SAHLGRENSKA ACADEMY

INSTITUTE OF NEUROSCIENCE AND PHYSIOLOGY

LISTENING EFFORT IN ADVERSE LISTENING CONDITIONS

A pupillometry study on hearing impaired listeners

Authors:

Elias Nyhlén

Johan Nyberg

Thesis:	Scientific thesis, 15hp
Program and course:	Programme in Audiology, AUD620
Level:	First Cycle
Semester/year:	St 2018
Supervisors:	Dorothea Wendt, Thomas Lunner
Examiner:	André Sadeghi
Report no:	xx (not to be filled in by the student/students)

UNIVERSITY OF GOTHENBURG

SAHLGRENSKA ACADEMY

INSTITUTE OF NEUROSCIENCE AND PHYSIOLOGY

Abstract

Keywords: Listening effort, motivation, pupillometry

Purpose: The purpose of this study is to investigate if the motivation of the test subject influences the pupil dilation in different listening conditions. The goal is to assess effort involved in speech recognition with both rating scales (as an indicator of perceived effort) and pupillometry (as an indicator of objective effort).

Method: Fourteen subjects participated in the study, of which 3 were women and 11 men. The subjects were presented three lists of Danish HINT (Hearing in noise-test) with 4-talker babble noise. HINT was performed on one easy, one medium and one hard condition. During the HINT, the subject wore Pupil Labs eye-tracking headset to measure the peak pupil dilation (PPD). After each list, the subjects rated their perceived effort and motivation on a questionnaire consisting of five questions.

Result: The results from this study indicates a significant increase in listening effort for both the perceived effort and the PPD from the easy to the medium condition. However from the medium to hard condition, the PPD decreased while the perceived effort showed no significant difference. The results also shows that the motivation decreases with increasing task demands.

Conclusion: This study shows that when task demands get higher, the subjects' motivation decreased. There was a difference between the perceived effort (as measured by the questionnaire) and the objective effort (as measured by the pupillometry). The perceived effort indicate that the hard condition is at least as effortful as the medium condition, while the objective effort indicate that the subjects spend less effort in the hard condition compared to the medium condition.

SAHLGRENSKA AKADEMIN

INSTITUTIONEN FÖR NEUROVETENSKAP OCH FYSIOLOGI

LYSSNINGSANSTRÄNGNING I UTMANANDE LYSSNINGSFÖRHÅLLANDEN

En pupillometristudie på personer med hörselnedsättning

Författare:

Elias Nyhlén

Johan Nyberg

Examensarbete:	15 hp
Program och/eller kurs:	Audionomprogrammet, AUD620
Nivå:	Grundnivå
Termin år:	Vt 2018
Handledare:	Dorothea Wendt, Thomas Lunner
Examinator:	André Sadeghi
Rapport nr:	xx (ifylles ej av studenten/studenterna)

GÖTEBORGS UNIVERSITET

SAHLGRENSKA AKADEMIN

INSTITUTIONEN FÖR
NEUROVETENSKAP OCH FYSIOLOGI

Abstrakt

Nyckelord: Lyssningsansträngning, motivation, pupillometri

Syfte: Syftet med denna studie är att undersöka om motivation hos testpersoner påverkar pupillens storlek i olika lyssningsförhållanden. Målet är även att bedöma ansträngning i olika lyssningsförhållanden med både skattningsskalor (som ett mått av upplevd ansträngning) samt pupillometri (som ett mått av objektiv ansträngning).

Metod: Fjorton testpersoner deltog i studien, varav 3 kvinnor och 11 män. Testpersonerna fick lyssna på tre listor av danska HINT (Hearing in noise-test) där bruset bestod av fyra talare. HINT genomfördes i ett lätt, ett medelsvårt och ett svårt förhållande. Under tiden HINT genomfördes mättes peak pupil dilation (PPD) med Pupil Labs eye-tracking headset. Efter varje lista fick testpersonerna skatta deras upplevda ansträngning och motivation i ett frågeformulär bestående av fem frågor.

Resultat: Resultaten från denna studie indikerar att det finns en signifikant ökning av lyssningsansträngning för både upplevd ansträngning samt PPD mellan det lätta och medelsvåra förhållandet. Mellan det medelsvåra och svåra förhållandet kunde man däremot se att PPD sjönk medans den upplevda ansträngningen ej visade någon signifikant skillnad. Resultaten visade även att motivationen sjönk med stigande svårighetsgrad på lyssningsförhållandet.

Konklusion: Denna studie visade att när lyssningsförhållandet blir svårare sjunker testpersonernas motivation. En signifikant skillnad uppmättes mellan upplevd ansträngning (mätt med frågeformulär) och objektiv ansträngning (mätt med pupillometri). Den upplevda ansträngningen indikerar att det svåra förhållandet är minst lika ansträngande som det medelsvåra, medans den objektiva ansträngningen indikerar att testpersonerna anstränger sig mindre på det svåra förhållandet jämfört med det medelsvåra.

Acknowledgement

We would like to thank our supervisors Dorothea Wendt and Thomas Lunner for guiding us through this project. We would also like to thank Oticon, Eriksholm Research Centre, Renskje K. Hietkamp, Sanne Mehrfeld Møller, and Patrycja Książek for the help and support.

The work was divided equally between the authors.

List of acronyms/abbreviations

AU: Arbitrary units

HINT: Hearing in noise test

NASA: National aeronautics and space administration

PPD: Peak pupil dilation

PTA4: Pure tone average for 0,5, 1, 2 and 4 kHz

VAC+: Voice aligned compression

VAS: Visual analogue scale

SD: Standard deviation

SNR: Signal to noise ratio

WDRC: Wide dynamic range compression

Definition of terms

Azimuth: The angular position of a certain object in relation to the listener in the horizontal plane. 0° azimuth refers to straight in front of the listener, and 180° refers to straight behind the listener.

Linear interpolation: A method to estimate missing data values. This is done by creating an estimated value based of the average of the value before and after the missing value.

Listening effort: The amount of cognitive resources spent during a listening task.

