C.A.L. Internal Report

LIBRARY COPY

COMMONWEALTH OF AUSTRALIA

Department of Health

COMMONWEALTH ACOUSTIC LABORATORI

Internal Report C.A.L. No. 5

Date: February, 1968

The Development of a speech hearing test using recorded sentences

by

P. Dudley

C.A.L.

Commonwealth Acoustic Laboratories

Sydney

Australia

R 534.771 a.s: 616.28-053.0

COMMONWEALTH ACOUSTIC LABORATORY LIBRARY
No.

COMMONWEALTH ACOUSTIC LABORATORIES

DEPARTMENT OF HEALTH

THE DEVELOPMENT OF A SPEECH HEARING TEST USING RECORDED SENTENCES

by

P. DUDLEY

COMMONWEALTH ACOUSTIC LABORATORY LIBRARY
No.

C.A.L. Internal Report No. 5

Address : No. 5 Hickson Road,
MILLER'S POINT - N.S.W.

THE PROBLEM

The object of the present study was to develop as a clinical tool a speech hearing test using recorded sentences. The need for some such speech hearing test, using simple material with a high degree of redundancy, arose from the new responsibility of the Commonwealth Acoustic Laboratories in administering a hearing-aid scheme for pensioners.

Speech tests of one kind or another are used for the following purposes:--

1. As a test of the degree of practical hearing impairment, independently of the pure-tone audiogram, which is considered the primary audiological test. In this function, speech hearing tests have historical precedence over pure tone tests and have the additional advantage of higher face validity. The "ideal" speech test for this purpose is one in which the task required of the subject is similar to that experienced in practical listening tasks of importance to the subject. The test described here comes closer to this ideal than most speech tests for the population in which it will be used.

2. As a check on the pure-tone audiogram. For many old people, speech hearing tests are less demanding than pure-tone threshold tests. Speech testin sometimes reveal that hearing loss for pure tones has been exaggerated. The ideal test for checking the pure-tone audiogram would be one with a high correlation with pure-tone thresholds. Sentence tests provide this; but wheth other speech tests may be better suited is not investigated in this study.

3. As a diagnostic test of the ability of the auditory systems to process complex material. The ideal test for this purpose would appear to be one consisting of material with a wide range of complexity, not necessarily speech, simila to the range of graded items in conventional tests of other abilities. However, the tests in use for this purpose such as recorded PBM lists are designed to be homogeneous in content. While this works over a wide range of discriminatory ability special tests are required for subjects who score too hi on too low on PBM tests. At the high end, the ceiling may be extended by using nonsense syllables or distorted speech. At the low end, an easier discriminatory task is required. Digits are sometimes used; sentence materia is also appropriate, but this application of sentence tests has not been explored in the present study.

4. As a guide to rehabilitative procedures. In particular, the selection of a hearing aid and prediction of the benefit to be gained from it are usuall based on the results of speech hearing tests. However, the applications of speech tests to these ends are mot limited than appears at first glance. While it is possible to devise test procedures which apparently measure the amount improvement in speech hearing obtainable from the use of a hearing aid there is evidence that among a big class of older hearing-impaired people, these do not predict whether or how much the aid will be worn or how much bene is subjectively experienced. (Dudley and Macrae 1967). Free field speech te

with and without a hearing aid, may be used to demonstrate to a potential hearing aid user the benefits and disadvantages of an aid. Either sentenc or "cold running speech" appear more suitable in this application than material consisting of isolated words.

This study is concerned only with the first two of these purposes; it describes the standardisation of a procedure which yields a "threshold for sentences" which can be used either as an independent measure of practical hearing impairment or as a check on the pure-tone audiogram.

The Test Materials.

The sentence lists used were those developed by Dr. D.B. Fry, Professor of Experimental Phonetics at London University. The sentences are grammatica simple and contain only familiar words. The sentences are intended to be unrelated to each other in meaning; in making the present recordings two slight alterations were made to the original order of the sentences because of apparent connections in meaning between consecutive sentences.

