Development of a new speech comprehension test with a phonological distance metric

Robert McAllister
Department of Linguistics
Stockholm University
E-mail: bob@ling.su.se

ABSTRACT

A new speech comprehension test has been developed. The test has been constructed with special consideration of test features, which enhance construct validity. These features include a careful selection of test stimuli from large databases enabling phonetic balancing that reflects word and phoneme frequency in the test language, Swedish. Words are presented in noise and the adaptive procedure results in a comprehension threshold expressed as a signal to noise ratio. An innovative feature is the use of a phonological distance metric, which gives the subject credit for good guesses. A group of native Swedish speakers with normal hearing was tested to establish a normal base line test result. A group of L2 users was tested and their results indicate that Perceptual Foreign Accent can be measured with the test. It was demonstrated with synthetic speech that speech quality could also be measured with this method.

1. INTRODUCTION

The first and foremost problem to be addressed in the setting up of guidelines for the construction of a comprehension test is that of the definition of ”speech comprehension”. The decoding of the speech signal and the subsequent understanding of the spoken message involves a complex of physiological and psychological mechanisms that makes this definition and the description of this phenomenon a formidable task indeed. The comprehension test, which is the object of this report, has been developed with a linguistically based view of this process. In this paper we will attempt to briefly describe the principle theoretical and practical bases for the features of the test. These theoretical and practical considerations provide then, in turn, the basis for a judgment of test validity, particularly that which Cronbach (1960)[3] in his classical analysis of validity refers to as "construct validity". The work on the development of this comprehension test has, as its point of departure, the identification of the need for a tool, which could provide an explicit measurement of speech comprehension in several contexts. An inventory of the need for explicit measurement of speech comprehension was carried out by McAllister and Dufberg (1989)[7] It was found that such methods would be useful in various research contexts as well as in some major clinical settings such as Audiology and Logopedics and Phoniatrics.

1.2 LEXICAL ACCESS IN SPEECH COMPREHENSION

A psycholinguistic view of speech perception or speech comprehension has yielded the assumption that the ultimate linguistic goal of the physiological and cognitive activity involved in speech comprehension is that which linguists refer to as ”lexical access”. The nature of the information contained in the lexicon for each lexical unit has been a subject of considerable linguistic debate and inquiry. Lexical access in the context of speech comprehension refers to the first step in the comprehension process which involves decoding the auditory information in order to recognize the words and sentences intended by the speaker and thereby the basic semantic content of the message. We would hasten to add that we are well aware that our view of the complex processes behind the everyday experience of understanding each other is not fully characterized by that which leads to what we here have termed lexical access. It is our hypothesis, however, that a valid test of lexical access ability may be a fair predictor of spoken language comprehension ability.

In order to access this vast storehouse of linguistic units, the listener utilizes various kinds of information. The construction of our test is based on the claim that this information is of two general types: signal dependent

 ISBN 1-876346-48-5

information and signal independent information. We would further claim that the inclusion of a measure of the ability to successfully handle these two information types might enhance the validity of a speech comprehension test

2. TEST CONSTRUCTION FEATURES

As our assumption is that lexical access is the goal of the process which we are attempting to test, it follows that we have constructed a word test composed of disyllabic items.

2.1 PHONEME AND WORD FREQUENCY

An important aspect of the construction of our test has been consideration of phoneme and word frequency in Swedish. We have reasoned that the phoneme distribution in the word material should reflect the distribution of the phonemes in the language being tested. This was the original idea in the principle of "phonetic balancing" first conceived by James P.Egan (1948)[4] at the Psychoacoustic Laboratory at Harvard University in the 1940's and followed by Lidén and Fant in the phonetically balanced tests for Swedish from 1954 (Liden&Fant, 1954)[6], which are still in use today. Our phoneme and word frequency data are drawn from the TMH text database, which is considerably larger than those used for the so-called FB-lists used in these Swedish speech audiometry tests. The TMH database is composed of 150 million words from various sources. Using the 500 most common Swedish words from this database, we have subsequently consulted the Spoken Language data base (Talspråksdatabas) compiled by Allwood and colleagues at Gothenburg University (Allwood, 1999)[1] and the phoneme frequency part of the TMH database. This selection process was an attempt to obtain test stimuli which reflected the word and phoneme frequency in Swedish and which would be uniformly familiar to native Swedish listeners.