Motivation: The desire or willingness to do a certain task. Specifically for listening, motivation is the desire or willingness to pay attention and listen to a certain task.

Objective listening effort: The listening effort measured through objective methods such as the peak pupil dilation.

Perceived effort: The subjective experience of how much effort is spent during a task.

4-talker babble: The noise used during the HINT. This noise consists of two male and two female talkers that are reading different texts simultaneously.

Table of Contents

1. INTRODUCTION

It is well known that listeners with hearing loss can experience difficulties when communicating and understanding speech. Communication difficulties as a result of hearing loss has shown by previous studies to be associated with reduced quality of life (Dalton, Cruickshanks, Klein, Klein, Wiley, & Nondahl, 2003). In order to better their chances of hearing, hearing impaired listeners have to spend more of their cognitive resources focusing on listening compared to listeners without hearing loss. This heightened focus of cognitive resources on listening has been mentioned as an increase of *listening effort* (Downs, 1982; McCoy, Tun, Cox, Colangelo, Stewart & Wingfield, 2005). As a result of increased listening effort hearing impaired listeners can be at extra risk of mental fatigue and distress (Kramer, Kapteyn, & Houtgast, 2006).

Listeners exposed to increased listening effort may stop putting in effort in order to avoid experiencing mental fatigue. In a recent workshop, effort was defined as "*the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task, with listening effort applying more specifically when tasks involve listening"* (Pichora-Fuller et al., 2016). In an attempt to explain the consequences of including motivation in the definition of effort, a 3D-plot of how motivation might work as a function together with demands to affect listening effort was created (Figure 1). Motivation is explained as the desire or willingness to do something, and this is the definition we will use when referring to motivation. Related to listening, motivation would be the willingness or desire to hear what is being said. The 3D-model suggests that listening effort increases based on task demands as well as motivation (Pichora-Fuller et al., 2016). With low task demands the listening effort will be at low levels as well, even though the listener's motivation is low. But when task demands gets higher it requires that the listener is motivated to experience high listening effort. If the listener's motivation is low during high task demands the objective listening effort will stay at low levels (Pichora-Fuller et al., 2016).

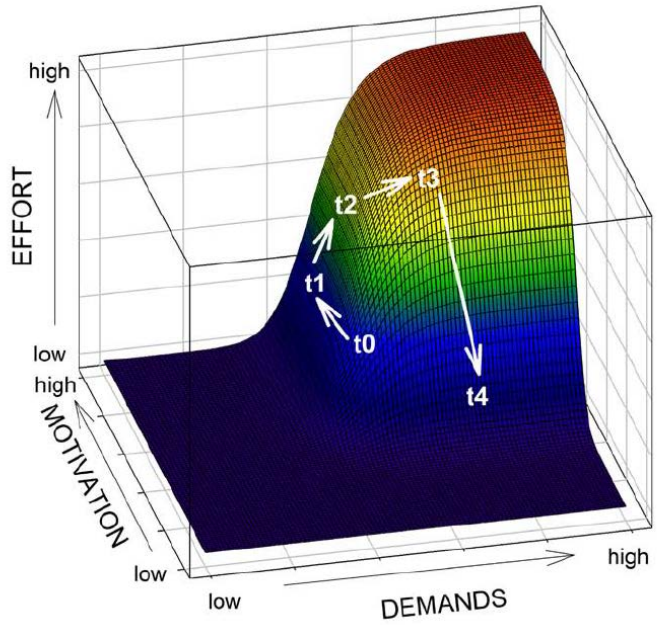


Figure 1. *Effort as a function of motivation and task demands (Pichora-Fuller et al., 2016).*

1.1 Subjective rating of perceived effort

One common method to measure the listening effort in subjects is by letting them rate their perceived effort (Larsby, Hällgren, Lyxell & Arlinger, 2005; Zekveld, Kramer & Festen, 2010; Rudner, Lunner, Behrens, Thorén & Rönnberg, 2012). In these studies, the test subjects are usually exposed to different listening tasks where they will then rate their perceived effort during these tasks. Lemke et al. (2016) describes the term perceived effort as the subjective experience of how taxing a task is or was. This is the definition of perceived effort we use in this report. As the objective task demand increases or decreases, we can then measure the perceived effort in different conditions. To create these different listening conditions one common way is using the Hearing in noise test (HINT) that was created to measure sentence intelligibility in quiet and in noise (Nilsson, Soli & Sullivan, 1994). HINT consists of 25 phonemically matched and balanced lists with ten sentences per list. The sentences are five words long and read by a male speaker. Subjects are presented one list at a time and are asked to repeat as many of the five words from each sentence as possible. HINT was originally created to be used as an adaptive method, where the SNR changes for each sentence based on the score of the previous sentence, but can also be used to find out a subject's speech intelligibility at a certain SNR. In this case the SNR will be the same during the whole list of ten sentences and the results presented as percentages of the words or sentences repeated

correctly. A Danish HINT consisting of ten lists, containing 20 sentences each, and three practice lists were created in 2011 based on the HINT (Nielsen & Dau, 2011).

Usually the participants are asked, at the end of a listening task, to "*rate your effort*" on a scale (Zekveld et al., 2010; Rudner et al., 2012). Questions regarding performance rating is also sometimes used together with the effort ratings. As task demands increases, the perceived effort increases and the subjective perceived performance decreases (Zekveld et al., 2010). The NASA task load index questionnaire is commonly used for assessing workload in general. This questionnaire includes several questions regarding different aspects of perceived workload. The NASA task load index is broadly used and can be applied to many different fields (Hart & Staveland, 1988). One of the questions in the NASA task load index questionnaire measures the perceived effort during a certain task. This question has been used in the hearing science field to evaluate the perceived listening effort (Wu, Stangl, Zhang, Perkins & Eilers, 2016; McGarrigle, Dawes, Stewart, Kuchinsky & Munro, 2017).

