

# STUDIES OF METHODS FOR THE MEASUREMENT OF SPEECH COMPREHENSION

Robert McAllister and Mats Dufberg

Dept. of Linguistics, Stockholm University, Sweden

## ABSTRACT

This paper is a report on ongoing research on the explicit measurement of speech comprehension. Our point of departure was a need for reliable speech comprehension tests with a higher degree of validity than existing measurement instruments. After several experimental studies which will be summarized here, we have developed a global comprehension test called the Question and Response (QAR) test. Using hard of hearing, second language learners as listeners/subjects we are at present attempting to assess the utility of this test as an instrument for the measurement of comprehension ability.

## 1. INTRODUCTION AND BACKGROUND

This paper is about the measurement of the ability to understand spoken language. This faculty has, in earlier research, been referred to by means of several different terms. We have chosen the term "speech comprehension" and wish to take a global approach to the definition of this ability. Figure 1 [5] is a graphic representation of some important aspects of one current view of speech comprehension upon which the work reported here is based. It is a very general model of the relationship between two major sources of information used by the listener to interpret a spoken utterance. One of these is the information contained in the speech signal rep-

resented in figure 1 as *signal dependent information* and often referred to in the past as "acoustic cues". The other source of information represented in the figure as *signal independent information* refers to the knowledge of the language spoken, knowledge of the world, the current communication situation, etc.. This information creates expectations on the part of the listener as to what meanings are to be communicated and thus greatly facilitates understanding of the message. It should be noted that while we are aware of the important role played by visual information in the interpretation of spoken utterances [7], only the acoustic/auditory component has been considered in this work.

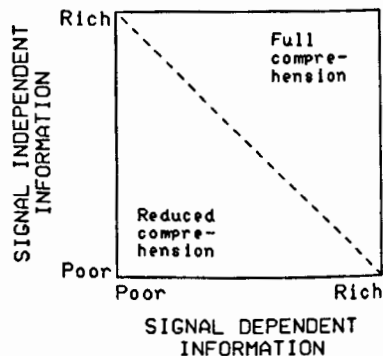


Figure 1. Model of the relationship between two major sources of information used by the listener to interpret a spoken utterance, upon which the work reported here is based.

Our point of departure was the discovery of the need for methods that could be used for the estimation of an individual's ability to understand language spoken in everyday situations. These methods would have several applications both clinical and pedagogical and could also be a useful research tool. There are a number of these tests used in various clinical settings including audiology and logopedics and phoniatrics as well as in the field of foreign and second language teaching where there is a need for methods for the measurement of learners comprehension ability and assessment of the quality of their pronunciation in terms of its comprehensibility. The problem with these existing methods, which is central to the work summarized here, concerns test validity.

The purpose of this work is to develop methods for the explicit measurement of speech comprehension with special consideration of content, construct, and concurrent validity [1]. At present, we are working on the development of a global comprehension test which in its present form is called the Question and Response (QAR) test. The rationale and methods used in the development of this test are summarized below. For a more detailed account of the experiments leading to the present design of the QAR test see McAllister and Dufberg, [7] and Dufberg [2].

## 2. METHODS

An inventory of comprehension testing methods currently in use motivated an emphasis at the outset on construct and content validity in our early research in the development of the QAR test. The model presented in fig. 1 is the theoretical basis for the structure of the QAR test. This model, however, is very general and vague in terms of specific perceptual mechanisms. It was therefore necessary to consider recent research in

comprehension testing in light of linguistic-phonetic theory. It was decided to begin with a method designed by Walker and Byrne [9] which could be said to be a version of a general test paradigm to assess speech reception threshold (SRT). Running speech was presented in noise and the test result was the signal to noise ratio at which the subject/listener "just barely understood the meaning of the text".

With the above experimental configuration (fig 2) we tested several important features of the original SRT paradigm.

### 2.1 Noise type

Several types of noise sources were used in these experiments. Our pilot studies narrowed these various maskers down to two main candidates for the QAR test. One was a colored, low frequency modulated noise whose long time average spectrum was approximately the same as a male voice. The other was a "cocktail party noise" achieved by overdubbing one male and one female voice many times to create the effect of a roomful of people engaged in lively conversation. The results of these experiments showed that the "babel", as we called the cocktail party noise, was

### GENERAL EXPERIMENTAL CONFIGURATION

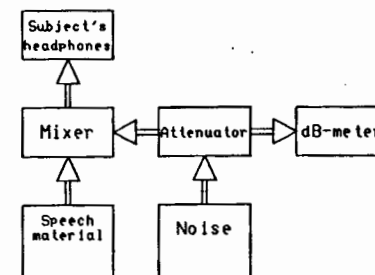


Figure 2.

the most effective masker and therefore we chose this noise for the QAR test.

