

February 01, 1961

Voice level: autophonic scale, perceived loudness, and effects of sidetone

H. L. Lane
Harvard University

A. C. Catania
Harvard University

S. S. Stevens
Harvard University

Recommended Citation

Lane, H. L.; Catania, A. C.; and Stevens, S. S., "Voice level: autophonic scale, perceived loudness, and effects of sidetone" (1961). *Psychology Faculty Publications*. Paper 22. <http://hdl.handle.net/2047/d20000869>

Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Sidetone*

H. L. LANE, A. C. CATANIA, AND S. S. STEVENS

Psycho-Acoustic Laboratory, Harvard University, Cambridge, Massachusetts

(Received August 22, 1960)

The speaker's numerical estimation of his own vocal level, the *autophonic response*, was found to grow as the 1.1 power of the actual sound pressure produced. When listeners judged the loudness of another speaker's vocalization (the phoneme [a]), the exponent was 0.7. The disparity between these exponents suggests that the speaker does not rely solely upon his perception of loudness in judging his own relative vocal level. The minor role played by loudness in the autophonic judgment is further demonstrated by the fact that the form and exponent of the subjective scale for autophonic responses remain relatively invariant under wide changes in auditory feedback.

The power laws governing the autophonic response (exponent

1.1) and loudness (exponent 0.6) were used to predict successfully the outcome of cross-modality comparisons in which subjects tried to match their vocal level to sounds of various intensities presented either by loudspeaker or by earphone. The slope of the matching function, relating the criterion SPL to the vocal SPL in log-log coordinates, is given by the ratio of the two exponents.

Unless the speaker tries deliberately to hold a constant level, the amount of sidetone gain with which the voice is fed back to the ears alters the voice level. The degree to which the speaker lowers his voice when the sidetone is increased is also predicted by the exponents governing the autophonic scale and the loudness scale.

ALTHOUGH most of the sounds we hear are generated by external sources, we generate a great many ourselves. Is the loudness scale different for a listener who is also his own source of sound? When judging his own vocal production, his *autophonic* output, does a person depend upon his perception of loudness, or does he judge some other variable, such as muscular effort? These questions set the stage for the present attempt to determine the speaker's subjective scale for his *autophonic response*.

The development of the sone scale for loudness¹ has served both as a prototype and as a catalyst for much current research on subjective scaling in other modalities. As a result, a variety of new techniques for ratio scaling and for its validation have become available. For the assessment of the relative psychological magnitudes of vocal levels, as judged by the speaker himself, the methods of magnitude production and magnitude estimation² were employed in the studies reported here. The validity of the subjective scale thereby obtained was tested with the aid of a third method, cross-modality matching.³ Finally, the role played by perceived loudness in determining the subjective magnitude function for autophonic level was assessed by varying the amount of auditory feedback and by masking the feedback completely.

PROCEDURE

Magnitude Production

The subject was asked to produce the vowel phoneme [a] at medium level. To the magnitude of this response, the experimenter assigned the numerical value 10. A series of values (2.5, 5, 10, 20, 30) were then named in irregular order and the speaker was asked to respond

to each with a vocal production of proportionate magnitude.

With this method, the subjective autophonic scale was determined under four conditions of auditory feedback: (1) the speaker vocalized in an anechoic space with the ears open; (2) the speaker wore a pair of disconnected PDR-8 earphones with Neoprene cushions MX-41/AR; (3) the earphones generated a masking noise of 110 db; (4) the speaker's voice was fed back to him over the earphones at three levels of amplification (0, 10, and 20 db above an arbitrary reference gain).

The speaker's vocal productions were picked up by a dynamic microphone, amplified, and sent to a graphic level recorder (Sound Apparatus Company). The level of each vocalization was read from the level recorder. The masking noise used to suppress auditory feedback was produced by passing a white noise through a filter (UTC-4C) with cutoff frequencies set at 100 and 2000 cps.

Magnitude Estimation

The method of magnitude estimation requires that the subject make a direct numerical estimate of the psychological magnitudes of a series of stimuli. Usually the experimenter presents a standard stimulus and assigns it a numerical value (such as 10). He then presents other stimuli and the subject assigns numbers to them proportional to their apparent magnitude. In the studies reported here, the stimuli were vocal levels of the vowel phoneme [a], produced by the subject himself. The subject watched the effect of his voice on the needle of a VU meter whose scale had been obscured, and his task was to center the pointer on the face of the meter. The experimenter controlled the gain in the microphone circuit and thereby determined the vocal level necessary for centering. The experimenter selected an intermediate gain setting as the standard and assigned the value 10 to the vocal level required to center the needle on the VU meter. The gain setting was then changed, and the subject was again asked to center the needle. The

* This research was carried out with the Office of Naval Research (Project Nr142-201, Report PNR-242).

¹ S. S. Stevens, *J. Acoust. Soc. Am.* **31**, 995 (1959).

² S. S. Stevens, *Psychol. Bull.* **54**, 177 (1958).

