existence of constant errors in the experimental procedures. Some of the factors which may introduce constant errors are the size of the interval, the position of the interval on the frequency scale, the order of presentation of the stimuli, the rate of presentation, the position of the variable tone before the observer begins the adjustment, and the effect of what is known as "absolute judgment," namely, a tendency to adjust the variable tone to a value which is to some extent independent of the two limiting tones, but dependent upon preceding judgments. The choice of a function to be used as a pitch scale will, therefore, be subject to revision whenever the sources of constant errors in the experimental procedures can be detected and eliminated.

The method of bisection has been applied to tonal intervals. The results of different investigators, however, have thus far been contradictory.[15] Some workers have insisted that the bisection is made at the arithmetic mean and some claim evidence for bisection at the geometric mean. A famous historical controversy was waged about this point.[16] From the form of the present pitch function it is evident that the bisection of an interval should depend upon the position of the interval on the frequency scale. Hence, both parties to the controversy may be correct.

[14] C. Stumpf, Tonpsychologie **I**, 250 (1883).

[15] C. C. Pratt, "Comparison of Tonal Distance," J. Exper. Psych. **11**, 77–87 (1928).

[16] E. B. Titchener, Experimental Psychology **2**, part 2, 232–248 (1905).