Testing listening effort for speech comprehension using the individuals’ cognitive listening spare capacity

N. Rönnberg,1,3 S. Stenfelt,1,3 M. Rudner2,3

1Technical Audiology, Department of Clinical and Experimental Medicine, Linköping university, Linköping; 2Department of Behavioural Sciences and Learning, Linköping university, Linköping; 3Linnaeus Centre HEAD and Swedish Institute for Disability Research, Linköping, Sweden

Introduction

Most hearing aid fittings today are almost solely based on the patient’s audiogram. Although the loss of gain in the cochlea is important, for a more optimal fitting, more individual parameters of the patient’s cochlear loss together with the patient’s cognitive abilities to process the auditory signal are required (Stenfelt & Rönnberg, 2009; Edwards, 2007). Moreover, the evaluation of the fitting is often based on a speech in noise task and the aim is to improve the individual patient’s signal to noise ratio (SNR) thresholds. As a consequence, hearing aid fitting may be seen as a process aimed to improve the patient’s SNR threshold rather than to improve communication ability. However, subsequent to a hearing aid fitting, there can be great differences in SNR improvement between patients that have identical hearing impairment in terms of threshold data (the audiogram). The reasons are certainly complex but one contributing factor may be the individual differences in cognitive capacity and associated listening effort.

Another way to think about amplified hearing is to ease a subject’s listening effort (Sarampalis, et al., 2009). When the speech signal is degraded by noise or by a hearing impairment, more high-order cognitive or top-down processes are required to perceive and understand the signal, and listening is therefore more effortful. It is assumed that a hearing aid would ease the listening effort for a hearing impaired person. However, it is not clear how to measure the listening effort. We here present a test that will tap into the different cognitive aspects of listening effort, the Auditory Inference Span Test (AIST). The AIST is a dual task hearing in noise test, that combines auditory and memory processing and is well suited as a clinical test for listening effort.

Background

Working memory appears to play a critical role in speech understanding and comprehension (Rönnberg, et al., 2008). Studies have shown significant correlations between cognitive capacity, i.e. working memory capacity, and hearing in noise (Akeroyd, 2008). Individuals with equal hearing impairment as measured by hearing thresholds can demonstrate quite different levels of speech understanding in noise due to differences in cognitive capacity.

An individual with higher cognitive capacity is likely to better cope with worse SNR, than an individual with a lower cognitive capacity (Lunner, 2003). In the study by Lunner (2003), cognitive capacity was measured by the reading span test that requires simultaneous cognitive processing and storage (Daneman & Carpenter, 1980). Rudner, et al., (2009) and Foo, et al., (2007) also showed significant correlations between cognitive capacity, as measured by the reading span test, and individual performance of aided speech recognition in noise using both Hagerman sentences and the Swedish version of the HINT.

It has been shown that the ability to hear speech in noise increases with increased cognitive capacity, i.e. greater working memory capacity. Moreover, the difference in speech in noise score correlated with cognitive capacity is more evident in a worse SNR, which indicates that individuals with lower cognitive capacity are more affected by the noise than individuals with higher cognitive capacity (Gatehouse, et al., 2003). The allocation of the individual’s limited capacity to the processing and storage functions of working memory varies with task demands, such as when the level of background noise increases during listening. When listening in a noiseless environment, processing information is easy, the individual can hear easily, and therefore can use more working memory capacity to store the information. However, when the noise increases the individual needs more processing to hear, and has less capacity left for storage of the information. Working memory capacity is limited and differs between individuals. An individual with greater working memory capacity will have more storage capacity compared to an individual with poorer working memory capacity, when listening in a noisy environment (Pichora-Fuller, 2007). On the basis of the above mentioned studies, as well as others showing similar results, it has been argued that audiologists should include cognitive tests in the clinic to help identify the most beneficial assistive technology for an individual and also the appropriate hearing aid fitting (Pichora-Fuller & Singh, 2006). The AIST was developed with this in mind.