³ S. S. Stevens, *J. Exptl. Psychol.* **57**, 201 (1959).

experimenter monitored the vocal productions on a separate meter (Ballantine VTVM). As soon as the subject held the needle centered within a 3-db range for approximately 2 sec, he received a signal (a light flash). Thereupon the subject stopped vocalizing and estimated the autophonic level of his production by assigning it a number proportional to its apparent magnitude. The method of magnitude estimation was used under three of the four conditions of auditory feedback described earlier: open ears, wearing earphones, and wearing earphones with masking noise.

In addition to the use of the method of magnitude estimation in determining the autophonic function, this method was employed in two experiments on the loudness of speech produced by an external source. In one experiment, a speaker gave vocal productions of the phoneme [a] at six sound pressure levels equally spaced over a 36-db range. As in the experiments on magnitude estimation of vocal effort, the speaker was required to center the needle on a VU meter while the experimenter controlled the gain in the microphone circuit. Judgments of the speaker's loudness were made by each of 10 listeners.

In a second experiment, a single production of the phoneme [a] at a medium level was recorded on a loop of magnetic tape and played back at about the same levels as those employed in the prior study of the loudness of live speech. Ten subjects, eight of whom had served in the first experiment, made loudness estimates.

Cross-Modality Matching

Under the method of cross-modality matching, the experimenter presents a series of criterion stimuli (in one modality) and the subject produces stimuli (in another modality) that seem equal in apparent intensity. For the experiments reported here, the criterion stimulus was a band of noise (100 to 2000 cps) presented at six levels over a 50-db range. The noise was generated by a loudspeaker in one experiment, and by earphones in another. The criterion stimulus was presented for 2 sec, and the subject then produced the phoneme [a] so that its apparent intensity seemed to him equal to that of the criterion.

The level of the criterion noise was controlled by means of calibrated attenuators and was monitored on a vacuum-tube voltmeter (Ballantine). The level of the vocalization produced to match the noise was read from the graphic level recorder.

In a related procedure, the criterion stimulus was the vocal sidetone produced by the subject, and the subject's task was to vocalize so as to compensate for, rather than match, changes in the intensity of the criterion. That is to say, the subject was instructed to hold the loudness of his voice, as he perceived it, constant under various levels of sidetone.

Some 40 different subjects participated in one or more of the various experiments. Twenty-four subjects were

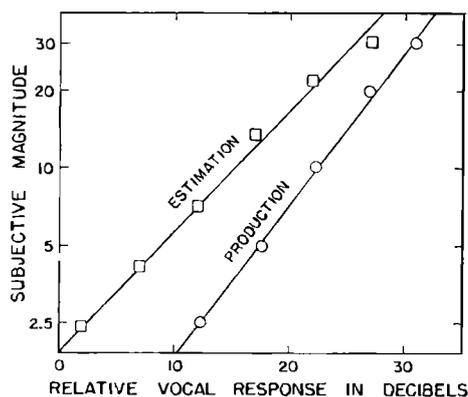


FIG. 1. The subjective scale of autophonic response determined by magnitude estimation and magnitude production. Each square represents the geometric mean of three judgments by each of 10 subjects. Each circle represents the mean decibel level of three responses by each of 24 subjects.

employed in the magnitude-production experiments with open ears and with masking noise. Groups of 10 subjects served in each of the other experiments. In the experiment on the matching of vocal level to a criterion noise presented over a loudspeaker, each subject made two responses at each stimulus level. In all other experiments each subject made three vocal productions, estimations, or matches, at each of the stimulus values presented.

SUBJECTIVE SCALE OF THE AUTOPHONIC RESPONSE

Figure 1 shows the speaker's subjective scale of autophonic response as determined by numerical estimation of the levels produced, and by vocal production made to numerical criteria. Each version of the scale is well represented by a straight line in log-log coordinates. Consequently, the relation between actual and perceived magnitude of vocalizing can be considered a power function whose exponent is given by the slope of the line of best fit to the geometric means of the data. The subjective scale of the autophonic response had an exponent of 1.17 when obtained by the method of magnitude production, and an exponent of 0.91 when obtained by the method of magnitude estimation.

The difference between the slopes (exponents) obtained with the two procedures may also be seen in the results of experiments, reported below, on the effect of auditory feedback on vocalizing. In general, the method of magnitude estimation tends to underestimate the exponent of subjective scales relative to the estimate obtained by the method of magnitude production.⁴ Both methods have systematic biases that are to some extent complementary, so that the best estimate of the exponent of the autophonic scale presumably lies somewhere between 0.9 and 1.2. As a first-order approximation, a slope of 1.1 may be chosen as representative.

⁴S. S. Stevens and E. C. Poulton, *J. Exptl. Psychol.* 51, 71 (1956).

