Japanese Psychological Research
1967, Vol. 9, No. 2, 70-77

THE ANALYSIS OF ENGLISH HEARING ABILITY IN TERMS OF SIGNAL DETECTION MODEL[1]

IKUKO OGASAWARA

University of Tokyo

This study was designed to determine how well Japanese students can evaluate the intelligibility of their own English hearing reception, applying the detectability measure of the signal detection model. 4 types of English hearing tests were given to the *Ss*: Aural perception (Test 1), words (Test 2), word in sentences (Test 3) and meaning (Test 4). The mean detectability score for Test 1 was lowest ($d'=0.742$) and for Test 3 was highest ($d'=1.403$). In Test 3 and 4, the relation between the detectability score and the achievement score was significant at 1% level, but it was not significant in Test 1 and 2. The data were analysed in terms of ROC curve. The result suggested some complex relationship between the achievement score and the detectability score, and the discussion was made on the possible factors affecting the detectability score.

During World War II, engineers developed a theory of detection which would apply to the detection of targets by radar receivers. Tanner and Swets (1953, 1954) and Smith and Wilson (1953) built a mathematical model and applied it to human detection. Signal-detection theory is a special case of the statistical decision theory. A detection involves a decision based on statistical considerations rather than a simple statement of the form, " Yes, I heard the signal," or " No, I did not hear the signal ".

Detection theory regards the experiment which attempts to determine the sensory threshold as a game between the subject and the experimenter. According to this view, the subject would always say he heard the signal, whether he heard it or not, if he knew in advance that it would be presented. This does not mean that subjects should simply be regarded as dishonest; it means that there is no sharp division between detecting a signal and not detecting a signal. So the subject is always forced to make decision based on probabilistic rather than certain information. How well can an ideal observer do in discriminating between signal plus noise and noise alone? Detection theory shows that the optimal performance possible depends on the relationship between the energy of the signal (E) and the power per unit bandwidth of the noise (N). To be more specific, the distance in standard deviation units between the means of the distribution of likelihood ratios for signal plus noise and noise alone on a logarithmic scale, is the square root of $2\,E/N$. This distance between the distributions is called d'.

The d' values so far obtained for human observers have shown good consistency for a given individual over a considerable range of values of a priori probability and pay off matrix. It is this consistency that represents perhaps the greatest contribution of the detection theory, for this consistency cannot be predicted by the traditional theory based on thresholds. The thre-

[1] The auther is indebted to Prof. Y. Takada of Tokyo Woman's Christian College and Prof. S. Shiba of Seikei University for their very helpful suggestions and advices for this research. Particular thanks are also due to Prof. J. D. Miller of Central Institute for the Deaf in St. Louis, Missouri, U. S. A., for the personal communication.

shold has been found to vary with conditions. Psychophysicists were coming to regard thresholds as significant only under the specified conditions in which the thresholds had been determined. The d' measure with its greater generality escapes this limitation. The d' determines a receiver operating curve or ROC. The ROC has been a useful device for predicting results in both vision and audition (Carterette and Cole, 1962); this could indeed be said for the whole of detection theory.

In many of the experimental procedures that were developed to study speech communication (Eagan and Clark, 1956 ; Pollack and Decker, 1958 ; Green and Birdsall, 1964 ; Clark, 1960), using the analytical techniques of theory of signal detection, the speech signals were mixed electrically with white noise. The signal-to-noise ratio was computed in each experimental condition. The receiver attempted to detect the signals over a noisy channel. The receiver could adopt a certain criterion of signal acceptance and after recording his response he could estimate with this criterion whether or not that his response was correct. The receiver operating characteristic, or ROC, curve shows the relation between the probability that the receiver will accept his response when in fact he heard the message correctly and the probability of an acceptance when in fact he recorded the message incorrectly.[2] The curve is generated by requiring the receiver to adopt various criteria for acceptance on successive tests conducted at the same signal-to-noise ratio (Peterson and Birdsall, 1953; Tanner and Swets, 1954). By using the concept of the ROC curve, it is possible to compare various methods of message transmission in terms of their efficiency. This procedure would be avairable not only for the physically noisy transmitting channel, but for the situations where some psychological noise or uncertainty is involved.