2.2 UNIFORMITY OF TEST STIMULI

Other aspects of the test stimuli which represent a deviation from the conventional form of speech comprehension tests is that the test words are not of uniform prosodic or morphological form. The TMH database has made it possible to select common disyllabic words with both lexical accent 1 and 2 in a proportion that reflects the general proportion of lexical tone accent occurrence in Swedish. Also, the test words are presented in their most common morphologically inflected forms. The reason for this is the aforementioned understanding of the lexical access process as a result of the utilization of both signal dependent and signal independent information. It follows that the test score is more likely to reflect speech comprehension ability if both signal dependent and signal independent information need to be used in the recognition of the test items. We hypothesize that this is more probable if the prosodic and morphological form of the test items is not predictable.

3. TEST ADMINISTRATION

For each test, the test program selects 40 words from the 100 potential test stimuli. Each selected list contains a predetermined proportion of disyllabic items with a prosodic structure, which reflects the distribution of stress and lexical tone accent in Swedish. A few dummy stimuli are then presented to the subject and he or she is asked to adjust the presentation volume to a comfortable level. Then a calibration tone is played and the output voltage at the previously calibrated earphones is registered. The test stimuli are then presented in a carrier phrase word by word via the earphones and the testee is asked to repeat the word heard. After the presentation of 10 stimuli without masking noise, the noise is initiated so that the signal to noise ratio is +30 dB. For every correct answer, which the test administrator registers, the noise level increases. In the beginning of the test, the noise level increases rapidly with each correct response until the subject is not able to accurately repeat the target word. For every unsuccessful attempt at repeating the target word the noise level is decreased, for every successful attempt, the level is increased. This adaptive procedure is followed until all the words have been presented and the final signal to noise ratio, interpreted as the lexical access threshold, is the test result.

3.1 THE PHONOLOGICAL DISTANCE METRIC

The measurement of the phonological distance between the target word and the word spoken by the testee is the product of an algorithm developed by B. Brodda (Brodda, 1966)[2]. The experiments, which were the basis for this distance measurement, were done between 1966 and 1970 as a consultation for the Swedish Patent Office, which was

interested in the possibility of the use of a computerized comparison of the phonological similarity between trademark names. New registration of a patented trademark was partly dependent on its not being phonologically too similar to any existing trademark. This method of an explicit measurement of phonological similarity proved to be useful in this application and is still used today as a commercial product with which a great number of word comparisons are carried out each year. The measurement is used in our test when the testee does not succeed in repeating the target word test stimulus. In this event, the test administrator transcribes the word actually said by the testee and this word is compared to the target word and the phonological distance is measured. The result of our version of this measurement is a number between 0 and 1 which has a role in the control of signal to noise ratio for the immediately subsequent test word. The phonological distance metric is a means by which the ability to use the signal independent information can be included in the test result. In other words, the test subject gets credit for a good guess that is phonologically similar to the target word.

4. NORMALIZATION

A group of native speakers of Swedish were selected on the basis of their hearing for the test normalization. Our criteria for normal hearing were rather stringent. To select subjects qualified as having normal hearing and thus for participation in the test normalization a standard pure tone screening procedure was used. An individual was required to have no more than a 20dB reduction at no more that one frequency in both ears. 45 subjects were selected according to this procedure and given the newly developed speech comprehension test. The age distribution of the normal hearing subject group is shown in table 1. The mean test result for this group was S/N=-8.4dB with a SD of 2.2.

Two additional subject groups have been given the test in an attempt to develop arguments for test validity. One of these groups was second language (L2) users of Swedish. Previous research has shown that normal hearing L2 users with high general L2 proficiency and several years experience with the L2 display a L2 comprehension deficit when compared to native users (Florentine, 1985[5], McAllister & Dufberg, 1989[7]). In some further research, McAllister (1997)[8] has verified this tendency and called it “Perceptual Foreign Accent”.

Age group	Normal n=45	L2 n=17
20-25	22	1
26-30	13	9
31-40	8	5
41-50	2	2

Table 1: Age distribution in the normal hearing native and L2 subject groups.

An appropriate indication of the validity of the comprehension test under consideration here might be to see whether it is sensitive enough to measure Perceptual Foreign Accent. 17 L2 users of Swedish were selected to participate in this trial. They were subject to the same auditory screening as the native subjects and were matched in terms of language education and length of residence in Sweden (3 to 5 years). The age distribution of these subjects is shown in table 1. The mean test result for this group, shown in figure 1, was -6.9dB, SD=1.9. A t-test performed between the native and the L2 group yielded a significant difference at the .005 level.

