Test Battery for the Measurement of Second Language Perception

Mats Dufberg and Robert McAllister
Department of Linguistics
University of Stockholm
S-106 91 Stockholm

ABSTRACT

This article describes the tests used in a project dealing with second language perception. The hearing tests could be divided into several types. Three tests are psychoacoustic tests used to estimate basic signal processing ability. Two tests are level based comprehension tests, and one test is a global comprehension test. One written self assessment test and one questionnaire on linguistic preferences are also used.

INTRODUCTION

In a project supported by HSFR we have attempted to measure L2 learners' explicit speech comprehension (in the target language) and to correlate that to certain background factors and self assessment. Their speech comprehension is measured explicitly by means of speech comprehension tests. I will here briefly describe the tests used in the present experiments. McAllister and Dufberg (1994, this volume) report some very preliminary results from the ongoing study.

We have also recorded spontaneous speech of all subjects to be able to, in future work, estimate production accent.

TEST BATTERY

The tests in the battery can be divided into six types:

- Background questionnaire and interview
- Audiometric screening
- Psychoacoustic tests (3 tests)
- Level based speech perception tests (2 tests)
- Global speech perception test
- Self assessment speech perception test

Background questionnaire and interview

The background factors investigated are first, the subject's linguistic preference of Swedish or her first language, and which of the two languages that dominate in her life, and second, when she started to use Swedish and at what age.

Schumann (1978) lists a number of factors that are suggested to affect second language acquisition. We have chosen to limit our study to the above mentioned factors.

Audiometric screening

Békésey pure tone audiometry is used for screening. Since we are, in this project, interested in the second language aspect of speech understanding, we want to make sure

that the subjects do not have any hearing impairments. Subjects with decibel mean of more than 20 dB will be discarded. The decibel mean is calculated as the mean of the levels at 500 Hz, 1000 Hz, 2000 Hz, and 3000 Hz.

Psychoacoustic tests

In the test battery we have included three psychoacoustic tests that are supposed to measure the ability to detect pitch differences, time resolution and signal type, respectively. The three tests have been implemented on a PC with an Ariel signal processor card by Annica Hovmark (Johansson 1988). All three tests have a common test procedure. Three signals are presented in a series to the subject, two of which are identical. The subject has a response box with four buttons. One button repeats the last series of signals. The other three buttons correspond to the three signals. The subject is asked to press the button that corresponds to the deviant signal. The program decreases and increases the difference adaptively until a discrimination threshold can be calculated. Hovmark has implemented two different adaptation principles, one faster and one more accurate (Johansson 1988:4-7). We have chosen the more accurate one. The pitch discrimination test starts with a large frequency difference between two sinusoidal tones. This difference is decreased when the subject gives a correct answer. The level of the tones varies randomly within +/- 3 dB. In the gap detection test, which is supposed to measure time resolution ability, the test signals are bandpass filtered noise (Hovmark 1991). One of the signals has a short break. The length of the break is adaptively shortened down to the threshold. The third test is a signal type test. Two signals are noise signals and one is a sinusoidal tone. The length of the signals are shortened until the subject's threshold is reached.

Since our subjects are and will be normal hearing persons, we expect them, firstly, to perform well on these tests, which means having a low threshold. Secondly, we expect the results to show a fairly small inter-group data spread.

Word Test — a word comprehension test

Our word comprehension test is one of our two level test. As opposed to the QAR test (below) this test is supposed to test a specific linguistic level.

The speech material of the Word Test consists of two-syllable words. The words have been chosen according to a number of criteria. All words belong to an open wordclass, i.e. nouns, verbs, adjectives or adverbs. The words have been chosen as the most common and frequent words of Swedish, that also fulfill the other criteria. They are equally divided between the two word accents of Swedish. And the phoneme distribution has been balanced.

The words are played back masked with noise, the same noise used in the QAR-test (Dufberg 1990). In the beginning the signal to noise ration is kept high (low noise level). The subject is asked to repeat the word as well as she can (we will return to how we judge the subject's response). We increase and decrease the level of the noise adaptively until the threshold can be calculated. This is the signal to noise ratio at which the subject repeats fifty per cent correctly.