In speech communication a person often evaluates the accuracy with which he receives words, sentences, and concepts as well as the accuracy with which he himself utters words and sentences, and uses these evaluation to communicate the concepts he experiences. If these evaluations are made earnestly, corrective action, (which takes the form of repetition, rephrasing or requests for repetition and rephrasing) may be taken. Whether or not communicating parties make evaluations about the effectiveness of the communication in which they are involved depends upon their ability to detect errors and, therefore, determine the need for corrective action. Corrective action, however, will generally be taken only when the value of accurate communication is high and the communicating parties are willing to admit that corrective action is needed.

A person's ability to detect errors in his own speech prediction depends, in part, upon his awareness of what constitute the " correct " or standard manner of producing the sequence of phonemes which make up the intended utterance as well as his ability to compare what he says with what he feels to be standard. A person's ability to detect receptive errors in communication seems to depend upon the function of his hearing mechanism and comparing processes which seem to be of a nature as those mentioned above. In the previous study (Ogasawara, 1966) it was reported that in communication between Japanese talkers and American listeners, the index of detectability was useful measure of their own speech evalua-

[2] Eagan and Clark (1963) and Pollack (1959) distinguish between two indices. The first is the index of signal discriminability, d_s which reflects the extent to which items of the information source are discriminated. The second is the index of response discriminability, d_r which reflects the extent to which a listener can discriminate the correctness of his own responses to the information source. Clarke, Birdsall, and Tanner (1959) have called the ROC curve based on response-response contingencies a " Type II ROC curve ".

TABLE 1

Five ratings listeners used for their identification of the messages

5.	I am certain I understood the message correctly.
4.	I am fairly certain I understood the message correctly.
3.	Undecided.
2.	I am fairly sure I did not understand the messsage correctly.
1.	I am certain I did not understand the message correctly.

tion. The present study was planned to explore further how Japanese listeners got the message from American talkers and evaluate their own reception. The method of data analysis was follows:

EXPERIMENT

Purpose :

The general purpose of this experiment is to investigate the nature of detectability of several English hearing tests. Specific objectives include measuring detectability and other parameters of ROC curve of four types of hearing tests, i.e., tests of perception, words, words in sentences, and meaning of sentences.

Procedure :

The general testing procedure employed was identical to the standard English intelligibility-test procedure with one addition : After the listener responded to each test item, he reported his judged accuracy of his reception. The rating scale with five categories shown in Table 1 was employed.

The 100 items for each test (except Test 4) were tape recorded by an American talker and presented to Japanese *S*s. About 5–8 sec separation between items permitted sufficient time for the listeners to record the answer and to record his accuracy rating.

About 120 university freshmen (Japanese) served as listeners. They were divided into 4 groups and tests in the hearing room. All the *S*s took all tests and the testing orders to the groups were randomized so that the effect of order was eliminated. It took about 30 min to finish each test.

Test 1 : The material used was " English aural perception test for Japanese students " by Lado. *S*s discriminate the difference of 3 short sentences.

Example : It was wrong.
It was wrong.
It was long.

Test 2 : The speech material used consisted of 50 monosyllabic English words, 25 words which were " easy " for Japanese people to pronounce and to understand, and 25 " difficult " words for Japanese people to say and understand. The criteria for selection of an easy word include that the word be familiar, that the word be composed of speech sounds found in the Japanese language or speech sounds which require movements similar to those found in the Japanese language, and that the word be easy to discriminate from other words. (Examples : you, high, day, key.) The criterion for selection of a difficult word is that the word did not meet the criteria described above for the selection of easy words. (Examples : run, light, rail sin.) Two lists of these 50 words with random order were given to the *S*s.

Test 3 : The speech materials used consisted of 50 short English sentences. The act of words which those sentences ended with was the same set used in Test 2. The listener recorded only the last word of each sentence and made judgment about his accuracy. Two random order lists of these 50 short English sentences were presented to the *S*s.
(Example : She has the *key*.)