1.2 Rating scales

As this is a subjective rating of perceived effort, one must rate the effort on some type of scale. For subjective ratings, different type of scales has been used. The Visual analogue scale (VAS) is a continuous scale that ranges from two points without discrete numbers or sections between these points. There are sometimes labels along the scale to help guide the subject in their rating. To the best of our knowledge, this scale together with the Likert scale is the most widely used in hearing science for evaluating perceived effort. The Likert scale is similar to the VAS, but differs in that it has discrete points instead of a continuous scale. The ranges and labels of the scales used in hearing research is different, and can range for example from "no effort" to "maximum possible effort", 0-10, 1-9, etc. (Zekveld et al., 2010; Rudner et al., 2012; McGarrigle et al., 2017).

Previous studies that assessed the perceived effort in different listening conditions found that with increasing speech intelligibility, the perceived effort decreased, see Figure 2. For the HINT sentences, increased speech intelligibility is achieved by increasing the signal to noise ratio.

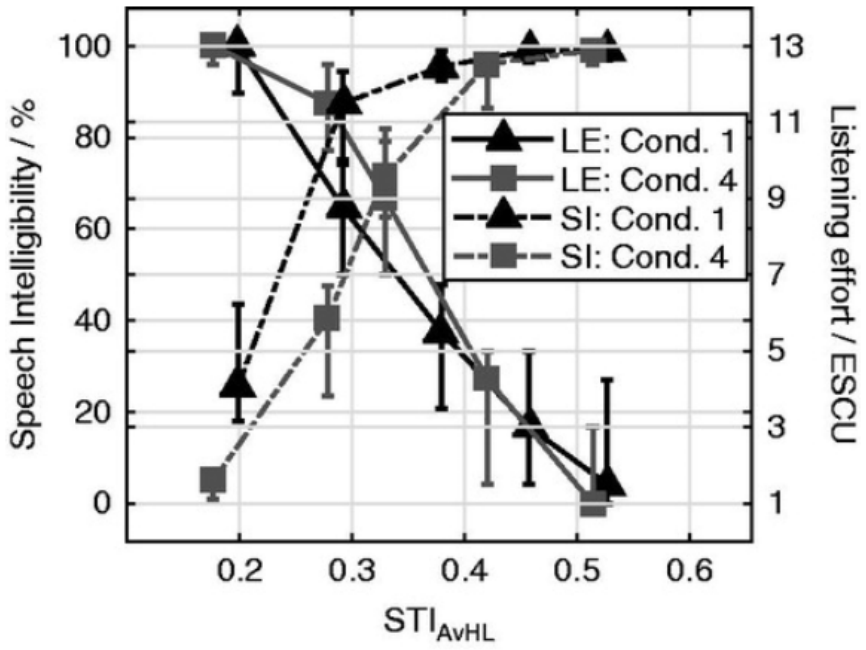


Figure 2. *LE=listening effort, SI=speech intelligibility. With decreasing speech intelligibility, the listening effort increases. (Schepker, Haeder, Rennies, & Holube, 2016).*

1.3 Pupillometry

The subjective rating for evaluating effort has been used for many years. However with new technology, objective methods for evaluating effort is now available. One method to objectively measure the listening effort is to measure the pupil dilation response. The subject is exposed to a listening task, while the dilation of the pupil is continuously measured. It has been suggested that the pupillary response is a good measure for effort (Kahneman, 1973). More recent research has then indicated that the pupillary response also could reflect the listening effort (Kramer, Kapteyn, Festen, & Kuik, 1997; Zekveld, & Kramer, 2014; Pichora-Fuller et al., 2016; McGarrigle et al., 2017). The increased pupil size from a task can be measured in different ways and to measure the momentary cognitive load the peak pupil dilation (PPD) is a good measure (Pichora-Fuller, 2016). The PPD is often defined as the change from resting state to the biggest dilation that can be measured from the response of a task. By changing the task demands, for example by increasing the SNR, we can then measure how the task demands influences the PPD. Studies show that the PPD reaches a turning point at medium task demands, where PPD is at max. From this point, lower SNR and higher SNR results in decreasing PPD, see figure 3.

