Proceedings of The Institute of Acoustics

A PSEUDO-CONVERSATION TEST FOR ASSESSING SPEECH LINKS

PROF. D. L. RICHARDS: R. W. WHORWOOD

THE UNIVERSITY OF ASTON IN BIRMINGHAM, U.K.

Introduction

Until the modern sophisticated methods of speech transmission became common in telephone networks it was permissible to use fairly crude methods for assessing whether telephone connections were likely to be acceptable by the telephone-using public.

Much assessment work on speech communication links has been carried out by means of listening-only type tests. That is speech material in the form, for example, of lists of nonsense words was spoken over the connection to listeners who wrote down what they heard (Ref. 1). The proportion correct gave a measure of the 'goodness' of the connection. This method is now little used for public telephone systems although another simpler form of listening-only test is still used, namely loudness balancing to ensure that the overall customer-to-customer transmission losses are suitable.

Of course a listening-only test using sentence material can be used for limited purposes; such a test usually requires the listener to express a quality judgement rather than reproduce what he heard. Where a full conversation test cannot be employed (because it is too costly) listening opinion tests may give useful information.

Conversation testing

It is clearly possible to imagine a connection over which conversation is scarcely possible and yet which would show up favourably on a listening-only-type test. Factors which harm conversations include echo, delay, cross-talk, side-tone, etc., but these are totally ignored by listener-only tests. The large telecommunications organisations now make extensive use of conversational methods of assessment. A conversational test enables all the many factors affecting the quality of a speech link to be assessed as a whole. By careful design of the experiments it is feasible to examine the relative importance of various factors not merely in isolation, but in the presence of other factors excluded in listening-only tests. Important interactions between factors can thus be investigated. The subjects, while listening, participate actively in the conversation they are assessing and by talking alternately with listening directly affect it (Ref 2).

But conversational assessment methods are extremely expensive to carry out and are very laborious. Also the results of such tests are not available quickly and so such tests are not suitable for network optimisation. Even the large organisations use conversational tests only for key experiments and interpolate from the results to other conditions where possible.

Against this background a method of carrying out a pseudo-conversation test, which is almost as easy to administer as a listener-only test, is attractive.

A PSEUDO-CONVERSATION TEST FOR ASSESSING SPEECH LINKS

The method to be described is still in the process of development but early results are most encouraging and are in line with known results from previous full-scale tests of similar circuits.

The objective of the pseudo-conversation method is to include as many features as possible of those to be found in a full conversation test. (The test cannot, of course, be used to examine the effects of delay). The principle is for the subject to 'converse' with a tape-recorder. The subject is located in an office-like environment and all the equipment involved is placed in a neighbouring room. The subject is therefore expected to act normally and not be inhibited in any way by the laboratory conditions.

A set of sentences is constructed. These sentences are all of a pre-determined form in the active voice, are syntactically correct, and consist of words extracted from a set of one hundred actual telephone conversations (Ref. 3). The words are classified into type, i.e. nouns, verbs, adjectives, etc., and are stored on a computer in these classes. The sentences at present being used are of the form determiner-noun-auxiliary verb-verb-determiner-adjective-noun. Other forms will be considered in the future. A computer program has been devised which produces sentences in the chosen form by selecting, at random, from the words stored in the computer. It follows that the sentences are nonsensical.

The sentences are recorded over the local telephone system as shown in figure 1B and the long-term r.m.s. level of the recorded sentences measured. The sentences are then re-recorded to ensure they are all at the same level. The re-recording stage can be avoided if the talker producing the sentences can be constrained to talk at the same level throughout the recording session.

Subjects are invited to listen to the sentences over the local telephone system shown in figure 1C and are asked to make a response requiring a certain amount of effort. How much effort it is wise to ask for is still a matter of study. At the present time subjects are asked to put the active voice sentence they hear into the passive voice. They are allowed seven seconds to respond before the next sentence is presented.

This response on the part of the subject forms his reply. A fairly quick response is expected from the subject in order to simulate the conversation situation.

At the end of each block of five sentences the presentation of sentences is stopped and the subject invited to indicate, over the telephone set being used, his opinion of the circuit. This opinion is given on a scale of listening effort A, B, C, D and E.

The subjects are asked to interpret these letters as follows: A = complete relaxation possible, no effort required; B = attention necessary, no appreciable effort required; C = moderate effort required; D = considerable effort required; E = no meaning understood with any feasible effort. The categories A to E are scored 4, 3, 2, 1 and 0 respectively. Opinion scales of this type are discussed in reference 4. All the replies made by the subject are recorded for subsequent analysis.

A PSEUDO-CONVERSATION TEST FOR ASSESSING SPEECH LINKS

Having recorded the subject's replies the circuit conditions are changed in a way determined by the factor or factors being studied and the task repeated for a further block of five sentences. A subject will in general be exposed to about six or eight sets of sentences and perhaps a dozen subjects used on an occasion.