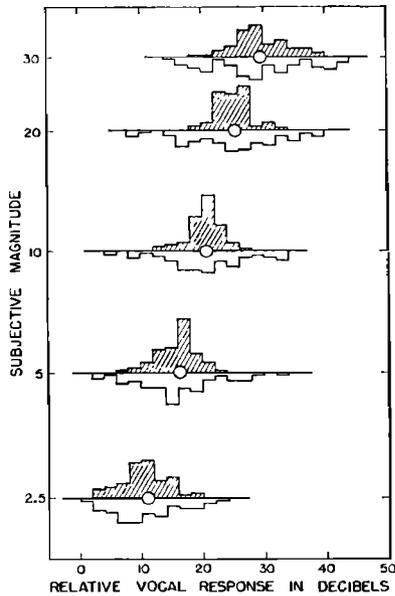


FIG. 2. Distributions of vocal productions before (white histograms) and after (shaded histograms) correction for the component of variability due to the subject's choice of modulus. The circles represent the means of the vocal levels (in db) produced for each of the five criterion values shown on the ordinate. Each histogram contains 72 scores.

Autophonic response may thus be added to the list of more than two dozen other continua on which psychological magnitude has been demonstrated to be a power function of the stimulus.⁵

It is of particular interest to compare the subjective scale for autophonic level with the subjective scale for loudness (same scale). The psychophysical law governing both modalities, speech and hearing, is a power function. Loudness, however, grows approximately as the 0.6 power of the sound pressure⁶ whereas the autophonic scale grows approximately as the 1.1 power of the sound pressure. These two exponents tell us that, when the speaker raises his voice by what he judges to be a factor of 2, his voice will not sound twice as loud to a listener. In other words, there is great difference from the subject's point of view between the relative subjective magnitude of sounds that he generates by his own vocal effort and those that are generated by an external source.

The exponent for autophonic output obtained by the methods of magnitude estimation and magnitude production is in general agreement with some early results obtained by the method of fractionation. In a study by Clark *et al.*,⁷ a single subject produced the vowel [a] alternately at levels that he considered maximum and half-maximum, for several trials. He then made alternate productions of what he judged to be one-half and one-quarter of the maximum level, and, finally, of one-

TABLE I. Standard deviations of vocal levels obtained by magnitude production and magnitude estimation. Correction was made for the component of variability attributable to the subject's choice of modulus.

Criterion value	Magnitude production ($N=72$)		Magnitude estimation ($N=30$)	
	Standard deviation of vocal level (db)	Un-corrected for modulus	Vocal level produced (relative db)	Standard deviation of estimates (log units)
2.5	5.2	3.4	5	0.16
5	6.3	3.3	10	0.15
10	6.6	2.6	15	0.13
20	6.9	2.8	20	0.08
30	6.7	4.0	25	0.08
			30	0.08

quarter and one-eighth maximum. The average decibel differences obtained in the three successive halvings were 4.5, 6.2, and 7.3 db. These results may be compared to those obtained when the method of magnitude production is employed. Because a power function describes the relation between actual and perceived magnitude of vocalizing (Fig. 1), a constant subjective ratio corresponds to a constant stimulus ratio (constant decibel difference). Since the exponent of the autophonic function obtained in the present study is 1.1, the decibel difference predicted for the subjective ratio of one half is 5.5 db, which is close to the average value for halving obtained by Clark *et al.*

VARIABILITY

Although our primary concern is with the exponent of the autophonic scale, it is of interest to examine the variability associated with each of the assessment techniques and to try to diagnose its sources. Table I presents the standard deviations for each of the geometric means plotted in Fig. 1. The second column shows the variability (in db) in the level of vocal production for each of the five criterion values. The last column shows the variability (in log units) of the numerical estimates of vocal level at each of the six levels employed.

Stevens⁸ has enumerated three sources of variability associated with ratio-scaling methods: (1) variability due to the subject's choice of modulus, i.e., his conception of the standard; (2) variability due to the subject's conception of a subjective ratio; and (3) variability due to differing sense-organ operating characteristics. The data obtained by both methods, magnitude production and magnitude estimation, are subject to all three sources of variability, but with the method of magnitude estimation as used here there is relatively little variability due to the choice of modulus. In the method of magnitude estimation the modulus was determined by requiring the speaker to produce a given vocal level (which centered the needle on the VU meter). This standard level was called 10. The only significant

⁵ S. S. Stevens, *Psychol. Rev.* 64, 153 (1957).

⁶ S. S. Stevens, *J. Acoust. Soc. Am.* 27, 815 (1955).

⁷ K. C. Clark, H. W. Rudmose, J. C. Eisenstein, F. D. Carlson, and R. A. Walker, *J. Acoust. Soc. Am.* 20, 776 (1948).

⁸ S. S. Stevens, *Am. Scientist* 48, 226 (1960).

source of variability due to the modulus was presumably the scatter attributable to the inability of the subject to remember the standard. In magnitude production, however, the subject was free to select his own modulus, with the sole constraint that the vocal production be of "medium" level. With the method of magnitude production, therefore, the standards varied from one speaker to another. That component of the total variance that was due to the subject's choice of modulus was removed by treating each vocal response in the following way. A grand mean of the vocal productions of the group was first computed. The mean of all the responses by a given subject was then subtracted from the grand mean and the difference was added to each one of that subject's vocal productions. This operation left unchanged the slope of each subject's magnitude function for autophonic output, but it minimized the sum of the squared deviations of his productions around the regression line for the group. The standard deviations of the productions so treated are shown in column 3 of Table I. We note that almost half the total variability obtained with the method of magnitude production is accounted for by the fact that each subject chooses a different modulus.