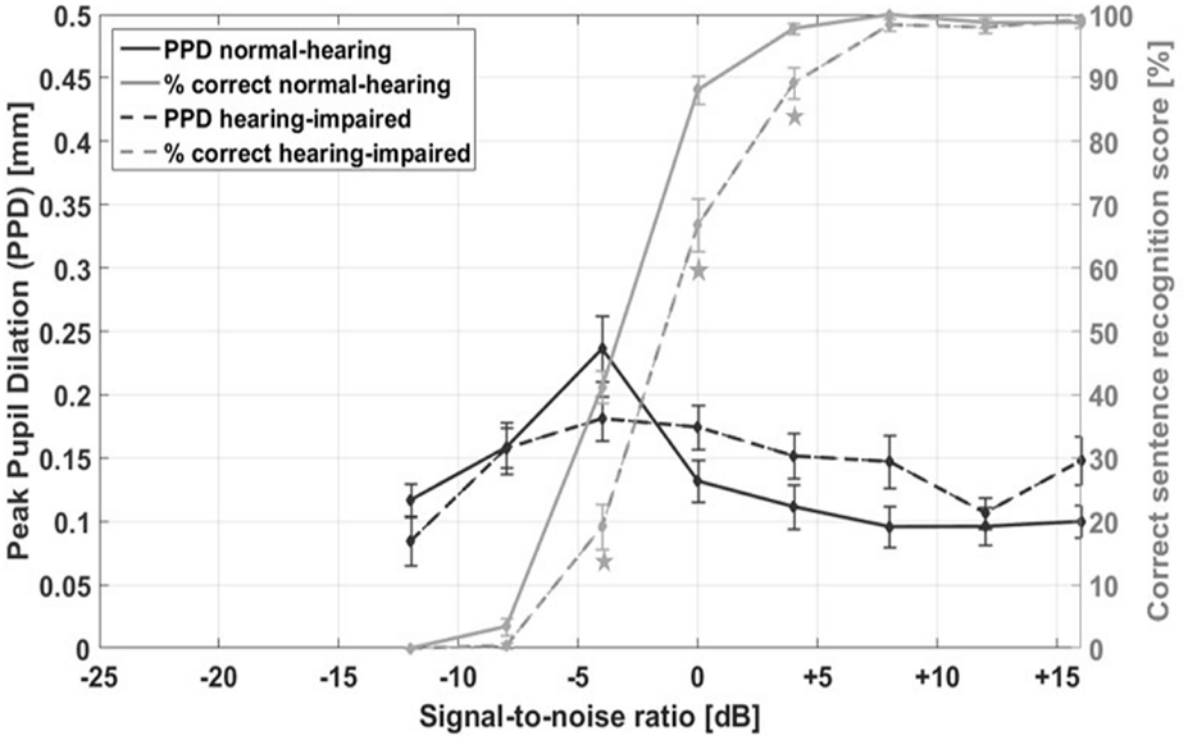


Figure 3. *Peak pupil dilation across a range of signal-to-noise ratios in normal-hearing and hearing-impaired listeners (Ohlenforst et al., 2017).*

As mentioned earlier, by decreasing the SNR of a listening task, the perceived listening effort as measured by the subjective rating increases (Figure 2). However, the listening effort as measured by pupillometry increases with decreasing SNR up to a certain turning point, and then decreases with further decreasing SNR. Other studies have shown that the perceived effort doesn't correlate to the physiological indicators of listening effort (Gosselin & Gagne, 2011; Mackersie & Cones, 2011; Koelewijn, Zekveld, Festen & Kramer, 2012).

One explanation of why the listening effort as measured by subjective rating versus pupillometry does not match could be that they do not measure the same thing. The pupillometry might measure the actual physiological resources spent during a listening task, while the subjective rating in certain scenarios measures the expected resources spent during a task. If a listening task is really difficult the subject might expect that he/she should be spending a lot of resources, and rates the listening effort as high, while the actual resources spent is low. The reason why the resources spent might be low is because the subject's motivation is low, and he/she has given up trying to listen. According to Pichora-Fuller et al. (2016), the motivation of the test subject might influence the listening effort. We believe that motivation could be the difference between these two measures of listening effort. Motivation

of the test subjects during listening tasks has been assessed in previous research, however only on normal hearing subjects (Zekveld et al., 2014). This needs to be studied on hearing impaired subjects to improve the relevance in the field of hearing science.

2. PURPOSE/AIM

The purpose of this study is to investigate if the motivation of the test subject influences the pupil dilation in different listening conditions. The goal is to assess effort involved in speech recognition with both rating scales (as an indicator of perceived effort) and pupillometry (as an indicator of objective listening effort).

3. RESEARCH QUESTIONS

1. How does perceived effort (as measured by the questionnaire) and objective listening effort (as measured by the PPD) change with increasing task demands?

2. How does perceived effort and objective listening effort relate to each other?

3. How does motivation change with increasing task demands?

4. METHOD AND MATERIALS

4.1 Participants

Thirty-three test-subjects with sensorineural bilateral hearing loss using hearing aids since at least 6 months ago were invited, out of which 14 were women and 19 were men. Twenty of the invited participated in the measurements. Some test subjects had to be excluded due to low scores in the speech intelligibility test. Fourteen subjects were then used for the analysis, whereof three were women and 11 were men. The ages of these 14 test subjects had a mean-average of 68 years and ranged from 47 to 75. Pure tone average for 0,5, 1, 2 and 4 kHz (PTA4) for the participants were not lower than 25 dB or higher than 70 dB. The mean PTA4 of the subjects were 46 dB (SD=9,3 dB) for right ear and 42 dB (SD=10 dB) for left ear. The subject's PTA4 difference between left and right ear was lower than 15 dB to avoid an asymmetric hearing loss. Figure 4 shows the mean hearing thresholds for the test subjects. None of the participants suffered from cataract, glaucoma or had previously gone through eye-surgery as this would complicate the assessment of the pupil. The participants were

recruited from Eriksholm Research Centre's patient database. They were sent a written information of the test before accepting being a part of the experiment.

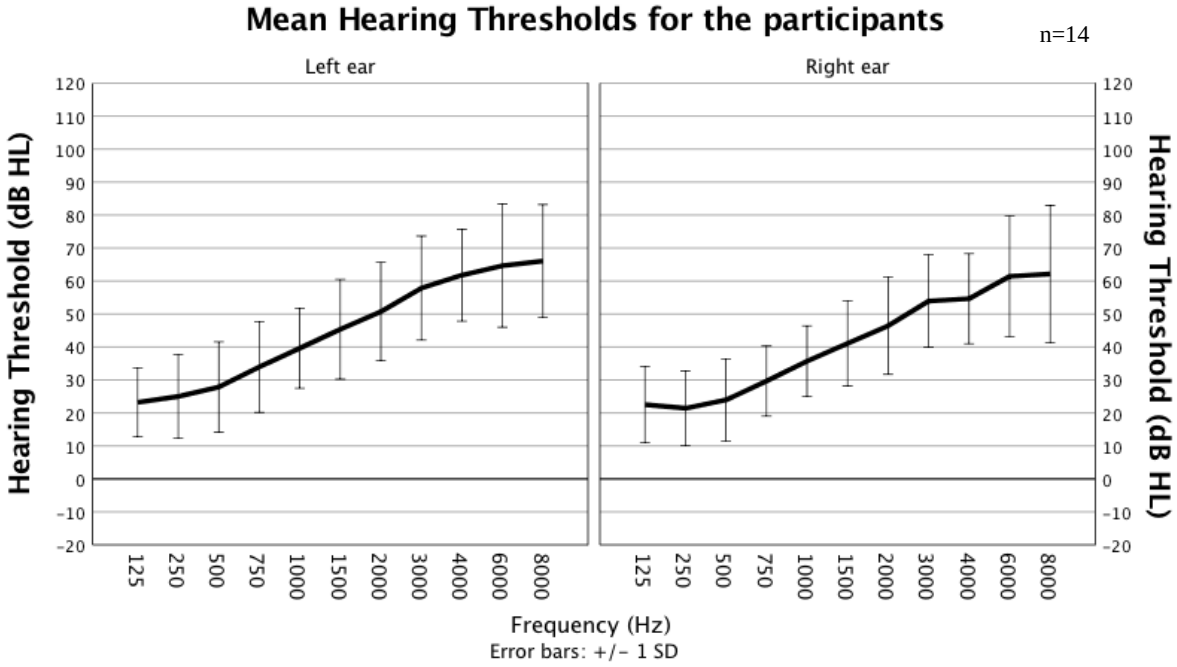


Figure 4. *Mean hearing thresholds for left and right ear of the participants in the study. n=14*