Each list of 25 sentences contains 100 "main" words, providing an alternati scoring basis; but in this study, scoring was on a whole sentence basis, ri or wrong.

Recordings were made of the sentences on tape. The speaker was John West of the Australian Broadcasting Commission whose voice was also used in the original C.A.L. recordings of PBM lists. Voice level was monitored during recording and timing was assisted by a Grason-Stadler interval timer which flashed a cue light every 7½ seconds. For convenience in administering th tests, the sentences were re-recorded on tape cartridges for use in an R C A Stereo tape playback unit. A cue signal was recorded on another ta channel so that a light on the tape playback lit immediately after each sen to indicate to the tester when to expect a response. Another channel carr noise chosen to simulate the "cocktail party" effect, but this was not use in the present study. A 1000 Hz calibrating tone was recorded on each ta cartridge at a level corresponding to the average peak level of the senten recording.

All intensities referred to in this report are the sound pressure levels obtained in a 6 cc coupler from this calibrating tone.

PROCEDURE

Subjects

Subjects were 46 people attending the N.S.W. Branch of the Commonwealth Ac Laboratory. No attempt was made to achieve homogeneity; subjects includ ex-servicemen, dockyard employees, schoolchildren and army entrants. Age range was 14 to 86 years. A number of cases with normal and better than normal hearing are included; the deafest had an average hearing level of 79 dB (1951 A.S.A. calculated over the 3 octave range - 500-4000Hz). In 28 subjects, both ears were tested and the results for the two ears treate separately in the analysis, giving a total of 74 ears tested. Subjects were excluded if their spoken English was inadequate for the test, or if masking was required in the non-test ear. A few cases were excluded bec of apparent exaggeration of hearing impairment.

Test procedure.

Pure tone thresholds were established using a C.A.L. D3 audiometer and a standard "descending" technique.

The speech tests were presented monaurally through the earphones of the audiometer. The output from the tape playback unit was taken to the tape input of the audiometer. Calibration was maintained by checking the acoustic output for the 1000 Hz calibrating tone in a 6cc coupler at the beginning and end of the experiment and the electrical output across the earphone was checked on the built-in meter before each test. The attenuator range on the audiometer allowed any level of presentation up to 130dB SPL; the range actually required was 20dB to 125dB SPL.

Instructions to the subject were as follows:

"This is a test of your hearing for speech. I will play you tape recordings of short sentences, like 'The boy broke the chair'. I want you to repeat each sentence as you hear it. Some of the sentences will be too quiet for you to hear properly, but say as much as you can of each sentence."

The instructions were repeated or further explained if required.

The first sentence was presented at an arbitrary level somewhat above the subject's presumed threshold. After each sentence correctly repeated the level was reduced 5dB; after each sentence not correctly repeated, the level was increased 5dB. Errors, scoring, and level of presentation were recorded on a scoring sheet consisting of a printed copy of the sentence list.

Scoring criteria were somewhat lenient. Errors in tense, number, or unimportant words were recorded, but the item was not scored wrong unless a significant word was missed or misreported or the sense of the sentence lost. Some examples follow:

Item	Response	Score
The police feared a riot	The police cleared the riot	Wrong
The woman made the cakes	The woman made a cake	Right
The teacher marks the exercises	Teacher marked the exercise	Right
The child was shown a picture	The child showed the picture	Wrong

The up-and-down record of level of presentation resembles a Bekesy audiogram, and an approximate threshold can be read from it by eye. More accurately, the average of the peaks can be easily calculated. This was the procedure adopted in the present study to obtain the measure referred to as "sentence threshold." By ignoring the early part of each test record, and counting the same number of high and low peaks, a stable estimate is obtained of the 50% point on the articulation curve for this material.

Method of Analysis.

The test procedure yielded for each subject a set of pure-tone thresholds (at 250, 500, 1000, 2000 and 4000 Hz) and a sentence threshold. The subject's age constituted a seventh variable. Product-moment correlations were calculated between these variables. Regression equations were calculated predicting the sentence threshold from various combinations of these variables.

Results.