Figure 2 presents the distributions of the vocal productions for each of the criterion values before (white histograms) and after (shaded histograms) the removal of the component of variability due to the choice of modulus. When this component is removed, the distributions are more nearly normal and considerably reduced in range.

CROSS-MODALITY MATCHING

Recent research on subjective scaling has shown that the relations among the exponents of the power functions governing different sense modalities may be verified by cross-modality matches.^{1,3} In general, if two continua are governed by the equations

$$\psi_1 = \phi_1^m \text{ and } \psi_2 = \phi_2^n$$

and if the psychological values, ψ_1 and ψ_2 are equated at various levels, it follows that the stimulus values ϕ_1 and ϕ_2 should stand in the relation

$$\log \phi_1 = (n/m) \log \phi_2.$$

In other words, cross-modality matches should produce a function that is a straight line when plotted in log-log coordinates and has a slope given by the ratio of the exponents n and m .

In order to verify the form and exponent of the subjective autophonic function by cross-modality comparisons, two experiments were performed in which subjects were asked to match their vocal level to a white noise, presented at various intensities by loudspeakers or by earphones. Figure 3 shows that the resulting equal-sensation function has a slope of about 0.5, which

approximates the ratio of the previously determined exponents for loudness and autophonic response.

The matching of the apparent intensity of autophonic production to that of an external sound closes a circle in a process of validation. The 10 subjects who matched vocal level to the loudness of noise presented over a loudspeaker (top curve, Fig. 3) also served in the experiment on magnitude production of autophonic level described above (Fig. 1) and in an experiment on loudness estimation. The slope of the autophonic function for this particular group was 1.06. When these subjects estimated the loudness of various levels of white noise, covering a range of 50 db, the resulting subjective scale of loudness had a slope of 0.59. Hence the predicted slope (exponent) for the equal-sensation function is 0.56; the obtained slope (line of best fit to the mean values shown in Fig. 3) turned out to be 0.51. When earphones, rather than a loudspeaker, served as the noise source, a slope of 0.52 was obtained.

Figure 3 also shows the results of an experiment by Black⁹ in which 24 subjects attempted to speak a monosyllable at the same loudness as that of an external sound whose level covered a 30-db range (20 db less than in the present experiments). Black employed two sound sources, "the output of a Clarkstan sweep frequency oscillator, 40-10 000 cps range repeated 20 times/sec" and a white noise. He also used three orders of stimulus presentation: random, descending, and ascending. Despite the differences in procedure, there is fair agree-

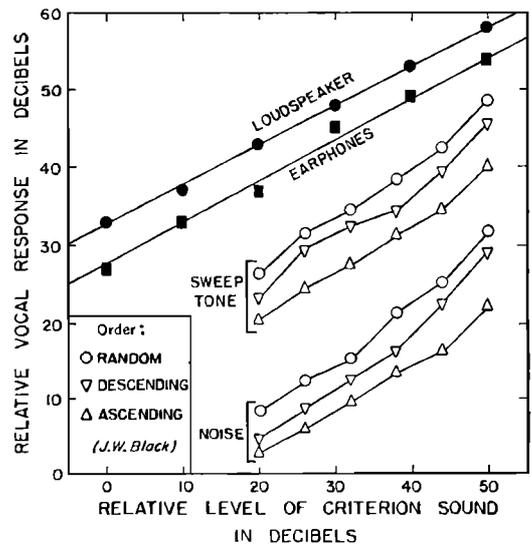


FIG. 3. Equal-sensation functions obtained by matching autophonic response to loudness. Subjects vocalized to match criterion stimuli (noises) presented over a loudspeaker (filled circles) and earphones (filled squares). Each point represents the mean level (in db) of two matches by each of 10 subjects (filled circles) or of three matches by each of 10 subjects (filled squares). The unfilled points are from a comparable experiment by Black.⁹

⁹ J. W. Black, Control of the sound pressure level of voice, Joint Project Rept. No. 42, Bureau of Medicine and Surgery, U. S. Navy Project No. NM 001 104 500.42 (February, 1955).

ment between Black's curves and those from the present study.

ROLE OF HEARING

Clearly a speaker does not judge the relative level of his own voice on the basis of loudness alone, for the power functions governing loudness and autophonic response have different exponents. In order to assess the role that loudness plays in determining the autophonic function, several experiments were performed in which auditory feedback (sidetone level) was increased or decreased relative to the standard condition (open ears). These experiments demonstrate that, under extensive changes in the auditory feedback that a speaker receives from his own voice, the scale of vocal effort remains relatively invariant in form and slope.

Figure 4 shows autophonic scales obtained by the method of magnitude production under both attenuated and amplified auditory feedback. The function labeled "with masking" was obtained from 24 speakers who were subjected to a masking noise of 110 db. This same group of speakers had served in the magnitude-production experiment discussed above (Fig. 1), which was performed under standard conditions in an anechoic space. The function labeled "wearing earphones" (unfilled circles in Fig. 4) was obtained from subjects who wore disconnected earphones mounted in MX-41/AR cushions, which presumably attenuated the air conducted sidetone by the amount attributable to these cushions.¹⁰ This experiment was replicated with a second group of 10 subjects and the function obtained is also shown in Fig. 4 (filled circles). Finally, this second group of subjects made vocal productions under 0, 10, and 20 db of auditory feedback (re an arbitrary standard). In these several experiments the effective level of the vocal sidetone varied, therefore, from practically zero under intense masking to as much as 110 db SPL when the subjects spoke loudly with the sidetone gain set at 20 db.