In other word repetition tests (e.g. the standard speech audiological HTT-test) the repetition is judged as either being correct or incorrect. If the target word is /mɛst/ then it is considered to be as incorrect to answer [bɛst] as to answer [høns]. Obviously, the subject has decoded a greater part of the heard word in the first case than in the second. To be able to measure the *distance* between the target and in the response in greater detail, we implemented a phonological distance metric developed by Benny Brodda

(1967). A response is judged to be incorrect (i.e. 0% correct) if the phonological distance between the target and response is on the random level. Between the two extremes incorrect and correct a value between 0 and 1 is assigned.

Top Down test

Our second level test is a test that measures the ability for using the context for phoneme restoration, i.e. a test that measures the use of *signal independent* information. (For a discussion of *signal dependent* and *signal independent* information see McAllister and Dufberg, 1989.) The test has been designed by Rolf Lindgren.

The speech material in this test consists of sentences with a specific *target phoneme*. The target phoneme is always a stop consonant and is always word initial. The target word (i.e. the word with the target phoneme) has a known probability given the context before the word, and this probability varies between target words in the test sentences. (For a description of the test material see Lindgren, 1982.) The target phoneme itself is not present on the test tape, but is substituted by a noise pulse. The task for the subject is to react on the target phoneme, i.e. to press a button as quickly as possible when she "hears" it. The rationale behind the test is that an increased probability will shorten the reaction time *if you can utilize the context*.

QAR — Question and response

The QAR-test, which stand for Question-and-Response, is a global comprehension test. The test material consists of short texts with accompanying yes–no questions. The material is read by a woman with a neutral voice and presented in masking noise. The procedure is adaptive, which means that the signal to noise ratio is lowered (more noise) if the subject has answered correctly, and increased if he/she has answered incorrectly. The design of the test was reported briefly in Dufberg (1990). Some modifications have been made since that paper.

The result of the QAR-test is the signal to noise ratio at which the subject gives correct answers to fifty per cent of the questions.

CONCLUSION

Our aim has been to use tests with high degree of validity. Since we will attempt to correlate the L2 learners' speech comprehension with their production and with certain background factors, it is fundamental that the comprehension tests do measure what we want them to test. In a previous article (McAllister and Dufberg 1989) we present the model of speech comprehension on which we have based our work. With this battery we have tried to satisfy both the validity requirement and the necessity to have manageable tests.

REFERENCES

Brodda, B. 1967. 'Perceptual distances between phonemes.' Unpublished paper, Dept. of Linguistics, Stockholm University.

Dufberg, M. 1990. 'The measurement of speech comprehension: A progress report of the construction of a global test.' *Phonum* 1, 96–99. Umeå: Dept. of Phonetics, University of Umeå.

Hovmark, A. 1991. 'Signal analysis and speech perception in normal and hearing-impaired listeners.' In *Perilus* XIII (Papers from the fifth National Phonetics Conference), 117–120. Stockholm: Dept. of Linguistics, Stockholm University.

Johansson (Hovmark), A. 1988. 'Psykoakustiska tester med dator'. Unpublished paper. Stockholm: Speech Transmission Laboratory, Royal Institute of Technology (KTH)

Lindgren, R. 1982. 'Informationsteoretiskt viktade testsatser för testning av talperception.' Stockholm: Dept. of Linguistics, Stockholm University and Dept. of Audiology, Södersjukhuset.

McAllister, R. and M. Dufberg. 1989. 'Some attempts to measure speech comprehension.' In *Perilus* IX, 29–52. Stockholm: Dept. of Linguistics, Stockholm University.

McAllister, R. and M. Dufberg. 1994. 'Aspect of the perception and production of a second language: A progress report.' (This volume.)

Schumann, J. H. 1978. 'Social and psychological factors in second language acquisition'. In Jack C. Richards (Ed.), *Understanding second and foreign language learning*, 163–178. Rowley, Mass.: Newbury House Publishers.