Some practical difficulties in administering the test were encountered: In a few subjects with particularly poor discrimination, a stable estimate of threshold was not given by this method. Data from 3 ears was discarded for this reason. For the procedure to work, it is necessary that the subject should be able to score near 100% at some level not too far above his threshold. This condition is met in most subjects but some modification of the procedure will be required for the remainder.

Scoring produced some difficulties; although a complete item analysis was not made, there were a number of errors so sommon that it may prove better to ignore them in scoring. In particular:

List	Item	Usual error
A	21	"use" for "choose"
B	19	"left" for "let"
C	16	"Fowls" for "cows"
D	23	"Fear" for "feel"
E	24	"band" or "man" for "van"
J	19	"extensive" or "effective" fo "expensive".

The test procedure appears to be accepted as valid by the subject, and is not over-demanding of the tester or the subject. However, if complete recording of responses and levels of presentation are made, a good deal of attention is required to avoid errors. One benefit from making complete records is that the articulation function can be derived approximately, with a small number of trials. The normal articulation function with this material is, of course, very steep, and with the subjects in this study, scores rose from near 0% to near 100% in 10-15 dB in most cases.

The product-moment correlations between the seven variables recorded for each subject are given in Table 1.

	TH_{500}	TH_{1000}	TH_{2000}	TH_{4000}	AGE	$TH_{Sentences}$
TH_{250}	.935	.809	.565	.442	.392	.813
TH_{500}		.898	.628	.478	.360	.862
TH_{1000}			.743	.570	.359	.916
TH_{2000}				.826	.465	.848
TH_{4000}					.576	.717
AGE						.465

TABLE 1. Product moment correlations between the variables age, sentence threshold, and pure-tone thresholds at 5 test frequencies.

It is seen from this table that, as expected, the best predictor of sentence threshold is the pure-tone threshold at 1000 Hz. Multiple regression equations were derived which predict the threshold for sentences from various combinations of the other variables. These equations are given in the following table together with the corresponding multiple correlation coefficients and standard errors of prediction.

Equation No.	Variables	Prediction Equation	Correlation or multiple Correlation R	Standard Error of Prediction dB
1	TH_{500} alone	$TH_5 = .92\ TH_{500} + 39.8$	.86	9.8
2	TH_{1000} alone	$TH_5 = .79\ TH_{1000} + 40.0$	.86	9.8
3	TH_{2000} alone	$TH_5 = .71\ TH_{2000} + 36.2$	.85	10.2
4	TH_{500} and TH_{1000}	$TH_5 = .22\ TH_{500} + .64\ TH_{1000} + 37.8$	.92	7.5
5	TH_{1000} and TH_{2000}	$TH_5 = .55\ TH_{1000} + .31\ TH_{2000} + 34.7$	.95	6.0
6	TH_{500} and TH_{2000}	$TH_5 = .58\ TH_{500} + .43\ TH_{2000} + 33.9$	.95	6.1
7	TH_{500} TH1000 and TH_{2000}	$TH_5 = .30\ TH_{500} + .32\ TH_{1000} + .34\ TH_{2000} + 34.2$	.96	5.6

TABLE 2 : Regression equations, multiple correlation coefficients and standard errors of prediction, to predict threshold for sentences from various combinations of thresholds at 500, 1000 and 2000 Hz designated by frequency.

The errors in prediction obtained using equation 7 are still positively correlated with thresholds at 250 and 4000 Hz and with age; but the correlations are small and do not approach significance according to the usual criteria.

The coefficients in equation 7 may be rounded off for convenience in calculations with only a small sacrifice in accuracy, leading to the "three-octave average" as a predictor of speech threshold.

$$\text{Equn. 8 : } TH_5 = \tfrac{1}{3}(TH_{500} + TH_{1000} + TH_{2000}) + 38$$

In this equation, as in others given, the final term is chosen so that the average error of prediction is zero for this sample. Equation 8 should not be taken to imply that 38 dB SPL is in any sense a "normal" speech threshold.

