BRITISH ACOUSTICAL SOCIETY

"SPRING MEETING" at Chelsea College, London, S.W.3 on
Wednesday 25th April / Friday 27th April, 1973.

SPEECH AND HEARING: Session 'A': Speech Production and Perception.

Paper No:

73SHA4

INTERFERENCE TEST FOR SPEECH FIDELITY EVALUATION

Lloyd H. Nakatani
Bell Laboratories
Murray Hill, New Jersey 07974 U.S.A.

ABSTRACT

The interference test, which measures the intelligibility of speech under adverse listening conditions, is proposed as a sensitive and objective procedure for speech fidelity evaluation. The test successfully discriminated fidelity differences among eight speech types ranging from telephone speech to high fidelity speech, even though all the speech types received very high and indistinguishable scores on conventional articulation tests. The results suggest the following three stages in speech fidelity evaluation: (1) articulation test, (2) interference test, and (3) subjective tests.

INTRODUCTION

Presently speech fidelity evaluation employs two types of tests: articulation test and subjective tests of speech quality. An articulation test is a performance test since it measures a listener's ability to understand the speech under evaluation. Subjective tests (e.g., rating and paired-comparison) are opinion tests since they measure opinions or judgments about subjective factors in speech which presumably cannot be objectively measured.

Generally speaking, an opinion test must be used to discriminate fidelity differences among speech types when the articulation test fails to discriminate because the scores are too high. Unfortunately, opinion tests suffer from some severe deficiencies. First, opinion tests often cannot be used in many situations where the speech types are not easily compared. For example, a person's hearing cannot be evaluated with an opinion test. Second, opinions are relative, so a given speech type may receive either a high or low opinion depending on what other speech types it is compared with. Third, how much a factor affects speech fidelity, in an absolute sense, is difficult to measure with an opinion test. For example, a factor B midway between a minor factor A and major factor C may have a large or small effect on opinions depending on whether factor A or C, respectively, also varies in the speech types under evaluation. Performance tests do not have these problems.

A new performance test is therefore proposed which can be used to discriminate fidelity differences when an articulation test no longer suffices. The availability of such a test introduces a third stage in speech fidelity evaluation which is intermediate to articulation and opinion tests. The proposed interference test measures speech intelligibility under adverse listening conditions, in contrast to the articulation test which measures speech intelligibility under ideal listening conditions.

SPEECH INTERFERENCE TEST

The speech interference test is essentially an articulation test run with interfering speech. A subject listens to two channels of speech at once. The signal channel contains test material which he must understand and write down, and the interference channel contains irrelevant material which he must ignore. The same talker and speech type is heard in both channels. The test determines, for each speech type, the interference threshold corresponding to the signal-to-interference intensity ratio (S/I) where the signal speech is 50% intelligible. The fidelity of a degraded speech type is measured by the difference between its threshold and the threshold for a high fidelity reference speech type. This measure, Q, increases as speech fidelity decreases.

ARTICULATION VS. INTERFERENCE TEST

An experiment was run to demonstrate the utility of the interference test in speech fidelity evaluation. The PB-word articulation test and interference test were used to evaluate the fidelity of speech types which differed obviously in fidelity but seemed highly intelligible. Three components of the telephone system (microphone, transmission channel and earphone), each with two alternatives (high fidelity or telephone), were combined as shown in Fig. 1 to generate eight speech types ranging from telephone speech (CLR) to high fidelity speech (EWH).

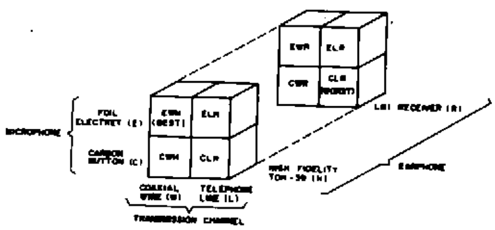

Fig. 1 Speech types generated by a 2 × 2 × 2 factorial design.

PB-word articulation test. The mean articulation scores averaged over ten subjects are shown in Fig. 2 for the speech types along with their 95% confidence intervals. (The predicted Q will be explained shortly.) The scores ranging from 92% to 98% confirmed that the speech types were all highly intelligible under ideal listening conditions, but, as expected, the scores failed to discriminate fidelity differences among the speech types. Note, for example, that telephone speech (CLR) received about the same score as high fidelity speech (EWH), even though these speech types were at the extremes of the fidelity range considered here.

