

## AN EXPERIMENTAL STUDY OF SPEECH PERCEPTION\*

*From the Department of Psychology of the University of Colorado*

---

T. H. HOWELLS AND J. B. SCHOOLLAND

---

Deafness was one of the earliest subjects of psychological investigation. Individual differences in reaction to auditory stimuli have always been very apparent, and also quite susceptible to measurement. Very simple techniques were adequate to demonstrate these differences, e.g., the familiar watch-tick method. In recent times the methods of measuring auditory acuity have been greatly refined and extended and a great deal of highly specialized apparatus has been produced. Instead of determining the threshold for a random, complex sound, such as a watch-tick, the present practice tends to analyze out the elementary auditory stimuli and responses, to determine the sensitivity to relatively pure tones, and to map the profile of acuity throughout the whole pitch-range. The danger in this analytic procedure, as in all reductive processes, is that it may cause one to overlook or even deny any possible combination effects.<sup>1</sup>

Along with the improvement in techniques for measuring auditory acuity has come also an appreciation of its practical importance, so that now the diagnosis of deafness is a recognized routine in child-welfare and guidance work. Much money and time has been devoted to extensive surveys of deafness of children. It seems plain that sensitivity of hearing is widely recognized as a very important factor in the genesis of personality.

### THE PROBLEM

So much attention has been given to auditory sensitivity or acuity that there is apparently a tendency to take it for granted that acuity

---

\*Accepted for publication by Carl Murchison of the Editorial Board and received in the Editorial Office, April 16, 1934.

<sup>1</sup>In an unpublished study T. H. Howells has obtained tentative evidence that the acuity threshold for certain pairs of tones is lower than for either one of them alone when the sound energy is held constant. This is implied also by the work of Wegel and Lane (12).

deficiency and auditory deficiency are one and the same thing, or in other words, that all auditory defects are acuity defects and that the terms can be used synonymously. It is a notable fact that the instruments most generally used for measuring acuity are termed audiometers. It is evident, of course, that this usage is not precise. Ear specialists are familiar with cases of word-deafness in which the patient is practically incapable of understanding the speech of those about him, although he can hear a clock tick and shows normal sensitivity when given an acuity test. It would be very much of a mistake to pronounce such hearing as normal on the basis of such a test, but the mistake has undoubtedly been made many times (3).

There are also less extreme deficiencies in hearing which are likewise not explainable in terms of deafness alone. Many people whose hearing is obviously keen enough have great difficulty in understanding speech at a talkie theatre, in a noisy factory, or at a social gathering where there is much conflicting sound. A telephone operator who was discharged because of inability to understand spoken numbers was found to be above normal on an acuity test. The fact that an increased volume of sound does not alleviate the difficulty seems to indicate that it is not due to insensitivity, but rather is the result of inability to select and organize the elements in the sounds heard, or in other words that it is due to imperception.

There is a surprising poverty of studies, either experimental or theoretical, on the subject of auditory perception. Stumpf (10), in his classic survey, discusses the problem of perception but outlines no experimental techniques. Troland (11) refers briefly to auditory perception as a problem in psychophysics. The problem is also raised by Ruckmick (7, 8) in his reviews of research in audition. Wiley (14) raises the problem of speech perception from a neurological point of view. Knudsen and Jones (4) compare perceptual and acuity explanations of paracusia, while Hall (3) cites cases of auditory imperception. The differential masking and emphasis effect of tones on each other is indicated by the work of Wegel and Lane (12). Wever (13) reviews evidence that a secondary sound affects the perceptual and acuity thresholds for other sounds. Meyer (5) denies the adequacy of explanations of audition in physical terms only. Clarke (1) observes the general inadequacy of hearing tests to diagnose hearing ability. The effect, in general, of artificial mutilation of speech, through changes in relative frequency content,

etc., has been observed by a number of people and perceptual difficulties studied. This work is described in detail by Fletcher (2). The problem of individual differences in perception of speech sounds or of the means of measuring them does not seem to have received attention.

#### THE MEASURE OF SPEECH PERCEPTION

In order to test the hypothesis of imperception as a hearing handicap a speech-perception test was devised which presented, through the medium of phonograph records, a series of speech sounds of varying difficulty in interpretation.<sup>2</sup> This test served as a means of classification of individuals with respect to their ability to understand or perceive spoken words. In addition it was necessary to measure differences in auditory acuity between these same individuals, and to determine the degree of correspondence between the perceptual and acuity factors.