Applying equation 8 to the experimental data gives as good a prediction as the derived regression equation, equation 7. The correlation between predicted and obtained sentence thresholds (both rounded to the nearest dB) is .96 and the RMS error of prediction is again 5.6 dB.

The three-frequency average hearing loss is of course widely used as an estimate of practical speech hearing impairment from the pure-tone audiogram. The preferred method of these laboratories has been an estimate of the average hearing level over the entire range 500 Hz to 4000 Hz. A comparison of the two methods (Farrant and Macrae 1961) has shown that the "C.A.L. average hearing loss" may be the better method of predicting social adequacy of hearing for speech and PBM threshold, but the "Three-frequency hearing loss" is the better predictor of spondee threshold.

The present study indicates that the "three-octave hearing-loss" method is also the better predictor of sentence threshold. The extent of the difference was shown by predicting sentence thresholds from the equation.

$$\text{Equn. 9} \quad TH_5 = 1/6\ TH_{500} + \tfrac{1}{3}\ TH_{1000} + \tfrac{1}{3}\ TH_{2000} + 1/6\ TH_{4000} + 29$$

The weighting used in Equation 9 is equivalent to the C.A.L. method except in the cases where a difference of 30 dB is shown between success octave points, when the C.A.L. method requires the mid-point threshold to be taken into account. On the present data, the correlation obtaine between the sentence thresholds predicted by equation 9, rounded to the nearest dB, and measured thresholds is .95 and the RMS error of predicti 5.8 dB. It is thus slightly inferior to the three-frequency method in respect.

Implicit in the analysis described so far is a particular mathematical model of the relation between pure-tone hearing and speech hearing. According to this model, the contributions to speech hearing in each frequency band are independent and unique to that band, and can be combined by means of a weighted sum, the weights being related to the information content of the band. It seems more plausible that with such highly redundant materials, with a fairly low standard of performance required, the different frequency bands should be considered instead as alternative sources of equivalent information, so that the contribution of information from any one band is not unique but depends on what is available from other bands and from non-auditory sources.

A rigorous examination of this hypothesis is not possible with the present data. However, a small change in the computational procedure allows the mathematical model to lean slightly in the desired direction. The change in procedure is to re-label each set of thresholds at 500, 1000 and 2000 Hz TH_a, TH_b and TH_c in order of elevations above normal, TH_a being the "best" of the three, and TH_c being the worst. Correlations between TH_a, TH_b, TH_c and TH_s, the threshold for sentences, were calculated, giving the following results.

	TH_b	TH_c	TH_s
TH_a	.940	.793	.913
TH_b		.842	.945
TH_c			.886

TABLE 3. Product-moment correlations between the best the median, and the worst of the 3 thresholds at 500, 1000 and 2000 Hz and the threshold for sentences.

It will be seen that the best predictor of sentence threshold is the second-best of the three thresholds at 500, 1000 and 2000 Hz. Comparison with Table 1 shows that this gives a better prediction than any threshold designated by frequency. Regression equations were derived to predict the threshold for sentences from different combinations of the variables TH_a TH_b and TH_c. These are given in Table 3, with the multiple correlation coefficients and associated error of prediction.

Equation	Variables	Prediction Equation	R	
10	TH_a only	$TH_s = .91\ TH_a + 43.3$	.91	7.9
11	TH_b only	$TH_s = .90\ TH_b + 36.2$	.94	6.3
12	TH_c only	$TH_s = .78\ TH_c + 31.1$	.89	8.9
13	TH_a and TH_b	$TH_s = .21\ TH_a + .70\ TH_b + 37.4$	.95	6.2
14	TH_b and TH_c	$TH_s = .65\ TH_b + .27\ TH_c + 32.9$	.92	7.4
15	TH_a and TH_c	$TH_s = .57\ TH_a + .38\ TH_c + 35.9$	.95	6.0
16	TH_a, TH_b and TH_c	$TH_s = .21\ TH_a + .46\ TH_b + .27\ TH_c + 34.2$	.96	5.3

TABLE 4 : Regression equations calculated from experimental data, predicting threshold for sentences from various combinations of pure-tone thresholds at 500, 1000 and 2000 Hz designated TH_a, TH_b and TH_c in order of elevation above normal threshold