The test setup

Initially, the AIST was implemented as a web-based test, which proved to be an easy way to reach test subjects. The drawback was that the web-based test was conducted in an uncontrolled environment with several uncertainties (e.g. choice of sound card and headphones, if subjects took a short break during the test, or even made notes...
The test was conducted in speech weighted noise at 0 dB SNR. Before the test started, questions about the subject’s age, sex, educational level (Elementary school, High school, College/University, Post graduate) were given and the subject rated themselves as having a hearing impairment or not (Table 1). The test required complete answers and besides answering the questions, the subjects stated their certainty level of the accuracy of the answers.

In total, the test consisted of six lists of three sentences. Before the test started, the participant was familiarized with a practice list. Each subset of the test comprised three sentences, three sentence questions, three information questions, and an estimation of the perceived listening effort (Figure 1).

**Speech material**

The test used the Swedish matrix sentences, the Hagerman sentences (Hagerman, 1982; Hagerman & Kinnefors, 1995). These sentences consist of five words in a structured sequence: name, verb, number, descriptor, item.

For example (translated from Swedish):
1. Britta moves eight black rings.
2. Elsa gave six new gloves.
3. Peter bought seven light bowls.

**Sentence questions**

After each Hagerman sentence there was a question to verify speech perception, termed sentence question. This asked which of three alternative words that occurred in the sentence. Any of the five words in the sentence were candidates. The aim with this question was to force processing of the auditory stimuli and prevent the test subject from only focusing on word/sentence storage. This question was also a control of word recognition (that should be close to 100%). The result was not used in the test but is here included for verification purpose (99.3% correct; n=41).

**Information questions and memory load levels**

Subsequent to the three sentences there were three information questions. The information questions were divided in three different memory load levels.

Level #1 questions were simple memory tests of a word in one of the sentences. Three choices were given where only one of the words was in one of the preceding sentences. An example of a level #1 question was: Which of the following names were used in the sentences?

Questions could also refer to numbers or items given in the preceding three sentences.

Level #2 questions were complex memory tests that referred to two words in one of the sentences. There was a need to relate and process the information given in the sentences to be able to give a correct answer. An example of a level #2 question was: Which item did <name> have? Where <name> was the name of one of the persons in the preceding three sentences and the correct answer was the item for that person according to the sentence.

Level #3 questions were complex memory and processing tests that referred to two words in each of the three sentences. There was a need to relate and process the information given in the sentences to be able to give a correct answer. An example of a level #3 question was: Who had the largest number of items that was even? Answer alternatives for this question were all names used in the preceding three sentences.

All information questions were three-choice questions and the answer alternatives were correct words from the closed-set of Hagerman sentences. However, only one alternative was correct to each question and in each subset. The answers given were recorded and scored (1 point for a correct response and 0 points for an incorrect response), and the reaction time was recorded. Only one memory load level was tested at a time following a set of three sentences. The order of memory load levels was balanced to eliminate order effects.

**Perceived listening effort**

After the information questions the subjects were asked to rate their listening effort. The listening effort was given on a ten-point scale, with the endpoints No effort and Greatest possible effort. The effort given was encoded as an integer between 0 and 9, where 0 indicated no effort and 9 the greatest possible effort.

**Results**

Analyses of variance were performed on the data, one for the accuracy and one for the reaction time. For both there were main effects of memory load level (accuracy: F(2,120) = 22.4, p<0.001; reaction time: F(2,120)=19.5, p<0.001). This demonstrates that the greater the load on the working memory the poorer the score, as shown in Figure 2.

The maximum accuracy score for each memory load level was six points, since there were six questions at each level. For level 1 questions the average score was 5.1 (85%) which was significantly (p<0.05) higher than for level 2 questions, 4.4 (73%). Level 3 questions had a significantly (p<0.001) lower score, 3.4 (57%). This showed that the greater the demand on working memory, the lower the score. Random guessing gives an average score of 2.0 (33%).