Auditory feedback was also varied in the experiments that used the method of magnitude estimation. Subjects vocalized and gave numerical estimates of their autophonic outputs (1) without earphones, (2) with earphones, and (3) with earphones energized by a masking noise. In one series of experiments, the autophonic productions that were estimated varied over a range of 40 db. Licklider and Miller¹¹ give 40 db as the range of average speech power between the loudest and the weakest vocalizing possible. Some of our subjects had difficulty producing the lowest and the highest vocal intensities required, and two subjects who could not

¹⁰ For a curve showing the attenuation of the earphone socket MX-41/AR, see J. Zwislocki, *J. Acoust. Soc. Am.* **27**, 155 (1955); see also "Ear protection," in *Handbook of Noise Control* edited by C. Harris (McGraw-Hill Book Company, Inc., New York, 1957), Chap. 8.

¹¹ J. C. R. Licklider and G. A. Miller, "The perception of speech," in *Handbook of Experimental Psychology*; edited by S. S. Stevens (John Wiley & Sons Inc., New York, 1951), p. 1040.

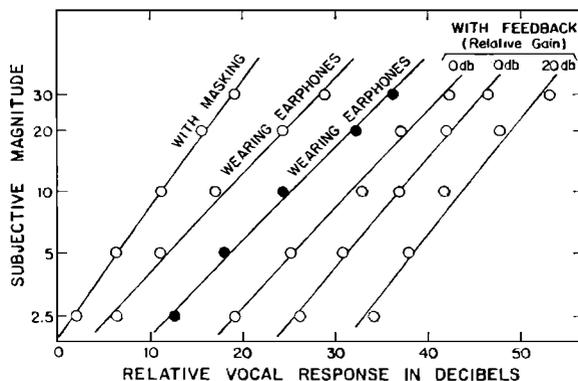


FIG. 4. Magnitude functions for the autophonic response obtained by the method of magnitude production with five conditions of auditory feedback as a parameter. Each point represents the mean level (in db) of three responses by each of 10 subjects, except that the points on the "with-masking" function are based on three responses by each of 24 subjects.

generate these intensities were dropped from the experiment. Figure 5 shows that, although the magnitude-estimation function obtained with the wide range of vocal levels is a power function over most of the 40-db range, it turns downward near threshold. In another series of experiments, the vocal productions that were estimated covered a range of 25 db. The magnitude-estimation function obtained with the narrower range of levels agrees well with the straight segments of the wide-range function in Fig. 5.

A digression is in order here concerning the sharp downward turn of the curves through the circles in Fig. 5. As Stevens and others have shown,¹² it is typical for subjective-magnitude functions to exhibit this curvature at their lower ends. A general formula descriptive of such functions relating subjective magnitude ψ to physical magnitude ϕ takes the form

$$\psi = k(\phi - \phi_0)^n$$

The value ϕ_0 can often be identified with the "threshold" that obtains under the circumstances of the experiment. On the assumption that an analogous effect occurs when the stimulus is a person's own vocalization, we could, if desired, straighten the functions in Fig. 5 by subtracting a value (ϕ_0) from each of the vocal levels.

Returning to the problem of feedback, we note in Figs. 4 and 5 that the amount of auditory feedback has a minor effect on the slope of the autophonic scale. Figure 4 shows that there is a small but monotonic increase in the slope as the amount of auditory feedback is increased. (The masking experiment constitutes an exception.) Figure 5 shows that the magnitude-estimation function also increases in slope as auditory feedback is increased.

¹² Compare S. S. Stevens, *J. Exptl. Psychol.* **57**, 210 (1959); B. Scharf and J. C. Stevens, The form of the loudness function near threshold, in *Proc. 3rd Internat. Congress of Acoustics* (to be published); J. C. Stevens and S. S. Stevens, *J. Exptl. Psychol.* **60**, 183 (1960).

Auditory feedback not only steepens the slope of the subjective function; it also changes the overall level at which a person vocalizes. (This change in over-all level as a function of auditory feedback is discussed in a later section.) When the autophonic response is assessed by the method of magnitude production, these two effects of auditory feedback turn out to be nonindependent. With decreased auditory feedback (masking), for example, the subject tends to choose a modulus whose intensity is high relative to that chosen under the standard conditions. Because there is an upper limit to the sound pressure that a subject can generate vocally, an upward shift in the level of the modulus decreases the subject's available range. If this shift in modulus is great enough, the subject can no longer produce a range of vocal levels appropriate to the designated criterion values and, as a result, the function describing the autophonic scale becomes artificially steepened.