A comparison with Table 2 shows that designating the three thresholds in order of elevation above threshold allows slightly better prediction by means of multiple regression equations than designating them by frequency. The difference is of theoretical interest but of little practical significance for present purposes, since when the coefficients in the last equation above are rounded for ease in calculation :

Eqn. 17 : $TH_s = .25\ TH_a + .5\ TH_b + .25\ TH_c + 33$

the correlation between predicted and obtained thresholds and the RMS error of prediction are no better than those given by equation 8 (the "three-frequency" method), i.e. a correlation of .96 and a RMS error of prediction of 5.6dB.

Discussion and Conclusion

This study has demonstrated the practicability of a clinical speech test using sentence material. The value of the test in practice has been suggested but not proved.

The relationship between pure-tone hearing and speech hearing was not the primary interest in this study. However, it is interesting to compare equation 7 with the corresponding regression equations derived in two other studies.

Equn 7 : $TH_{Sentences} = .30\ TH_{500} + .32\ TH_{1000} + .34\ TH_{2000} + 34.2$

From Farrimond (1961)

Hearing loss for sentences $= .38 TH_{500} + .43\ TH_{1000} + .07\ TH_{2000} + 4.6$

From Quiggle et al (1957)

Hearing loss for Spondees $= .22 TH_{500} + .47\ TH_{1000} + .09\ TH_{2000} + 6.9$

The differences in the final term arise from differences in methods of measuring threshold and are of no significance. The differences in the weight assigned to 2000Hz in this study compared with the other two is the surprising difference. Both of the other studies used larger number of subjects from a much more homogeneous population, but there is no obvious reason why the results should differ in this particular respect.

It is hoped that results from field trials of the test in its present form will enable a more complete evaluation of the procedure to be made. In its present form the test is suitable for use in the following applications :

(1) as a test of practical hearing loss independent of the pure-tone audiogram

(11) as a check on the validity of the pure tone audiogram.

The recommended procedure for clinical use of the test and the test material itself are given in appendixes to the report.

REFERENCES.

Dudley, P. F. and Macrae, J. H. : The Measurement of improvement in Speech Hearing and its relation to hearing aid use. C.A.L. Internal Report No. 4 (1968)

Farrant, R. H. and Macrae, J. H. : Comparison of C.A.L. Average Hearing Loss (Three Octave Method) and the Three Frequencies Averaging Hearing Loss as Predictors of Deficiencies in Hearing for Speech Etc. C.A.L. Informal Report No. 24 (1961)

Farrimond, T. : Prediction of Speech Hearing Loss for Older Industrial Workers. Gerontologia 5, 65-87 (1961).

Fry, D. B. Word and Sentence Tests for Use in Speech Audiometry Lancet 2, 197-199 (1961)

Quiggle, R. R., Glorig, A., Delk, S. H. and Summerfield, A. R. Predicting hearing loss for Speech from pure tone audiograms Laryngoscope. 67, 1-15 (1957).

APPENDIX 1.

RECOMMENDED PROCEDURE FOR CLINICAL USE OF TEST MATERIAL.

1. Calibration.

It is suggested that presentation levels and obtained thresholds for speech tests should be recorded in dB SPL as measured on the calibration tone. The calibration procedure determines what correction figure should be added to the audiometer attenuator marking to give dB SPL in the earphone. The following procedure is applicable to the C.A.L. D3 audiometer, calibrated to 1951 A.S.A. Standards.

It is assumed that the audiometer has been calibrated acoustically for pure tones. Check the output VU meter when the 1000Hz tone from the audiometer oscillator is keyed, with the attenuator set at 80. This meter reading corresponds to 96.5 dB SPL in the earphone. Then with the 1000 Hz calibrating tone from the disc or tape played through the same earphone and the attenuator set at 70, adjust the calibration knob to give a meter reading 3.5 dB higher than that obtained previously. This corresponds to 100 dB SPL in the earphone. At other settings of the attenuator add 30 to the attenuator marking to obtain speech levels in dB SPL in the earphones.