4.2 Test setup and overview

During the whole experiment, the participants were sitting in a double walled sound proof room wearing hearing aids. The speaker setup was similar to what is described in Wendt et al. (2017), see figure 5. Five loudspeakers (Genelec 8030A) were placed in a free field setup in the sound proof room, with the test subject in the middle. The speakers were each placed 1.2 meters from the test subject. One in front, 0° azimuth, that presented the speech material, and four to the side and in the back at 90°, 150°, 210° and 270° azimuth that presented the noise. The noise and HINT sentences were presented through a MATLAB script. The participant was asked to repeat as many of the words as possible during the short break between each sentence. After each list of 25 sentences, the test subject filled in a questionnaire. The questionnaire contained questions regarding effort, performance and motivation. This procedure was repeated three times in different signal-to-noise ratios. The PPD was monitored throughout the experiment by using the Pupil Labs eye tracking headset, see figure 6.

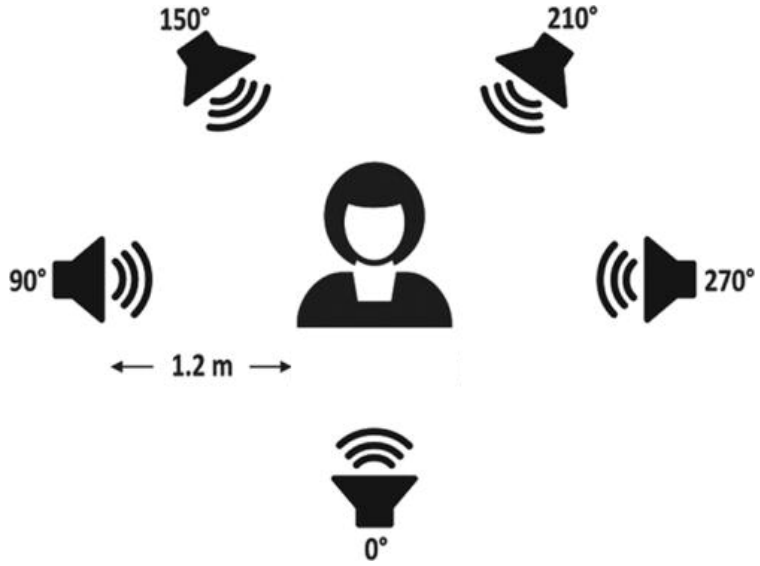


Figure 5. *The speaker setup used in this study. Modified figure, originally from Wendt et al. (2017).*

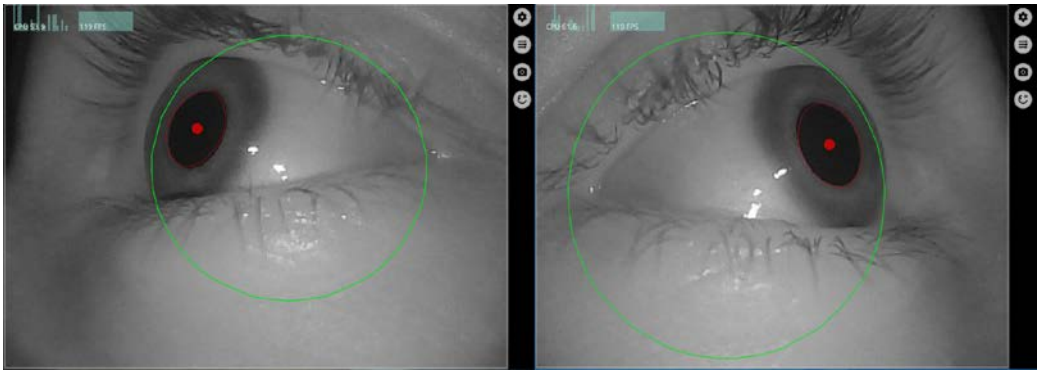


Figure 6. *Screenshot from Pupil Labs eye-tracking software. The red circle outlining the pupil indicates the estimated pupil dilation.*

4.3 Hearing aid fittings

The test subjects wore hearing aids throughout the experiment that were programmed specifically for this study. All the subjects were fitted bilaterally with Oticon Opn1, speaker unit 85 and Oticon Grip Tip with no ventilation. The amplification method used was Voice Aligned Compression (VAC+) individually fitted to each participant's hearing loss. The VAC+ approach falls within the family of wide dynamic range compression (WDRC). WDRC is essentially the standard for fitting sensorineural hearing loss. The amplification was set to experienced user. Noise reduction was turned off, and the microphones set to omnidirectional. The push button on the hearing aids were deactivated.

4.4 Measurements

4.4.1 Speech intelligibility test

Speech intelligibility was measured using the Danish HINT (Nielsen & Dau, 2011), with fixed SNR's. Before testing, the subjects received both a verbal and written instruction (Appendix 3) on how to perform the test. The subjects were presented three lists with 25 sentences consisting of five words each. Competing noise was presented three seconds before the HINT-sentence until three seconds after. The noise consisted of Danish 4-talker babble with two male speakers and two female speakers, all of them calibrated to equal sound-levels. HINT-sentences were presented from the front speaker while the noise was presented from the four speakers on the side and back. After each sentence, the subjects were asked to repeat as many of the words as possible before continuing with the next sentence. They were encouraged to guess if not sure. The results were based on how many percentages of the words they repeated correctly.