Test 4 : The speech material used consisted of 56 short English sentences. There were 3 different categories of these sentences ; the statement is true statement, the statement is

TABLE 2
Results of Test 1, 2, 3 and 4

	Test 1	Test 2	Test 3	Test 4
1. No. of *S*s	121	104	125	123
2. Mean of d'	0.742	1.125	1.403	1.287
3. Mean Achievement Score (%)	45.5	52.6	49.3	58.9
4. Correlation between 2 and 3	0.099	0.013	0.232**	0.376**

$^{**}p<.01$

false statement, and nonsense statement. For example, " The sun is larger than the moon.", " milk is red.", and " pig pog pin." There were 24 true statements, 26 false statements and 6 nonsense statements. The listener circled the T if the statement was true, F if the statement was false, and DK if the statement was nonsense on the answer sheet. Two random order lists of these 56 short English sentences were presented to the *S*s.

Results

Means of d', means of achievement scores and correlations between d' and achievement score were summarised in Table 2. The exact method of analysis of data was developed from the theory of signal detection (Eagan, 1963). The mean of d' became higher from Test 1 to Test 3, but d' of Test 4 was lower than d' of Test 3.

The correlation between d' and achievement score in Test 3 and Test 4 was significant at 1% level, but it was not significant in Test 1 and Test 2.

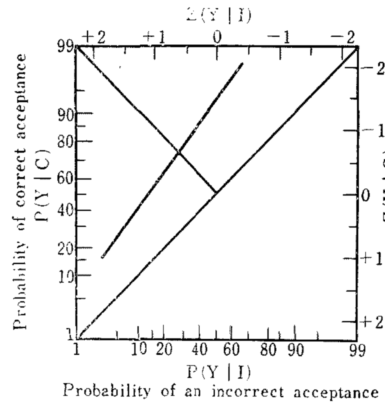

FIG. 2. Mean operating characteristics for 104 *S*s in the word hearing test. Slope $s=1.384$ and $d_r=1.294$

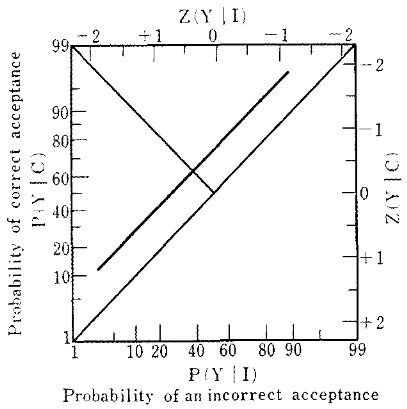

FIG. 1. Mean operating characteristics for 121 *S*s in the aural perception test. Slope $s=0.981$ and $d_r=0.733$

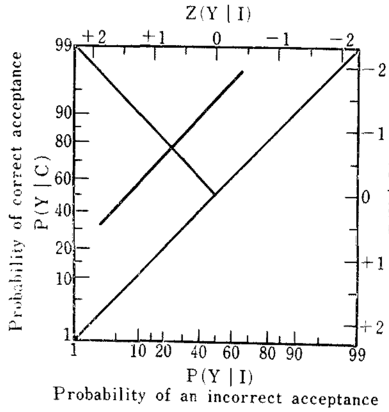

FIG. 3. Mean operating characteristics for 125 *S*s in the sentence hearing test. Slope $s=1.128$ and $d_r=1.416$

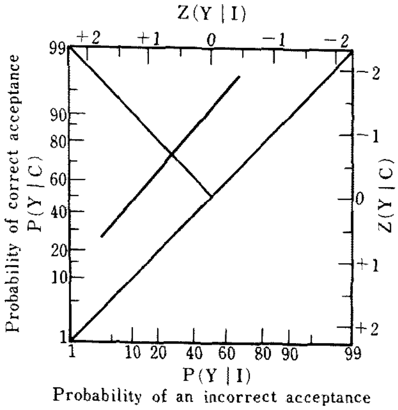

FIG. 4. Mean operating characteristics for 123 *Ss* in the True and False test. Slope $s=1.114$ and $d_r=1.355$