The relation between reaction time and memory load level is shown in Figure 3. Between level 1 and level 2 questions there was no significant difference in reaction time (4.7 respectively 5.0 seconds). However, the accuracy score differed 0.7 points, as shown in Figure 2,
indicating that it was more demanding to remember a level 2 question compared to a level 1 question, but the processing time did not differ.

There were significantly longer reaction times for the level 3 questions, with an average of 8.3 seconds (p<0.05). This indicates that a greater amount of processing of the information was required before an answer could be given. It should be noted that the reaction time also included the time to read and answer the question.

Figure 4 shows the self-rated listening effort for the three levels of memory load. As can be seen, the perceived effort differed between subjects but was on average similar for all levels. For level 1 the average listening effort was 4.4 (SD=2.2), for level 2 it was 4.5 (SD=2.1), for level 3 it was 4.7 (SD=2.1), and the grand average for all levels of memory load was 4.6 (SD=2.1). There were no significant differences in perceived listening effort between the different levels of memory load.

Since the sentences were all presented at the same SNR, the perceived listening effort should be the same for all memory load levels. We controlled this by comparing the rated listening efforts within each subject, shown in Figure 5. Each memory load level was tested twice (which gives a total of 6 tests), and the rating at the two memory load 1 tests was first controlled. The box-plot on the left-hand side in Figure 5 shows the difference in listening effort rating between the two memory load level 1 tests. As can be seen, the median difference is zero and almost all subjects are within 1 rating unit between the two testing times. The next two box-plots in Figure 5 shows the difference in averaged rated listening effort between memory load level 2 and 1 (middle box plot) and memory load level 3 and 1 (right-hand side of Figure 5). These data show small differences in self-rated listening effort between different memory load levels for each subject. However, when analysing Figure 4 there was a great spread in self-rated listening effort between subjects.

The median for intra subject listening effort for each level of memory load is at 0 and the standard deviation for all intra-subject self-rated listening effort was 0.5. Moreover, 36% of the test subjects did not change their stated perceived listing effort between the lists of sentences or levels of memory load. This indicates that the subjects were consistent when stating the perceived listening effort.

**Discussion and Conclusions**

The presented study has shown that AIST performance decreased with increasing levels of memory load, and that reaction time increased for memory load level 3. According to the current results and previous studies, it is reasonable to assume that the current test probes the cognitive capacity of an individual.

The aim of the test is to probe cognitive aspects of listening effort. Some of the test subjects did not change their stated perceived listening effort between the levels of memory load. The likely explanation is that since the SNR was the same for all sentences, the perceived listening effort was stated as the same. This is interpreted as an indication that the test subjects were consistent in stating the listening effort. However, there were large variations of stated listening effort between subjects. This indicates that subjects experience the same SNR to make different demands, that may be related to both auditory and cognitive capacities. It should be remembered that all subjects had normal hearing and differences in hearing ability could not explain the results.

In a future study, to verify that the AIST is sensitive to cognitive capacity, the test will be evaluated with measurements of the subjects’ cognitive capacity as well as the their hearing thresholds. We believe that the test will show correlations between SNR of the speech material, number of correct answers, the reaction time, and the perceived effort. For example, we believe that when the SNR gets worse the score will also get worse, the reaction time longer, and the perceived effort...
greater (a relation also shown by Larsby, et al., 2005). It is also expected that the score of the AIST will correlate to the test subjects' cognitive capacity. For example, a test subject with poorer cognitive capacity will score worse, have longer reaction time, and perceive greater effort, than a test subject with greater cognitive capacity. It is also expected that this difference will be greater when the SNR gets worse (a relation also shown by Gatehouse, 2003).

For a clinical test the requirement is that it is fast and easily administered. The AIST takes no more than fifteen minutes to complete, and the aim is to further shorten the time and adapt the test for clinical use. This ensures that the AIST will be a useable instrument for testing listening effort using the individuals' cognitive spare capacity.

References