The phonograph records were prepared by dictating 124 words into a recording phonograph, in much the same way that they would be pronounced before a spelling class. During this same time a barrage of conflicting noise was produced which was also recorded. This noise was obtained by shaking objects in wood boxes, by crumpling paper held near the microphone, by a confused background of speech, etc. While the interference was purposely made more intense for some of the words than for others, it was impossible, of course, to predetermine just what amount of interference there would be for any particular word. Once recorded, however, the interference remained the same always for each particular word. The general imperfection in recording and reproducing speech, which is ordinarily a fault, became a virtue in this case, since it increased the difficulty of the test. The general effect was to produce a suitable range of difficulty in recognizing the words when the records were played.

The words were familiar and unequivocal with equal proportions of one, two and three syllable forms, and arranged in random order. It should be understood that in the present instance the purpose was to provide a typical or representative variety of speech sounds and also of sound interferences. It was not the object of this study to select or limit the sound elements either in the stimulus words or in

---

<sup>2</sup>A similar technique was used by Fletcher (2) to measure the influence of frequency components etc. on intelligibility.

the conflicting sounds, since it was primarily an investigation of differences in individuals rather than situations as in the case of Fletcher (2). It was necessary, of course, that the sound situations be as uniform and as normal as possible. While no graphic analysis was attempted the recording was electrical and there were no obvious gaps in frequency components.

The test was given by playing the records in a small room before groups of university students of about fifteen each. They were arranged about equally distant from the phonograph. Outside noises were minimized and conditions were standardized in so far as possible. The volume of sound was purposely maintained at a uniformly loud level. The words had been dictated and were reproduced at four-second intervals. The subjects were asked to write down the words as heard on a form provided. The number of words correctly written, from the standpoint of sound, constituted the score. Scores were obtained for 110 subjects.

A satisfactory range of talent was obtained, approximating the normal curve of distribution. Two words only of the 124 were missed by everyone. One word was recognized by 97 per cent of the group, and two by 92 per cent. Seven words were heard once only. Of the 3700 recognitions, 26 per cent were of one syllable words, 22 per cent of two syllable words, and 52 per cent were of three syllable words. The mean score was 43 words. The highest score was 67 words and the lowest was seven, with a standard deviation of 13.2. The correlation of the scores on the odd-numbered words with the scores on the even-numbered words was .85, affording a coefficient of reliability, according to Brown's formula, of .92. It seems evident that the test was measuring some ability with considerable accuracy.

#### SPEECH PERCEPTION AND ACUITY

The next procedure was to measure the auditory acuity of the same subjects for the purpose of comparison. They were tested both by the watch-tick method and also with the Seashore electrical audiometer. The scores on each test were the result of averaging ten trial scores. These trial scores were obtained by combining the individual scores for both ears, allowing the better ear four times the weighting of the poorer one in the combination. The score for each ear was obtained by averaging the ascending and descending thresholds. In

the case of the watch-tick test the watch was mounted on a string which ran over pulleys with a scale between them, so that the distance of the watch could be varied silently and at the same rate. The coefficients of reliability, obtained by correlating the odd trials against the even trials and applying Brown's formula, was .94 for the watch-tick method, and .97 for the Seashore audiometer.

It is an interesting fact that the correlation between the scores on these two reliable acuity tests was only .41, indicating apparently that acuities are specifically related to specific stimuli, and are common, perhaps, only in so far as the stimulating elements are common.

There are no reliable norms, apparently, for either of these acuity tests. The watch-tick test obviously could not be standardized, except for a particular watch with a given adjustment in a specific laboratory situation. The range of acuity as measured by the Seashore audiometer seemed to be about normal according to norms that were available. There were, however, no subjects who were found to be extremely deaf according to either test. This would be expected, of course, since they had all been able to achieve college rank in a normal school system. This fact may indicate that the distribution would be somewhat more restricted than normal. There were, however, a few subjects who were almost completely deaf in one ear.

As was stated before one of the principal objectives of this study was to determine the part played by auditory acuity as affecting the efficiency of hearing in normal situations. The coefficient of correlation of the scores on the described test of speech perception with the scores on the Seashore acuity test was .08 for 110 cases, while with the watch-tick test the coefficient was .06. It seems apparent that, for the subjects used and for the intensity of the sound provided by the phonograph, variations in sensitivity or acuity did not appreciably affect efficiency in hearing.