To make sure the subjects understood the procedure before starting the measuring, they were presented a practice trial at the start of the experiment. The practice list was presented at SNR +4 dB and contained 20 sentences.

After the practice list, the subjects were presented three different HINT-list in different conditions. One easy condition with SNR +7 dB, one medium condition with SNR -2 dB and one hard condition with SNR -7 dB. These SNRs were based of subjects' performance in the study by Ohlenforst et al. (2017), see figure 3. The order in which the conditions were presented were balanced over the subjects to make sure there were no pattern in how the HINT-lists or conditions were presented as this might lead to an unwanted bias. Each condition had a different 4-talker-babble setup where the position of the four talkers were different. For each setup, the talkers were located so that no speaker of the same gender were next to each other. HINT-lists were balanced equally across the subjects and conditions.

4.4.2 Subjective ratings

The subjects were presented a questionnaire (See Appendix 1) on paper four times, one directly after each HINT-list, and one after the practice trial. The questionnaire consisted of the same five questions each time and was inspired by Zekveld et al. (2014). The questions were presented in Danish. All the questions were asked to be rated on an eleven point visual

analogue scale. In the first question (Q1) the subjects were asked “How much effort did you spend when listening to the sentences?” with labels ranging from “no effort” to “very high effort”. Furthermore, the participants were asked in question two (Q2) to rate their performance by answering “How many words do you think you understood correctly?” ranging from “none” to “all”. The third to fifth question were intended to rate the subjects’ motivation. The third question (Q3) was formulated “How often did you give up trying to perceive the sentences?” with labels ranging from “I never gave up listening” to “I always gave up listening”. In the fourth question (Q4), the subjects were asked “How motivated were you to answering the sentences correctly at the last few sentences of the list compared to the first sentences of the list?” ranging from “Less motivated at the end” to “More motivated at the end”. The final question (Q5), the subjects were asked “To what degree did you feel that you could affect the results by putting in more effort?” with labels from “Not much” to “Very much”. An overview of the questions used can be seen in table 1.

Table 1. *Questions used in the questionnaire for rating the perceived effort.*

Q1	How much effort did you spend when listening to the sentences?
Q2	How many words do you think you understood correctly?
Q3	How often did you give up trying to perceive the sentences
Q4	How motivated were you to answering the sentences correctly at the last few sentences of the list compared to the first sentences of the list?
Q5	To what degree did you feel that you could affect the results by putting in more effort?

4.4.3 Pupillometry

Pupillometry was used during the speech test to measure the subject’s PPD. Before starting the practice trial, the subjects were seated down and put on Pupil Labs eye-tracking headset. The headset is connected to a laptop in front of the subject. Pupil Labs headset is worn like a pair of glasses but with cameras about two centimeters from each eye. The cameras are able to slide closer or further away from the subject, and they are also connected to a ball joint which makes it possible for them to rotate. Both cameras have adjustable lenses, making it possible to calibrate the focus to the pupil. When put on, the eye-tracker were adjusted and calibrated

to the subject's eyes. The subjects were instructed to focus on a fixed point in front of them and also told to avoid blinking during the listening task. In the pause between the sentences, the subjects could blink if needed.

The pupil dilation was measured during each sentence from the onset of the noise, to the offset of the noise. This makes a total of 6-7 seconds of pupil dilation data for each sentence. This data was then analyzed through a MATLAB script. Blinks and eye movements were removed by linear interpolation. The pupil dilation data was processed with a smoothing filter based of the moving average 14. A baseline was then calculated as the mean average of the pupil dilation during the two seconds right before the onset of the sentence. The PPD was defined as the maximum dilation in relation to the baseline after the onset of the sentence, i.e. second 3-6 of the pupil data. This was done for all the sentences in each condition, and an average was then created. A similar method was used in the study by Wendt et al. (2017), see figure 7. The only difference is that Wendt et al. (2017) used a baseline calculated of the second before the onset of the sentence, whereas in this study the baseline was calculated of the two seconds before the onset of the sentence.

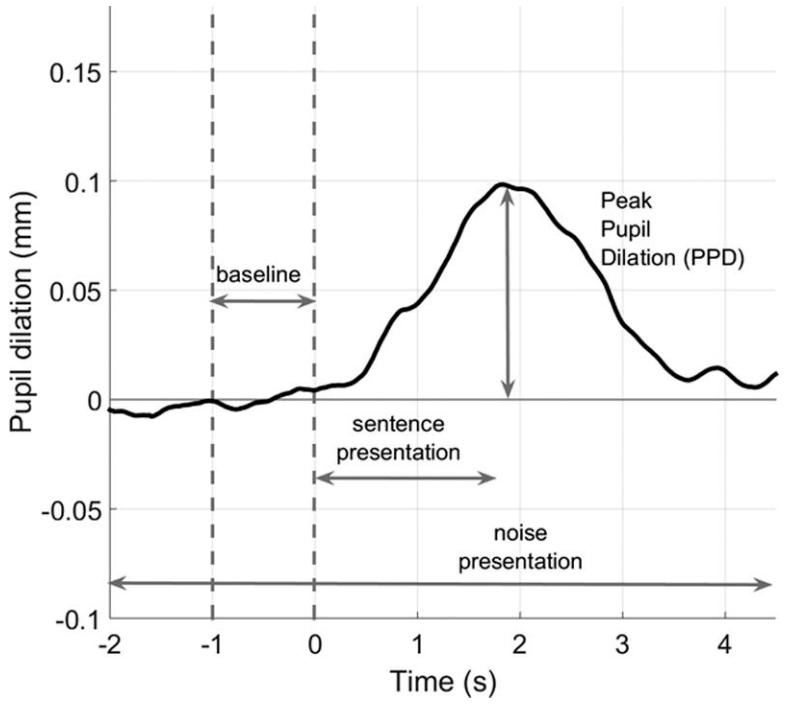


Figure 7. *The figure displays the definition of the peak pupil dilation (PPD) used in the study by Wendt et al. (2017). A similar method was used in this study. The only difference is that Wendt et al. (2017) used a baseline of one second, where in this study it was calculated based of 2 seconds of pupil data.*

4.5 Ethical considerations

The test subjects participated in the study voluntarily. Before accepting the invitation to the study, the test subjects received a written information about the study's purpose and an explanation of how it is performed. They also signed a declaration of consent before the testing. They were offered travel reimbursement. Their personal information and test data was coded and handled anonymously according to the General Data Protection Regulations (EU) (2016/679). There was no economic interest involved in the study. The study was partly funded by Oticon A/S.