## 2.2 Speech material

From a point of view of validity, we judged the original method and therefore the speech material in the SRT test to be inappropriate. Instead of a subjective judgement as to the noise level at which the connected speech was "just barely understood" we chose to ask simple questions whose answers would be clear to anyone who knew the language and had understood the question. It was assumed that this would be a more valid method for determining whether or not the listener/subject had understood the running speech. A further development of this idea resulted in the material for the QAR test. This material consisted of a series of short texts averaging 31 words (5 sentences) per text and constructed systematically according to text linguistic principles of script theory [8].

## 2.4 Subjects

Three subject groups were used in the experiments which led to the QAR test. These were: Foreigners who used Swedish, the language of the tests, as a second language, hearing impaired persons, and normal native speakers of Swedish as controls.

## 2.5 The QAR test and test comparisons

The methods summarized above were used in the development of the QAR test. This test, then, used "cocktail party" noise mixed with short texts. Subjects were asked to answer questions, also presented in noise, about the text and the S/N ratio was adaptively adjusted according to the subjects performance in the answering task. The test result is a S/N ration which represents a threshold of 50 per cent speech comprehension.

### 2.5.1 Tests for comparison

At the time of the writing this paper, the QAR test is being given to new subject groups together with the following tests for comparison purposes.

*Bekësy pure tone threshold audiometry.* This is a standard hearing test used in clinics in Sweden and elsewhere.

*Hearing Threshold for Speech.* Also a standard audiometric test in swedish clinics. Part of the test repertoire called "Speech Audiometry", this test is supposed to indicate speech comprehension ability.

The Hagerman Test. Also part of the "Speech Audiometry", test repertoire supposed to indicate speech comprehension ability [4].

*Sentence Completion Test and Word Completion Test.* Two paper and pencil tests in which subjects, within limited time frames, had to use syntactic and lexical information to complete various tasks.

*Modified Hearing Measurement Scale.* A self assessment test based on the Swedish version of the Hearing Measurement Scale [3] in which subjects estimate their comprehension of speech in a number of hypothetical everyday situations.

## 3. RESULTS

The results presented here are based on only 5 normal and 3 immigrant subjects. At the time of the ICPhS we will have more data to present from these subject groups as well as from a hearing impaired subject group.

The results from the various tests that were run in these experiments were of different types and could not be directly compared by means of parametric statis-

tics. The subjects were rank ordered for each of the seven tests mentioned above i.e. resulting in seven rank orders. A Spearman rank order correlation was performed between the QAR test and all the other tests given to the subjects. The only test with which the QAR seemed to be correlated was the self assessment test here called the modified Hearing Measurement Scale. This correlation could only be said to be modest with  $r_s=.52$ . All the other test showed a very low positive or negative correlation.

## 4. DISCUSSION

It was not entirely unexpected that the QAR test would not show a high correlation to the other tests used in these experiments. It has been our contention, on the basis of validity arguments, that it could be quite possible that these tests do not, in fact, test speech comprehension as we have defined it here. We would suspect that important components in speech comprehension are tested by them, but that the global aspects of this ability may not be captured. The moderate correlation of the QAR test with the self assessment test could be seen as encouraging if it holds. We would, at present, be very cautious in lending a great deal of credibility to these preliminary results. At the time of the ICPhS we will be able to present similar data based on larger groups of subjects and hopefully be able to make more definitive statements about the utility of the QAR test.

## 5. REFERENCES

[1] CRONBACH, L.J., (1961) *Essentials of Psychological Testing*, New York, Harper & Row.

[2] DUFBERG, M. (1990), "The measurement of speech comprehension: a progress report of the construction of a global test", *PHONUM* 1, Dept of Phonetics, Univ of Umeå.

[3] ERIKSSON-MANGOLD, M. and HALLBERG, L. (1990), "Hörselskada och upplevt handikapp. Hearing Measurement Scale på svenska" (the Hearing Measurement Scale in Swedish), Report from the Dept of Psychology, Gothenberg University, nr. 3.

[4] HAGERMAN, B. (1984), *Some aspects of methodology in speech audiometry*, (dissertation) Karolinska inst., Stockholm.

[5] LINDBLOM, B. (1987), "Adaptive Variability and Absolute Constancy in Speech Signals: Two Themes in the Quest for Phonetic Invariance", *Proceedings of the XI International Congress of Phonetic Sciences*, Tallinn, August 1987.

[6] McALLISTER, R. and DUFBERG, M. (1989), "Some Attempts to Measure Speech Comprehension", *PERILUS IX*, Institute of Linguistics, University of Stockholm.

[7] MIDDELWEERD, M. J. and PLOMP, R. (1987), "The effect of speech reading on the speech reception threshold of sentences in noise" *JASA* 82, nr. 6, 2145-2147.

[8] SAMUELSSON, S. and RÖNNBERG, J. (1990), "Script activation in lipreading" *Scandinavian Journal of Psychology* 13.

[9] WALKER, G. and BYRNE, D. (1985), "Reliability of Speech Intelligibility Estimation for Measuring Speech Reception Threshold in Quiet and Noise", *Australian J Audiology* 7:1.