#### CORRESPONDENCE WITH OTHER FACTORS

An associated problem which arises naturally in connection with the study of speech perception is that of its correspondence with other types of perception. Scores were obtained for 38 of the subjects on a test of visual perception of printed letters and words.<sup>3</sup> This test consisted of letters and words of varying degrees of blurredness,

---

<sup>3</sup>Developed by T. H. Howells in an unpublished study.

which were photographically printed upon a series of charts by means of an enlarging lantern which was adjusted so that the images were more or less out of focus, much as in an imperfect eye. The aim was to hold distinctness comparatively constant for all the subjects and to classify them with respect to ability to interpret the blurred images. The fact that the problem consisted in organizing blurred imagery seems to justify the term, perception. This test showed a reliability coefficient of .86. The correlation with speech perception scores was .07.

Scores were also available for 44 of the subjects on a test of ability to recognize faces.<sup>4</sup> In this test three photographs of the faces of 42 different persons were grouped on a card, so that there were 42 cards with three portraits of a given individual on each. Corresponding groups of three different portraits of each of these persons were placed in random order on a large chart. Subjects taking the test looked at the face on a given card for ten seconds, laid it down, and then attempted to find the same individual from among the groups of portraits on the chart. The number of correct identifications constituted the score. The portraits on the chart differed in facial expression, angle of view, size, clothing etc., from the corresponding ones on the chart, which necessitated identification in terms of facial pattern, rather than through discrete common elements. Chance factors served to bring about a suitable range of difficulty. The observed reliability coefficient of this test for 134 cases was .88.

The fact that this test, in addition to measuring perceptual or organizing ability, was concerned with stimuli which were primarily social would seem to indicate that it might have much in common with the test of speech perception. There was a correlation of .14 between the scores on these tests for 44 cases, which would seem to suggest that this is not the case.

Since the tests just described deal with visual phenomena it might be presumed that there would be little correspondence with the speech-perception scores, and that auditory measures might correlate more highly. The Seashore Tests of Musical Aptitude (7) were given to an average of 80 of the subjects who had taken the speech-

---

<sup>4</sup>Preliminary report by T. H. Howells in the Journal of the Colorado-Wyoming Academy of Science for 1933.

STUDY OF SPEECH PERCEPTION

343

perception test. The coefficients of correlation of the speech-perception scores on the various units of the Seashore test are presented in Table 1.

TABLE 1  
CORRELATIONS WITH SPEECH-PERCEPTION SCORES (Rel. .92)

Test	Reliability	No. of cases	Coefficient	P.E.
Acuity (audiometer)	.97	110	.08	.064
Acuity (watch-tick)	.94	110	.06	.064
Facial memory	.88	44	.14	.102
Reading blurred print	.86	38	.07	.106
Spelling	.90	62	.05	.086
Intensity		75	.005	.078
Consonance		86	.03	.071
Rhythm		88	.06	.071
Pitch		92	.09	.070
Time sense		65	.12	.080
Tonal memory		69	.21	.077
Grades		100	.20	.065
Intelligence	.96	71	.37	.069

It is evident that the correspondence is very slight. Only in the case of tonal memory does it approach significance.

It has sometimes been suspected that deficiency in spelling ability may be due to a perceptual difficulty, either auditory or visual or both. A preliminary form of the speech-perception test was given to eighty-four subjects who were also given a spelling list of 92 words taken from a commercial spelling book. Reliabilities were not ascertained, but there was a coefficient of .35 between the measures. Sixty-two of the subjects taking the final form of the speech-perception test were given a list of 150 words selected at random from the freshman spelling list used at the University of Colorado. The reliability of the spelling scores was .90, and their correlation with the speech-perception scores was .05. It seems that there is little in common.

There was more correspondence with the more complex and inclusive functions of grade achievement and general intelligence. The correlation of speech-perception scores with a composite of grades achieved by subjects in university courses was .20 for 100 cases. The correlation with intelligence scores obtained from the American Council Test was .37 for 71 cases. When the intelligence factor is partialled out the correlation of speech perception scores with grades becomes .10.

The data on sex differences are not conclusive. The test was taken by 46 men and by 64 women. The mean score for the men was 40.9, while for the women it was 44.8. The difference between these means was 3.9, which is 2.3 times the probable error. The difference seems suggestive, but it is hardly significant.