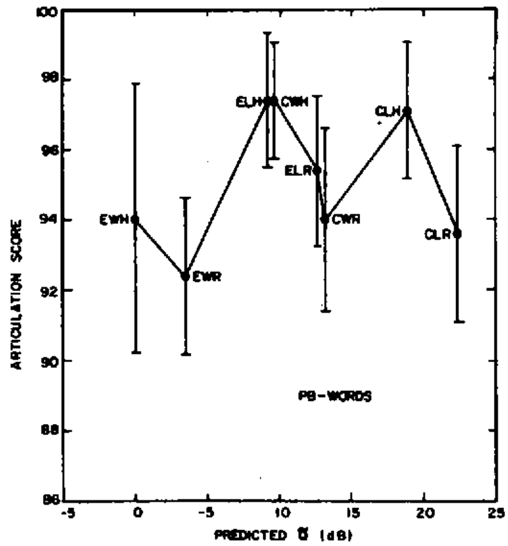

Fig. 2 Mean PB-word articulation test scores and 95% confidence intervals for the eight speech types.

Interference test. Figure 3 shows the mean Q values and 95% confidence intervals obtained for the speech types. Each data point is based on six estimates from each of five subjects. The wide spread among the Q values indicates that the speech types differ significantly in fidelity. The fidelity differences were subtle and small enough so that communication performance was impaired only under the adverse listening conditions of the interference test, but not under the ideal listening conditions of the PB-word articulation test.

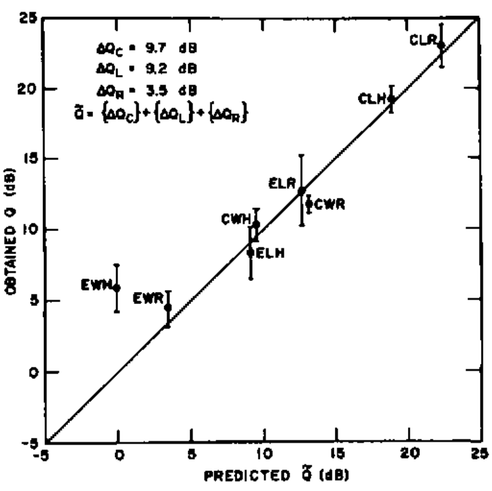

Fig. 3 Mean interference test Q values and 95% confidence intervals as a function of predicted $\tilde{Q}$ values for the eight speech types.

Additivity of component degradations. The amount by which each telephone component degraded speech is apparent in the difference between the Q values for pairs of speech types where only one component was changed. For example, the differences between the Q values for the pairs (CLR, ELR), (CLR, CWR), and (CLR, CLH) are estimates of the improvement one may expect in the telephone speech (CLR) when a high fidelity component is substituted for a telephone component. Based on such pair-wise differences, it was estimated that the carbon button microphone, telephone line, and receiver degraded speech by

9.7, 9.2, and 3.5 dB, respectively; these ΔQ values are shown in Fig. 3.

The method of generating the speech types made it possible to check whether any component reduced or exaggerated the degradation of another component. There was very little evidence of such interactions, which means that the component degradations were additive. This implies that the Q value for a speech type could be predicted by simply adding up the ΔQ's corresponding to the telephone components involved in the speech type. The prediction formula is shown in Fig. 3. The linear relationship between the obtained Q and predicted Q values in Fig. 3 indicates that additivity was a reasonable hypothesis for the components.

Validity check. In order to check the validity of the interference test, opinion tests were run since such tests are generally regarded as providing valid measures of speech fidelity. The five subjects of the interference test were asked to rate the quality of the speech types on a nine point scale, and to state their preference between the speech types in a paired-comparison test. Within each earphone alternative, the ordering of the speech types from best to worst was identical for the interference, rating, and paired-comparison tests, thus supporting the validity of the interference test. The opinion test results for different earphones could not be combined for the validity check, because the act of physically interchanging headsets made earphone comparisons very difficult.

CONCLUSIONS

A three stage program for speech fidelity evaluation is proposed where the successive stages are (1) articulation test, (2) interference test, and (3) opinion test. Stages (1) and (2) measure the intelligibility of speech under ideal and adverse listening conditions, respectively, and stage (3) taps the subjective factors which may differentiate among speech types.

Experimental evidence was described to show that the interference test can differentiate fidelity differences among speech types which cannot be differentiated by articulation tests. Comparisons between the result of the interference test and opinion tests (rating and paired-comparison) indicate that the Q values are valid measures of speech fidelity. It was also argued that performance tests have many advantages over opinion tests. Collectively, these results and considerations argue for the inclusion of the interference test as an intermediary stage between the two traditional stages of intelligibility and quality testing for speech fidelity evaluation.