Assessment

During the tests tape recordings are made of signals at various points on the circuit. As a minimum it is necessary to record: (a) the sentence actually heard by the subject; (b) the reply made by the subject to each sentence; (c) the subject's opinion score.

The experimenter can choose the points where recordings are made in accordance with the purpose of a given test. The recordings can be used to measure the subject's talking level and determine how this changes as the test conditions are changed. Factors such as side-tone and noise effect talking level and may be studied. Examination of the replies made by subjects may be made on the basis of scoring the proportion of words used which are correct. In this regard it has been found essential to distinguish between key words like verbs and nouns and minor words like determiners and auxiliary verbs. Subjects appear to ignore the minor words and concentrate on major key words; this is consistent with every-day conversational behaviour. In the results reported here the score was based upon the proportion of key words which were correct. A rather harsher method would be to score sentences as correct only if all the key words in it were correct.

Figures 2 and 3 show some results from an early experiment. The smooth curves on figure 2 are obtained from previous work carried out by the U.K.P.O. Research Centre on identical circuits using the CATPASS computer simulation (ref. 5) of a full conversation test. In figure 2 the subject's opinion score is plotted against psophometrically weighted noise which was injected at the point shown in figure 1. Only three points were obtained at each of three speech levels.

It will be noticed that the points at the lowest noise level on the abscissa lie well below the respective curves. This arose because an excessive amount of hum, not included in the noise measurement, was present. However the other points are generally in line with the CATPASS estimate. Figure 3 shows how subjects' opinion score (Y_{LE}) accords with their actual performance. The lowest and highest received speech levels indicated on the diagram by crosses and circles respectively form straight lines. The middle value for speech level gives less consistent results.

Conclusion

The technique described is intended to find a place between the listening-only tests, which are relatively easy to conduct but are suitable for assessing only a limited range of factors, and the very general conversation tests, which require very elaborate facilities. It is expected that the method will enable many aspects of talker-behaviour to be taken into account as well as the purely listening factors like loss, noise and distortion.

A PSEUDO-CONVERSATION TEST FOR ASSESSING SPEECH LINKS

References

(1) F. MARKMAN 1953 Ericsson Tech, 9,8-50. 'Assessment of transmission properties of telephone instruments based on articulation tests'.
(2) D. L. RICHARDS 1964 Proc. IEE 111, 931-940. 'Transmission performance assessment for telephone network planning'.
(3) J. BERRY 1953 Butterworths Scientific Publications, 392-401. 'Some statistical aspects of conversational speech'.
(4) D. L. RICHARDS 1973 'Telecommunication by Speech', Chapter 3. 4, Butterworths.
(5) D. L. RICHARDS and P. K. WEBB 1976 National Telecommunication Conference, Dallas, Texas. 'Catpass - a model for estimating customer satisfaction'.

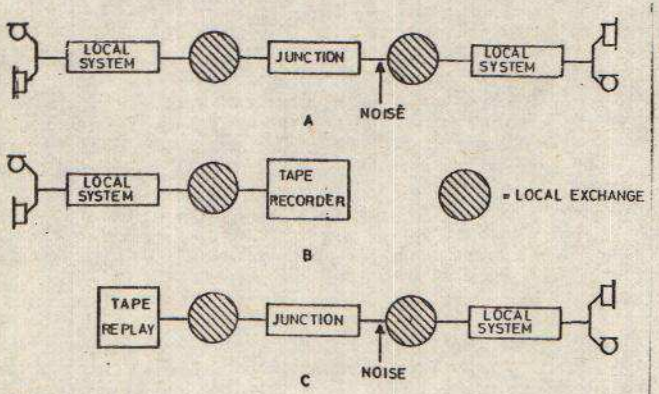

FIG. 1

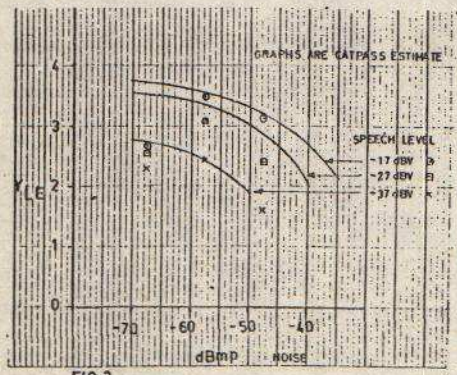

FIG. 2

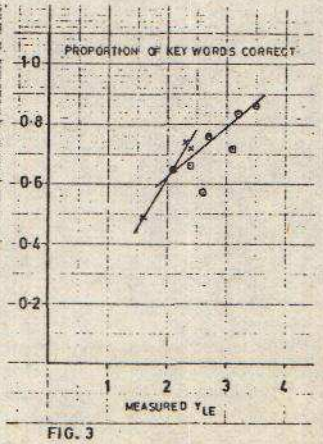

FIG. 3