The increase in slope resulting from a shift in the intensity of the subject's chosen modulus was most pronounced when the auditory feedback was minimal (magnitude production under masking, the steep function in Fig. 4).¹³ Under the standard conditions the subjective function has a slope of 1.17 (Fig. 1), and under masking a slope of 1.36 (Fig. 4). An analysis of the vocal-production data obtained under masking showed that this difference in the exponents for the subjective scale is correlated with an upward shift in the subject's overall intensity. The mean vocal intensity for each subject was computed and the data were separated into three equal groups: those showing the greatest upward shift in overall level from the standard to the masking condition, those showing an intermediate shift, and those showing the smallest amount of shift. The slope of the resulting function was computed for each group, and it was found that the first group, with a mean

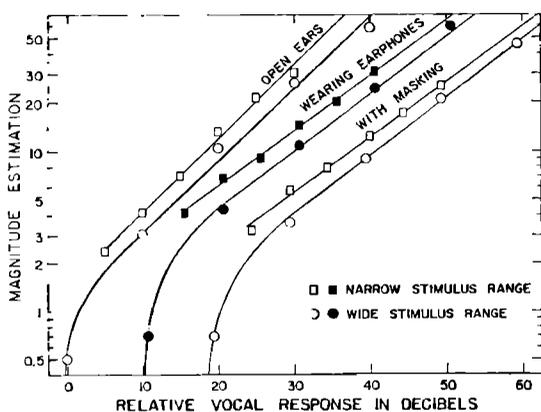


FIG. 5. Magnitude functions for the autophonic response obtained from numerical estimates of vocal productions, with three degrees of auditory feedback as a parameter. Each point represents the geometric mean of three magnitude estimations by each of 10 subjects.

¹³ The average modulus intensity produced under masking was 3.7 db higher than that produced when the ears were open. This finding is in general agreement with the data reported by Atkinson. See C. J. Atkinson, *J. Speech Hear. Disord.* 17, 419 (1952).

shift of +9.6 db, had a slope of 1.50; the second group, with a mean shift of +3.4 db, had a slope of 1.26; and the third group, with a mean shift of -2.1 db, had a slope of 1.11.

Figure 5 shows that the magnitude-estimation function, like the magnitude-production function, is steepened by an increase in auditory feedback. Since, in magnitude estimation, vocal intensity is controlled by a fixed set of criterion levels and is therefore not free to shift about, we may infer that the effect of auditory feedback is not restricted to that produced by changes in vocal intensity.

The various amounts of auditory feedback caused the slope (exponent) of the autophonic scale to vary over a range of only about 0.2, a range that is no larger than the differences produced by the various ratio-scaling methods. We may conclude, therefore, that auditory feedback plays a secondary role in the determination of the exponent of the subjective scale for autophonic level. Since the feedback from proprioception and from bone conduction presumably remains invariant when air-conducted feedback is attenuated or amplified, it is not entirely surprising that sidetone gain has little effect on a person's assessment of the relative levels of his own voice.

The minor role played by loudness in the judgment of autophonic level is further demonstrated by the finding that the loudness function for heard speech has a different exponent from the autophonic function. In one experiment, 10 listeners estimated the loudness of different playback levels of the phoneme [a] recorded at a moderate level. The measured exponent was 0.7. In a second experiment, the "sound of effort" was introduced: the same group of listeners estimated the loudness of *live* vocal productions of the phoneme [a] in which the speaker varied his voice level over a 30-db range. Under such wide changes in voice level, the quality of the sound inevitably alters, but this fact did not alter the measured exponent, which was again 0.7.¹⁴

Figure 6 gives a graphic comparison of the two loudness functions for speech (exponent 0.7), the sone scale (exponent 0.6), and the autophonic function (exponent 1.1).

EFFECT OF SIDETONE LEVEL ON VOCAL LEVEL

An earlier section described the outcome of cross-modality comparisons in which subjects tried to match their vocal level to sounds of various intensities. As a

¹⁴ In a study by I. Pollack, [*J. Acoust. Soc. Am.* 24, 323 (1952)] the level of a tape recording of spoken text was varied by subjects in order to halve and double the apparent loudness. The subjective scale obtained was not linear over its entire range when plotted in log-log coordinates, but the straight line of best fit had a slope of approximately 0.4. Pollack suggests that the departure from the sone scale (slope 0.6) may be due to speech-constancy effects, analogous to size-constancy effects in vision. Pollack's results also depart from the slope of 0.7 found for the loudness of a tape-recorded vowel phoneme. A single phoneme, however, may not resemble spoken text to an extent that would introduce speech-constancy effects.

complementary experiment, the speaker may be instructed to compensate for, rather than match, the changes in the apparent intensity of a criterion stimulus, so that some perceived magnitude is held constant. In the experiments to be reported the criterion stimulus was the "sidetone" that was fed back to the speaker's ears in an interphone system, and the speaker's task was to vocalize so as to compensate for changes in the amplification introduced into the sidetone channel. Thus the speaker was instructed to hold the loudness of his voice constant, as he perceived it, under various levels of sidetone.

Each of 10 subjects produced the phoneme [a] under nine sidetone levels, presented in irregular order. The sidetone levels were equally spaced over an 80-db range, four above and four below a reference level at which the subject judged the amplified sidetone to be equal to the natural sidetone heard when no earphones were worn. For five of the subjects the experiment was conducted first with a single earphone and then with a binaural headset. This order was reversed for the other five subjects.