ROC curves of four tests were shown in Fig. 1, 2, 3 and 4. The parameters of the cureves were obtained by compiling individual operating characteristics of *Ss* for each test and then finding the means. These were Type II ROC curve plotted on normal-normal probability paper. The ordinate shows the probability of the listener's acceptance of his identification response as correct, given that his identification response was, in fact, correct. The abscissa indicates the probability of a listener's accepting of his identification response as correct, given that it was incorrect. Two parameters describing each ROC curve d_r and s were also shown. The diagonal line connecting the upper right and the lower left corners of the graph represents chance performance. For a description of the analysis used on the data of the present study, the reader is referred to Pollack and Decker (1958).

DISCUSSION

Achievement Scores

The mean achievement score of each test in Table 2 must be interpreted in light of the facts that there were 100 items in each test (except Test 4) and that the listeners could not view the talkers while listening to the messages. Since our 4 tests were not standardized, we can not compare the scores of each test.

Miller, Heise and Lichten (1951) reported the value of the sentence as context for the correct perception of words, and the hearing scores of words in sentences were better than same words in isolation. In Test 3, the listener's task was the identification of words in same as those in Test 2. In this situation, the listener would likely judge the whole meaning of sentences and thereby identify the message. Because the elimination of unlikely possibilities occurs so quickly and so automatically, it is difficult to imagine how the process takes place. The nature of the situation implied by the whole sentence somehow influences what a listener expects and from the relatively limited range of expectations he chose the one that seems to him most probable.

In our experiment, the context did not affect the achievement score of Test 3. For the listener who does not understand the meaning of sentences, the context of sentences do not affect his achievement at all. In this case, the word hearing test (Test 2) may easier than Test 3, because he concentrates to catch each word item itself. The effect of context may be depend on the English ability of listeners.

Listener's Confidence Rating

Pollack and Decker (1958) reported that the assignment of confidence ratings does not interfere with the accuracy of message reception. In addition, the confidence rating is directly related to the average accuracy of message reception and this relationship is relatively invariant over a range of signal to noise ratios.

Fig. 5, 6 and 7 present the results of analysis of word hearing test. The ordinate Fig. 5 is $p(R_j)$— the probability of employing a selected response category, irrespective of correctness of message reception. The abscissa of Fig. 5 to 7 is the

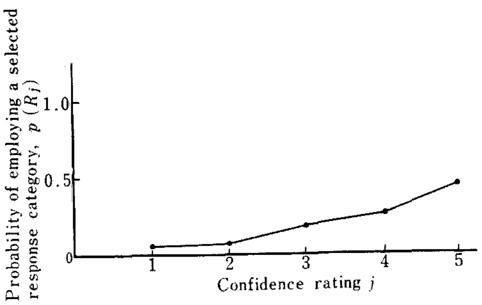

FIG. 5. Confidence ratings assigned at word hearing test. The abscissa is the confidence rating by the listener. The ordinate is the relative frequency of category assignment, irrespective of correctness of message reception. Average results for 104 listeners.

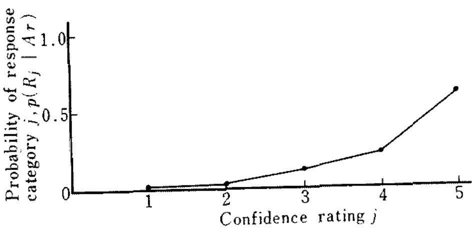

FIG. 6. Confidence ratings assigned to correct message reception only, at word hearing tests. The abscissa is the confidence rating assigned by the listeners. The ordinate is the relative frequency of category assignment among the correct message receptions. Average results for 104 listeners.