4.6 Data processing / Statistical analyses

During testing, the results from the questionnaire and HINT scores were documented on a paper for each participant. After the test was done, the data was transferred to an excel-file on the computer used for the analysis. The pupil data were extracted from the testing computer and also transferred to the analysis computer.

The pupil data from three subjects was excluded due to technical problems and noisy pupil data. All the participants that scored less than 15% on the medium condition were also excluded from the analysis as the PPD of these subjects might not correspond to what we are aiming for in this study. After excluding these subjects, the analysis was based on 14 subjects, out of which 11 had pupil data good enough for the PPD analysis.

The results from each measurement were analyzed using the Shapiro-Wilk test for normal distribution. For the data that was normal distributed, a paired T-test was used for assessing statistical significance. If the data was not normal distributed, the Wilcoxon signed ranks test was instead used. The perceived effort ratings and the pupil data was then compared using the Spearman’s rank correlation coefficient.

5. RESULTS

5.1 Hearing in noise test

Table 2 shows the mean average of words repeated correctly during the HINT in easy, medium and hard listening conditions. Correctly repeated words for the easy condition had a mean of 98,0% (SD=1,7%). Moving on to the medium condition, results for the test group decreased to a mean of 41,4% (SD=18,5%) of the words repeated correctly. Finally, for the hard condition the average for correctly repeated words was 4,0% (SD=7,1%). Wilcoxon

signed ranks test showed a significant difference between the HINT results for the easy condition versus the medium condition (Z= -3,297, p= 0,001), the medium condition versus the hard condition (Z= -3,297, p= 0,001) and also the easy condition versus the hard condition (Z= -3,301, p= 0,001).

Table 2. *The results from the HINT.*

	Mean HINT results	Median HINT results	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Easy condition	98.0%	98.0%	1,7	-	p=0,001	14
Medium condition	41,4%	40,4%	18,5	p=0,001	-	14
Hard condition	4.0%	1.0%	7,1	p=0,001	p=0,001	14

5.2 Subjective ratings

Table 3 shows the mean average for question one (Q1) on each condition, where the subjects were asked to rate their perceived effort. On the easy condition, the test subjects' mean average for perceived effort was 5,0 (SD= 2,1), on the medium condition the perceived effort increased to 8,6 (SD= 1,2) and on the hard condition the perceived effort slightly increased to 9,1 (SD= 0,8). Wilcoxon signed ranks test shows a significant difference between the perceived effort for the easy condition versus the medium condition (Z= -3,041, p= 0,002). Comparing the perceived effort for the medium condition and the hard condition shows no significant difference (Z= -1,578, p= 0,12). Finally, comparing the rated effort for the easy condition and the hard condition shows a significant difference (Z= -3,181, p= 0,001).

Table 3. *The results from question one (Q1) from the questionnaire, "How much effort did you spend when listening to the sentences?".*

	Mean	Median	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Q1 Easy condition	5,0	5,3	2,1	-	p=0,002	14
Q1 Medium condition	8,6	8,7	1,2	p=0,002	-	14
Q1 Hard condition	9,1	9,0	0,8	p=0,001	p=0,12	14

In question two (Q2) the subjects were asked for each condition to rate how many words they thought they understood correctly, ranging from none to all of them on an eleven point scale. Table 4 shows the results for question two in each condition. For the easy condition, the mean

average was 9,4 (SD= 0,6), which translates to 94% of the words understood correctly. For the medium condition the mean average dropped to 3,6 (SD= 2,7) which translates to 36%. The mean average continues to decrease in the hard condition with a rated mean average of 1,1 (SD= 1,1), translated to 11% of the words understood correctly. Wilcoxon signed ranks test shows there is a significant difference between the estimated words understood correctly for the easy versus medium condition (Z= -3,298, p= 0,001), the medium versus hard condition (Z= -2,902, p= 0,004) as well as the easy condition versus the hard condition (Z= -3,302, p= 0,001).

Table 4. *The results from question two (Q2) from the questionnaire, "How many words do you think you understood correctly?".*

	Mean	Median	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Q2 Easy condition	9,4	9,6	0,6	-	p=0,001	14
Q2 Medium condition	3,6	2,3	2,7	p=0,001	-	14
Q2 Hard condition	1,1	1,0	1,1	p=0,001	p=0,004	14

Table 5 displays the mean average results of question three (Q3), asking how many times the subjects gave up trying to perceive a sentence. For the easy condition, the mean average was 0,3 (SD= 0,4), indicating that the group almost never gave up. In the medium condition the mean average increased to 6,1 (SD= 2,9) indicating the subjects gave up trying to perceive the sentences more than half of the times. For the hard condition the mean average increased further to 8,2 (SD= 2,7) indicating the subjects gave up trying to perceive the sentences close to all of the times. Wilcoxon signed ranks test reveals a significant difference between the amount of times the subjects gave up trying to perceive the sentences during the easy condition versus the medium condition (Z= -3,300, p= 0,001). It also shows there is a significant difference between the medium and hard condition (Z= -2,132, p= 0,033) as well as the easy versus the hard condition (Z= -3,238, p= 0,001).