2. Instructions to Subject.

Variations in wording of instructions are probably not important as long as they do not affect the subject's expectation of his performance on the test. The following instructions were used in this study:

> "This is a test of your hearing for speech. I will play you a recording of short sentences, like "The boy broke the chair" I want you to repeat each sentence as you hear it. Some of the sentences will be too quiet for you to hear properly, but say as much as you can of each sentence".

The instructions may be repeated or explained as necessary; note that the word "sentences" is very difficult to lip-read.

3. Level of Presentation.

The first sentence should be presented at a level some 10-15dB above the subject's presumed threshold. Be prepared to stop the tape to repeat instructions if necessary or to encourage a part-response. After every response scored right, decrease the level 5dB by means of the audiometer attenuator; after every response scored wrong increase the level 5 dB.

4. <u>Scoring and recording.</u>

At this stage of development of the test, complete recording of responses and levels of presentation is recommended. Record error and omissions by altering or crossing out the sentences printed on the score sheet, marking with a tick all sentences scored correct. Enter the level of presentation of each item in the left-hand margin; calculation of threshold is made easier if the higher levels are recorded further to the left than lower levels, as in the example record appended.

The principle involved in scoring has already been explained in the body of the report. More precise definition of right and wrong responses will be given in the light of further field experience with the test.

5. <u>Calculation of threshold.</u>

Calculate the mean of the presentation levels which lead to a reversal of scoring (from right to wrong or vice versa). These are the peaks in the zig-zag line drawn in the example. There are a number of short cuts which may be used in making this calculation; one is suggested on the example.

6. <u>Interpretation of Results.</u>

The following example illustrated the suggested form of reporting the result of this test.

Threshold for speech (sentences) = 58 dB SPL = 25 dB loss (consistent with PTA)

For the purpose of converting threshold in dB SPL into dB hearing loss for speech, regard 33 dB SPL as the "normal" threshold as obtained by this procedure.

For the purpose of determining whether the obtained threshold is "consistent with the pure-tone audiogram", calculate the "three-octave average hearing loss" (average of thresholds at 500, 1000 and 2000 Hz) and add 38. This figure is within 10 dB of the obtained threshold in 90% of tests.

The following arbitrary rule is therefore recommended:

That when the predicted and obtained thresholds differ by 10dB or more, the speech test result be regarded as "better/worse than indicated by the pure tone audiogram." Results of other studies indicate that this sort of discrepancy has diagnostic significance, but as this stage, such a discrepancy should be taken merely as an indication that further diagnostic tests are required.

EXAMPLE OF SCORING

SENTENCE LIST B

70	✓ 1. The police feared a riot
55	✓ 2. The puppy chewed the ball of string
60	3. The teacher marks (ed) the exercises
55	4. The child ~~was shown a picture~~ (showed ——)
60	5. ~~The park-keeper~~ locked ~~the main gate~~
65	✓ 6. The gamblers tried their luck
60	✓ 7. The black cat sharpened ~~his~~ (its) claws
55	✓ 8. The man drank his beer
50	9. ~~The car skidded on the wet road~~
55	10. The mother ~~needed a lengthy rest~~ (neared ——)
60	✓ 11. The procession passed the house
55	✓ 12. They stayed here twelve months
50	13. ~~The fire engine~~ went rushing by
55	✓ 14. The clergyman preached (s) a long sermon
50	✓ 15. The taxi ran into a bus
45	16. ~~The doctor arrived in time~~
50	17. The whole family ~~enjoyed the play~~
55	✓ 18. The boy took off his coat
50	19. The agent ~~let~~ (left) the old house
55	20. The ~~thief~~ took the wrist-watch
50	✓ 21. All rabbits eat lettuce
55	22. The ~~gardener~~ (governor) cuts the lawn
60	✓ 23. Soldiers wear khaki uniform
55	24. The magistrate ~~bound~~ (found) her ~~over~~ (owner)
60	✓ 25. The woman made the cakes.

45	1
50	3
55	4
60	4