The filled circles and squares in Fig. 7 show the vocal levels produced under various sidetone levels with binaural and monaural sidetone. The log-log slope of the "compensation function" for binaural sidetone is -0.46 . Had the subjects managed to hold the level of the sidetone constant, the slope (exponent) of the compensation function would have been -1.0 . The exponent actually obtained for cross-modality compensation (-0.46) is approximately equal in magnitude but opposite in sign to the exponent ($+0.5$) obtained for the cross-modality matching of autophonic response to loudness. In both compensating and matching, the subject seems to equate changes in autophonic level to changes in loudness, and this equation appears to be in accordance with the two subjective magnitude functions, the ratio of whose exponents determines the absolute value of the resultant log-log slope. The difference in the sign of the exponents for the matching and compensation functions follows from the fact that in one

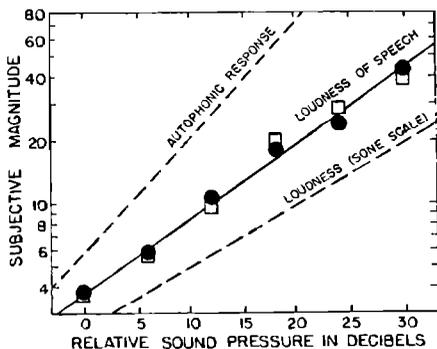


FIG. 6. Comparison of the autophonic function, the sone scale, and the loudness functions for live speech (filled circles) and recorded speech (unfilled squares). Each point on the loudness functions for speech represents the geometric mean of three magnitude estimations by each of ten subjects.

case the subject is instructed to match, in the other to compensate for, changes in loudness.

The abrupt leveling off of the compensation functions in Fig. 7 at low sidetone intensities suggests a threshold effect. There is a level below which the sidetone does not influence the level of the voice.

The function relating vocal level to monaural sidetone (Fig. 7, filled squares) has a flatter slope (-0.4) than that for binaural sidetone. This may be due to the fact that the subjective scale for monaural loudness has a slightly lower exponent (about 0.54 rather than 0.6).¹⁵ It is also possible that having one ear open to the natural acoustic sidetone leads the speaker to keep the level of his voice more constant than when both ears are subjected to the artificial sidetone.

Additional data on monaural sidetone are provided in an early study by Fletcher *et al.*¹⁶ (unfilled squares) who instructed the subjects to read monosyllables at a conversational level into a telephone microphone under different amounts of sidetone fed back through a monaural receiver.

Figure 7 also brings together the results of several studies in which subjects vocalized under various levels of binaural sidetone, with no explicit instructions to hold their voice at a constant loudness. The data presented demonstrate two points: (1) the speaker adjusts his voice by about 1 db in response to a 2 db change in sidetone gain; and (2) the compensation functions obtained are consistent with the assumption that the relation between the autophonic function and the loudness function determines the changes in vocal level that occur when the sidetone level varies.

The uppermost curve in Fig. 7 (diamonds) shows the effect of changes in sidetone level on the intensity of the subject's "modulus," i.e., the voice level he considers medium. These data are from experiments reported above on the effect of augmented sidetone on the autophonic function. When asked to vocalize at a medium level, the subject produces a sound pressure that varies inversely with the sidetone gain.

One means of attenuating the effective sidetone is to produce a temporary hearing loss. Black¹⁷ subjected his talkers to 2 hours of noise at 110 db and then measured, at 3-min intervals after the termination of the noise, both the hearing loss and the vocal level produced in the reading of phrases. As shown by the unfilled triangles, the subjects compensated for the hearing loss by raising their voices. In Fig. 7 hearing loss is plotted as increasing toward the left, i.e., in the direction of decreasing sidetone level.

¹⁵ G. S. Reynolds and S. S. Stevens, *J. Acoust. Soc. Am.* **32**, 1337 (1960).

¹⁶ H. Fletcher, G. M. Raff, and F. Parmley, Study of the effect of different amounts of side tone in the telephone set, Western Electric Company, Rept. No. 19412, Case No. 120622 (1918).

¹⁷ J. W. Black, The effect of noise-induced temporary deafness upon vocal intensity, Joint Project Rept. No. 7, Bureau of Medicine and Surgery, U. S. Navy Project No. NM 001 064.01.07 (January, 1951).

In two other studies, subjects spoke into an aircraft interphone under a constant noise level but with different amounts of sidetone. One study was a joint wartime effort by the Psycho-Acoustic Laboratory and the Aircraft Radio Laboratory to determine the performance of various kinds and arrangements of aircraft interphone components under actual flight conditions, including unpressurized flight at 35 000 ft.¹⁸ Crews of trained talkers and listeners flew more than a score of missions at Eglin Field, Florida, in the course of which some 55 000 words were read over aircraft interphones and recorded by listeners. Among other things measured was the effect of sidetone level on voice level during flight at altitudes of 5000 and 35 000 ft. One of the conclusions reached in the final report was the following.