#### SUMMARY AND REVIEW

By way of summary it appears that:

1. The common assumption that ability to hear speech is analyzable into elementary factors, or acuity thresholds, is unwarranted on the basis of available evidence.
2. A test of speech-perception, involving the recognition of words pronounced by a phonograph, was constructed and administered to 110 subjects, and found reliable.
3. The speech-perception scores showed little correspondence with scores on two reliable tests of auditory acuity, indicating that the acuity factor in hearing has been overestimated.
4. There was little or no relationship between speech perception scores and the data from (1) a facial memory test, (2) a test of ability to read blurred print, (3) a spelling test, (4) the Seashore Tests of Musical Aptitude: namely, *consonance*, *rhythm*, *pitch-sense*, *time-sense*, and *tonal-memory*.
5. The correlation of speech-perception scores with composites of grades in university courses was .20, while with intelligence scores the coefficient was .37, indicating some elements in common in each case.
6. The data suggest, but do not prove, that women excel men in speech-perception.

The most significant outcome of the experiment seems to be the demonstration and measurement of a non-sensory, auditory function which has much to do with ability to understand conversation. Supporting this conclusion is the fact that ability to hear in the test-situation was not dependent upon auditory acuity as evidenced by a negligible correlation with two reliable acuity tests. The implication seems to be that acuity has been overestimated as a factor in hearing and that more attention should be paid to the auditory function as a totality. As a correlate of this review it is worthy of note that Reudiger (6) found similarly in the case of vision that acuity



has little to do with differences between individuals in visual perception.

It should be noted that the present study is of differences between individuals in reaction to a fairly normal pattern of auditory stimuli, while Fletcher's studies (2) were of differences in the pattern of stimuli as regards their influence upon the perception of the normal or average individual. Either, or various combinations of both techniques could be devised for the study of various problems, as, for example: (1) influence of the time (e.g., between words) allowed for the perception, (2) the influence of the amount or nature of context, or meaningful relationship, (3) the influence of different social settings when giving the test, (4) the relationship between perception and such factors as phase difference, frequency content, intensity, etc., (5) the effect of practice on ability to supply missing perceptual elements (subjective tones etc.), and of disuse or adaptation on the disappearance of these elements, or (6) of the differential effect of these elements on each other. With a properly graded series of tests, similar to the one described, it should be possible to study the genesis of speech-perception in the child, and to detect abnormality at an early stage.

The lack of inter-consistency between the various perceptual tests, especially when viewed in light of their high intra-consistency, seems to constitute one of the most perplexing problems of psychology. The evidence in this study is merely a supplement to that supplied by many others. There seems to be no common perceptual factor. Whatever slight intercorrelation there is can be explained, apparently, on the basis of a common permeating factor of intelligence.

The most obvious and practical implication of the study seems to be the need for a test, or series of tests, similar in general to the one described, which should be refined and thoroughly standardized, with different forms of the test and with different norms developed for various social and age-groups. Such a test could be provided in the form of records manufactured and distributed by one of the phonograph companies as in the case of the Seashore tests. It seems that such a test should prove of much value in the more exact diagnosis of hearing defects.

# REFERENCES

1. CLARKE, T. A. On hearing tests. *Proc. Roy. Soc. Med.*, 1929, **22**, 361-376.
2. FLETCHER, H. Speech and hearing. New York: Van Nostrand. Pp. ix+331.
3. HALL, S. B., & HALL, M. B. Auditory imperception, illustrated by description of three clinical cases. *J. Neur. and Psychopathol.*, 1931, **11**, 304-317.
4. KNUDSEN, V. O., & JONES, I. H. The problem of paracusia. *Laryngoscope*, 1926, **36**, 623-633.
5. MEYER, M. F. The delusiveness of filtering compound sounds. *Science*, 1932, **76**, 54-55.
6. REUDIGER, W. C. The field of distinct vision. *Arch. Psychol.*, 1907, **1**, 56-61.
7. RUCKMICK, C. A. Facts and theories of audition. *Psychol. Bull.*, 1928, **25**, 229-244.
8. ———. Recent research in the field of audition. *Psychol. Bull.*, 1930, **27**, 271-297.
9. SEASHORE, C. E. Psychology of musical talent. Boston: Silver, Burdette, 1919. Pp. xvi+288.
10. STUMPF, C. Tonpsychologie. (2 vols.) Leipzig: Hirzel, 1883, 1890. Pp. xviii+427, xiv+582.
11. TROLAND, L. T. The psychophysiology of auditory qualities, *J. Gen. Psychol.*, 1929, **2**, 28-58.
12. WEGEL, R. L., & LANE, P. E. Auditory masking of one pure tone by another. *Phys. Rev.*, 1924, **23**, 266-285.
13. WEVER, E. G. Effect of a secondary sound on hearing. *Science*, 1928, **67**, 612-613.
14. WILEY, L. E. The function of the brain in audition. *J. Comp. Neur.*, 1932, **54**, 109-141.