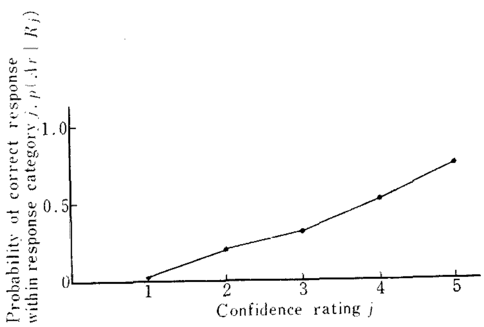

FIG. 7. The probability of a correct message reception within given rating categories at word hearing test. The abscissa is the confidence rating assigned by the listeners. The ordinate is the proportion of correct message receptions of the total number of responses assigned each rating category. Average results for 104 listeners.

rating category employed. The ordinate of Fig. 6 is $p(R_j|A_r)$, the conditional probability of employing a selected response category, given that the message was received correctly. The net effect of these two results yields the relationship of Fig. 7. The ordinate of Fig. 7 is $p(A_r|R_j)$— the probability of receiving a word correctly, given that the word were assigned a selected rating category. From these results, *Ss* apperantly employ the good approximation of assigning their confidence ratings to achieve fixed average levels of message reception, and these results are coincident with the case of higher signal to noise ratio in the experiment of Pollack and Decker (1958). The confidence rating of Test 1, 3 and 4 showed the similar tendency to the result of Test 2.

Relation Between Detectability and Achievement Scores

This experiment is an initial attempt to measure the detectability of the Japanese student to make prediction of the intelligibility of their hearing of English test. The interpretations of the results must be tempered by the facts that the listeners were freshmen of a university and did not have any previous training for the experiment.

Operating characteristics curve for listeners in Fig. 1 (Test 1) shows that *Ss'* ability to judge accuracy of their own reception is relatively close to the chance level. In Fig. 2 (Test 2), the slope of the ROC is steeper and ROC shows better predictive ability than in Test 1. In Fig. 3 (Test 3) ROC has the slope close to one and shows best prediction ability among four tests. Fig. 4 (Test 4) indicates relatively good predictive ability.

The results of our experiment revealed the complex relationship between achievement score and detectability score. In Test 1, aural perception test, the task required of the listener was the discrimination of English sounds and not necessarily the understanding the meaning of message.

The correlation coefficient between achievement score and detectability score was not significant. In Test 2, the correlation coefficient was not significant at this stage too. In Test 3, the listener's task was the identification of words in sentences. The correlation between achievement score and detectability score of Test 3 was significant at 1% level, but it was very low. In Test 4, the listener's task was to discriminate true from false meaning of sentences. At this stage, the correlation coefficient was significant at 1% level but it was very low. From all these results, we assume there is a complex relationship between achievement score and the detectability score and the relationship varies with the testing material.

In learning process, detectability appears to be an independent measure from achievement score, because there is only very low correlation between them. The field of education and ability assessment abounds with questions related to the evaluation of achievement and the detection of errors of the learner. The ability to detect errors and to make self-evaluation should be an important factors of learning in educational situation. The measurement of this ability has not yet been tried in to educational fields. Conventional achievement tests evaluate only the achievement score of students. In the process of learning, there must be different stages of progresses for the learner to evaluate himself accurately. The learning methods for each learner should be different, depending on this ability. Perhaps detectability involves the interaction of multiple situational factors, such as the nature of material, the attitude and personality of listeners, their developmental stages, and factors relating to the cues of judgment. How these factors cooperate in determining self-evaluation is not yet clearly known. Exploring how these factors are mutually related in detectability and clarifying the process of self-evaluation in learning will be our future problem.

REFERENCES

BIRDSALL, T. G., & ROBERTS, R. A. 1965 Theory of signal detectability: Deferred-detection theory. *J. Acoust. Soc. Amer.*, **37**, 1064–1074.

CARTERETTE, E. C., 1958 Message repetition and receiver confirmation of message in noise. *J. Acoust. Soc. Amer.*, **30**, 846–855.

CARTERETTE, E. C., & COLE, M. 1962 Comparison of the receiver-operating characteristics for messages received by ear and by eye. *J. Acoust. Soc. Amer.*, **34**, 172–178.