"When the speaker adjusts his speech effort in accordance with the sound of his voice in his own earphones ('sidetone monitoring'), changing the gain of the interphone amplifier does not, in general, cause an equal change in the signal level at the headsets. Nor, in general, is the speaker's compensatory adjustment so exact that there is no change at all in the voltage across the headsets. Instead, a compromise takes place in which the level at the headsets is altered by an amount somewhere between zero and the amount by which the gain is shifted" (see work cited in footnote 18, p. 54). Signal levels produced at the interphone headset were measured for nine speakers during articulation tests with various interphone voltage gains. The headset voltages have two components, one of which is due to the gain setting of the interphone and one to the vocal level produced. The vocal levels shown in Fig. 7 (half-circles) were obtained by subtracting the changes in voltage gain introduced by the experimenter from the changes in voltage measured at the headset. The difference gives the vocal level produced by the subject. It should be noted that, for a given interphone gain, the voltage measured across the earphones was 4 to 5 db greater at 5000 than at 35 000 ft.

In a comparable laboratory experiment, Lightfoot and Morrill¹⁹ employed 16 subjects, each of whom read one articulation test at each of four levels of sidetone intensity. "Two of these levels were above and two below the 'fixed' output level of the interphone amplifier in basic training planes . . . The reader wore a service headset with conventional dynamic headphones and doughnut ear cushions. A constant level of in-circuit noise was in the headphones all of the time, comparable to that induced by an open microphone in an airplane." The speakers were not airborne. The data are shown as unfilled circles in Fig. 7.

¹⁸ J. C. R. Licklider and K. C. Kryter, Articulation tests of standard and modified interphones conducted during flight at 5000 and 35 000 ft, OSRD Rept. 1976, PB 5505 (1944).

¹⁹ C. Lightfoot and S. N. Morrill, Loudness of speaking; the effect of the intensity of side tone upon the intensity of the speaker, Joint Project Rept. No. 4, Bureau of Medicine and Surgery, U. S. Navy Project No. 001.053 (September, 1949).

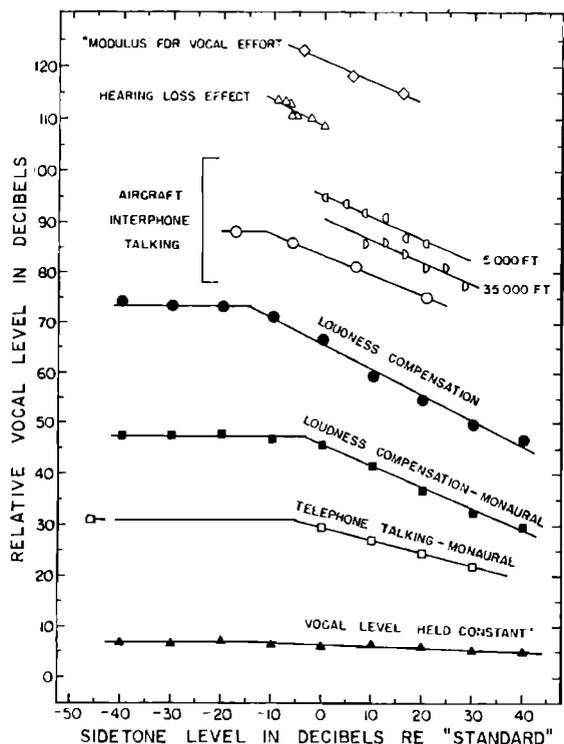


FIG. 7. The effect of sidetone level on voice level under a wide variety of conditions. Diamonds: Vocal productions of the phoneme [a] by 10 subjects instructed to produce a "medium" level. Unfilled triangles: Phrases read by 144 subjects with noise-induced temporary hearing loss. Half circles: Sentences read by nine subjects during flight-testing of aircraft interphones at 5000 and 35 000 ft. Unfilled circles: Sentences read by 16 subjects during a laboratory test of an aircraft interphone. Filled circles: Vocal productions of [a] by 10 subjects instructed to compensate for changes in binaural sidetone level. Filled squares: Vocal productions of [a] by 10 subjects instructed to compensate for changes in monaural sidetone level. Unfilled squares: Monosyllables read into a telephone microphone by 60 subjects; monaural presentation of sidetone through a telephone receiver. Filled triangles: Vocal productions of [a] by 10 subjects instructed to hold voice level constant.

An attempt was made to place the curves so that 0-db sidetone level corresponds approximately to that of normal open-ear talking.

We now come to the bottom curve in Fig. 7 (filled triangles). There we see that, despite the tendency of speakers to lower their voices when the sidetone gain is raised, most speakers can maintain a fairly constant voice level if instructed to do so. Under instructions to hold vocal level constant despite sidetone changes, the average vocal level of ten subjects changed less than 3 db under an 80-db change in sidetone gain. Comparison of the two functions, compensation and constancy, shows that, when he is instructed to hold sidetone loudness constant, the subject cannot ignore vocal effort, but when he is instructed to hold his vocal effort constant he can ignore sidetone loudness. This finding adds further support to the view that effort plays a major role in a speaker's judgment of his own vocal loudness, but that vocal loudness plays only a minor role in his judgment of autophonic output.