*University of Colorado  
Boulder, Colorado*

## UNE ÉTUDE EXPÉRIMENTALE DE LA PERCEPTION DE LA PAROLE

(Résumé)

On croit ordinairement que cela va sans dire qu'un test d'acuité auditive ou de sensibilité est aussi un test de la capacité d'entendre les mots parlés. Il est évident, cependant, que la difficulté que l'on trouve ordinairement à distinguer les mots parlés à haute voix mais imparfaitement, ou dans une confusion d'autres sons, n'est pas due aux oreilles insensibles, mais à l'imperceptibilité.

On a fait un test qui se compose de stimuli ordinaires de la parole de difficulté variante, en dictant des mots ordinaires et en les enregistrant sur des disques phonographiques, avec les divers sons en conflit, aussi. On a fait jouer les disques d'une manière retentissante devant 110 sujets. Le nombre des mots compris a constitué le résultat. La constance de la

mesure a été de 0,92. On a trouvé des cas importants de l'imperceptibilité. La correspondance avec plusieurs autres mesures s'est montrée légère, sauf pour une corrélation de 0,37 avec l'intelligence. Le résultat le plus significatif a été une corrélation négligeable entre les résultats de la perception de la parole et ceux de deux tests d'acuité auditive statistiquement de valeur.

On conclut que l'on a beaucoup exagéré l'acuité auditive comme facteur de l'ouïe, et que l'on devrait faire plus d'attention à la fonction auditive comme totalité. On suggère que l'on devrait construire des tests semblables et les standardiser pour divers groupes sociaux. On pourrait se servir de tels tests dans les études de la philologie comparative, ainsi que dans le diagnostic des difficultés de l'ouïe.

HOWELLS ET SCHOOLLAND

#### EINE EXPERIMENTELLE UNTERSUCHUNG DER SPRACHWAHRNEHMUNG

(Referat)

Es wird allgemein angenommen, dass ein Merkmal der Gehörschärfe oder Empfindlichkeit auch ein Merkmal der Fähigkeit sei, gesprochene Worte zu hören. Es ist aber einleuchtend, dass die Schwierigkeit, die gewöhnlich bei der Unterscheidung von laut oder mangelhaft gesprochenen Worten oder bei einer Verwirrung von anderen Lauten entsteht, nicht die Folge unempfindlicher Ohren ist, sondern einer mangelhaften Wahrnehmung.

Eine Prüfung, die aus gewöhnlichen Sprachreizen von verschiedenartiger Schwierigkeit bestand, wurde durch das Diktieren von gewöhnlichen Worten und deren Aufzeichnung mit verschiedenen widerstehenden Lauten auf Grammophonplatten hergestellt. Die Platten wurden vor 110 Vpn. vorgespielt. Die Anzahl der verstandenen Wörter ergab den Wert für die Prüfung. Die Zuverlässigkeit des Masses war 0,92. Hervorragende Fälle der mangelhaften Wahrnehmung wurden entdeckt. Die Uebereinstimmung mit einigen anderen Massen war gering, ausser einer Korrelation von 0,37 mit der Intelligenz. Der bedeutsamste Befund war eine geringfügige Korrelation zwischen den Sprachwahrnehmungswerten und den Werten von zwei statistisch zuverlässigen Prüfungen der Gehörschärfe.

Es wird gefolgert, dass die Gehörschärfe als ein Faktor des Hörens sehr überschätzt worden ist und dass mehr Aufmerksamkeit der Gehörfunktion als Ganzheit zugewandt werden sollte. Es wird vorgeschlagen, dass ähnliche Prüfungen aufgestellt und standardisiert bei verschiedenen Sozialgruppen werden sollten. Solche Prüfungen könnten auch in Studien der vergleichenden Philologie sowie in der Diagnose von Gehörschwierigkeiten verwandt werden.

HOWELLS UND SCHOOLLAND