CLARK, F. R. 1960 Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. *J. Acoust. Soc. Amer.*, **32**, 35–46.

CLARK, F. R., BIRDSALL, T. G., and TANNER, W. P., Jr., 1959 Two types of ROC curves and definitions of parameters. *J. Acoust. Soc. Amer.*, **31**, 629–630.

DAVIS, H., & SILVERMAN, S. R. (Eds.) 1960 *Hearing and deafness.* New York: Holt, Rinehart, and Winston, Inc.

EAGAN, J. P. 1957 Monitoring task in speech communication. *J. Acoust, Soc. Amer.*, **29**, 482–489.

EAGAN, J. P., & CLARK, F. R. 1956 Source and receiver behavior. *J. Acoust. Soc. Amer.*, **28**, 1267–1269.

EAGAN, J. P., & CLARK, F. R. 1963 *Psychophysics and signal detection, in experimental method and instruments in psychology.* J. B. Sidowski, Ed., New York: McGraw Hill.

GREEN, D. M., & BIRDSALL, T. G. 1964 The effect of vocabulary size on articulation score. *Operational Applications Laboratory of the U.S. Air Forch. Tech. Repts*, No. 81.

GREEN, D. M., & SWETS, J. A. 1966 *Signal detection theory and psychophysics.* New York: John Wiley.

HATORI, H. 1965 Relation between some English abilities. Tokyo Teacher's College: *University Bulletin*, No. 17, 71–76.

LADO, R. 1959 *Test of aural perception for Japanese students, in "Linguistic science and language tests".* Taishukan Tokyo.

MILLER, G. A., HEIGE, G. A. & LICHTEN, W. 1951 The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.* **41**, 329–335.

MURDOCK, B. B. 1965 Signal-detection theory and short-term memory. *J. exp. Psychol.* **70**, 443–447.

OGASAWARA, I. 1966 Analysis of speech communication by signal detection. *Proc. the Meeting of J.E.P.A.* (In Japanese).

OGASAWARA, I. 1967 Analysis of speech communication by signal detection. (*Dokkyo University: University Bulletin No.* 1.) 175–193.

PETERSON, H. W., & BIRDSALL, T. G. 1953 The theory of signal detectability. *University of Michigan: Electronic Defense Group, Technical Report.* No. 13.

POLLACK, I. 1959 Identification of elementary auditory displays and the method of recognition memory. *J. Acoust. Soc. Amer.* **31**, 1126–1128.

POLLACK, I., & DECKER, L. R. 1958 Confidence ratings, message reception, and the receiver's operating characteristic. *J. Acoust. Soc. Amer.* **30**, 286–292.

POLLACK, I., RUBENSTEIN, H., & DECKER, L. R. 1959 Intelligibility of known and unknown message sets. *J. Acoust. Soc. Amer.* **31**, 273–279.

SCHERER, G. A. G. & WERTHEIMER, M. 1964 *A psycholinguistic experiment in foregin-language teaching.* New York: McGraw-Hill.

SMITH, M., & WILSON, E. 1953 A model of the auditory threshold and its application to the problem of the multiple observer. *Psychol. Monogr.* 67.

SAWADA, K., HIDANO, T., JIMBO, S., & HATORI, H. 1960 A study of understanding of English words from the viewpoint of educational psychology. *Jap. J. of Educ. Psychol.* **8**, 74–81.

SWETS, J. A. (Ed) 1964 *Signal Detection & Recognition by Human Observers.* New York: John Wiley.

TANNER, W., & SWETS, J. A. 1953 A new theory of visual detection. *Electronic Defense Group, Univ. of Michigan, Technical Report.* **18**.

TANNER, W. P., & SWETS, J. A. 1954 A decision-making theory of visual detection. *Psychol. Rev.* **61**, 401–409.

VEGELY, C. A. 1963 Monitoring of monosyllabic words of deaf children. *M. A. thesis, Washington Univ. St. Louis.* Missouri U.S.A.

(Received May 11, 1967)