We have also made an attempt to demonstrate that the newly developed comprehension test could be used not only to test the comprehension ability of individual language users but also to test the quality of speech. 19 of the native Swedish users with normal hearing were given the test with the test stimuli composed of synthetic speech. The speech synthesis was provided by Lars Jäderberg of Speech Craft AB and was of average quality. We reasoned that this could not only demonstrate this use of the test as a estimator of speech quality but also provide still another argument for the validity of our test methodology. The interesting question was whether the test would be sensitive enough to detect the comprehension difficulty which the synthetic speech was assumed to introduce. As seen in figure 1, the mean test result for the synthetic group was .75dB, SD=2,6. A t-test between the normal group and the synthesis group yielded a difference at the significance level of .000.

ISBN 1-876346-48-5 © 2003 UAB

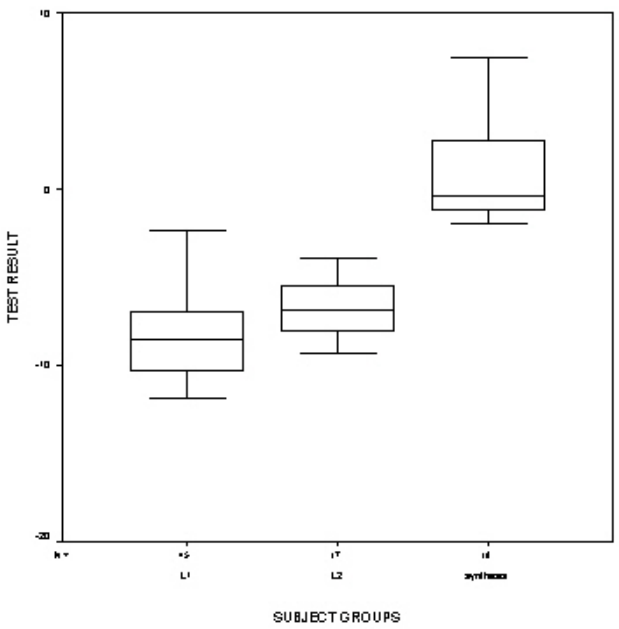

Figure 1: Test results in S/N ratio for normal hearing native speakers of Swedish (L1), normal hearing L2 users of Swedish (L2) and normal hearing native speakers of Swedish with synthetic speech as test stimuli (synthesis).

With support of the above test construction features we would argue that this new speech comprehension test has a high degree of construct validity (Cronbach, 1960[3). We would also contend that the test results of the L2 users and of the trial with synthetic speech would indicate not only that the test may also have concurrent validity and can be used to measure the quality of speech with regard to intelligibility. In the next stage of test development we hope to be able to continue our investigation of test validity with experiments leading to a judgment of concurrent validity. We hope to acquire a basis for this judgment through a comparison of the results of our test with the results of other speech comprehension tests run on the same subjects.

.

REFERENCES

[1] Allwood, J (ed.) *Talspråksfrekvenser frekvenser för ord och kollokationer i svenskt tal- och skriftspråk : frekvenser för ord, ordklasser och kollokationer : jämförelser mellan tal och skrift* Göteborgs universitet, 1999.

[2] Brodda, B., "En algoritm för att bestämma 'avståndet' mellan ord". Forskningsgruppen för Kvantitativ Lingvistik, *KVAL PM* nr 266, 1966

[3] Cronbach, L , *The Essentials of Psychological Testing.* New York, Harper and Brothers, 1960.

[4]Egan, J.P., Articulation Testing Methods. *The Laryngoscope*, Vol 58, no 9, 1948.

[5] Florentine, M., (1985). ”Non native listeners´ perception of American English in noise”. *Proceedings of Inter-noise ´85.* Bremerhaven: Wirtschaftsverlag NW, 1985.

[6] Lidén, G & Fant, G, “Swedish word material for speech audiometry and articulation testing”. *Acta Otolaryng*. suppl. 116, p. 189, 1954.

[7] McAllister, R & Dufberg, M, “Some attempts to measure speech comprehension”, *PERILUS*, vol IX, 1989.

[8] McAllister,. ”Perceptual foreign accent: L2 users comprehension ability”. In: J. Leather & A. James (eds.) *Second-Language Speech*. Berlin: Mouton de Gruyter, 1997.