Table 5. *The results from question three (Q3) from the questionnaire, "How often did you give up trying to perceive the sentences?".*

	Mean	Median	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Q3 Easy condition	0,3	0,1	0,4	-	p=0,001	14
Q3 Medium condition	6,1	7,3	2,9	p=0,001	-	14
Q3 Hard condition	8,2	9,0	2,7	p=0,001	p=0,033	14

Table 6 shows the mean average ratings for the subjects on question four (Q4) for each condition. The subjects were asked how motivated they were to answer correct on the last sentences of the list compared to the first sentences of the list. For the easy condition the mean average of the test subjects were 5,8 (SD= 1,7), for the medium condition 6,0 (SD= 2,0) and for the hard condition 6,1 (SD= 2,1). Wilcoxon signed ranks test showed that there is no significant difference between the easy condition versus the medium condition (Z= -0,730, p= 0,465), the medium condition versus the hard condition (Z= -0,271, p= 0,786) or the easy condition and the hard condition (Z= -1,166, p= 0,244).

Table 6. *The results from question four (Q4) from the questionnaire, "How motivated were you to answering the sentences correctly at the last few sentences of the list compared to the first sentences of the list?".*

	Mean	Median	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Q4 Easy condition	5,8	5,0	1,7	-	p=0,465	14
Q4 Medium condition	6,0	5,0	2,0	p=0,465	-	14
Q4 Hard condition	6,1	5,0	2,1	p=0,244	p=0,786	14

Table 7 shows the average for the subjects' rating of question five (Q5) in each condition. The subjects were asked how much they felt like they could affect the results by putting in more effort. For the easy condition the mean average was 3,6 (SD= 2,7), for the medium condition 3,2 (SD= 2,3) and for the hard condition 2,1 (SD= 2,4). Wilcoxon signed ranks test showed that there is no significant difference between the easy condition versus the medium condition (Z= -1,021, p= 0,307), neither the medium condition nor the hard condition (Z= -1,275, p= 0,202). However it did show a significant difference between the easy condition versus the hard condition (Z= -2,091, p= 0,036).

Table 7. *The results from question five (Q5) from the questionnaire, "To what degree did you feel that you could affect the results by putting in more effort?".*

	Mean	Median	Standard Deviation	Sig. compared to easy condition	Sig. compared to medium condition	Total n
Q5 Easy condition	3,6	3,0	2,7	-	p=0,307	14
Q5 Medium condition	3,2	4,0	2,3	p=0,307	-	14
Q5 Hard condition	2,1	1,0	2,4	p=0,036	p=0,202	14

5.3 Pupillometry

The PPD for the different conditions are shown in Figure 8. The mean PPD for the easy condition were 1,35 AU (SD= 0,8). For the medium condition, the PPD were 2,25 AU (SD= 0,86). In the hard condition, the PPD were 1,29 AU (SD= 0,5). A paired T-test showed a statistical significant difference between the easy and medium condition (T= -2,638, p= 0.03), and also a statistical significant difference was achieved for the medium versus the hard condition (T= 5,365, p= <0,001). However no significant difference between the easy and hard condition was measured (T= 0,405, p= 0,67).

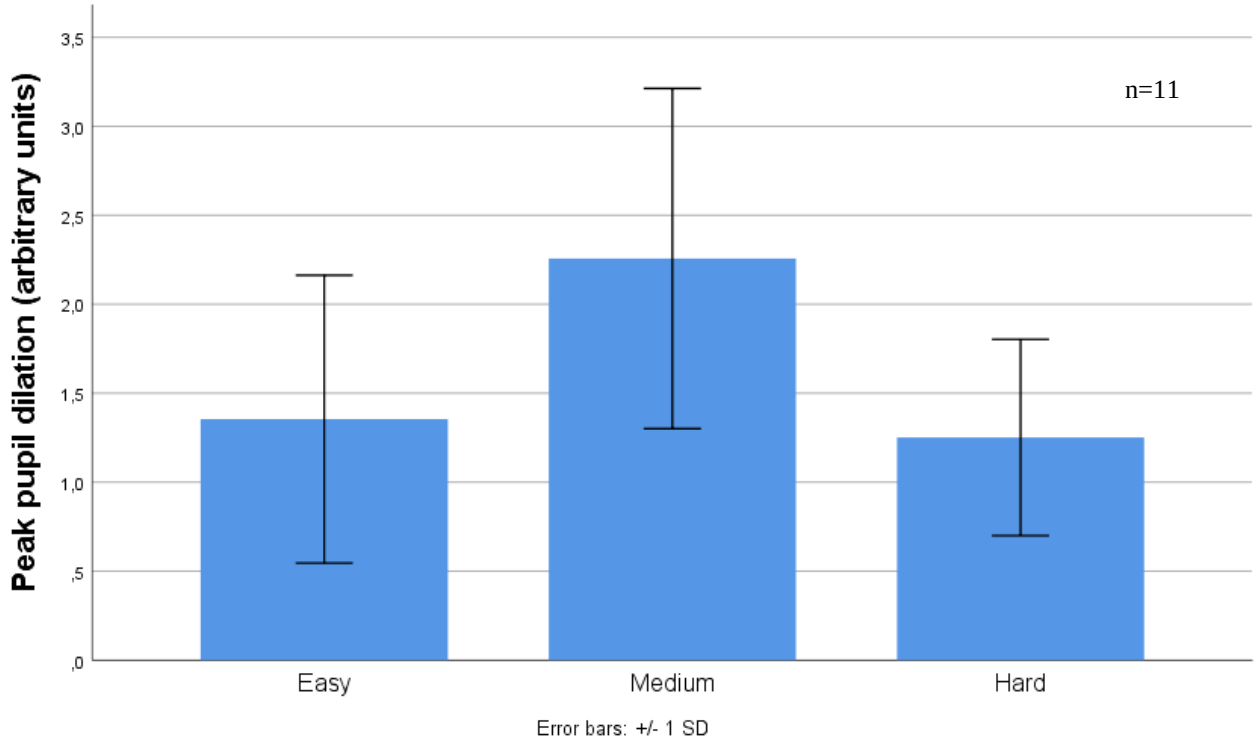


Figure 8. *The figure displays the peak pupil dilation for the easy, medium and hard condition. n=11*

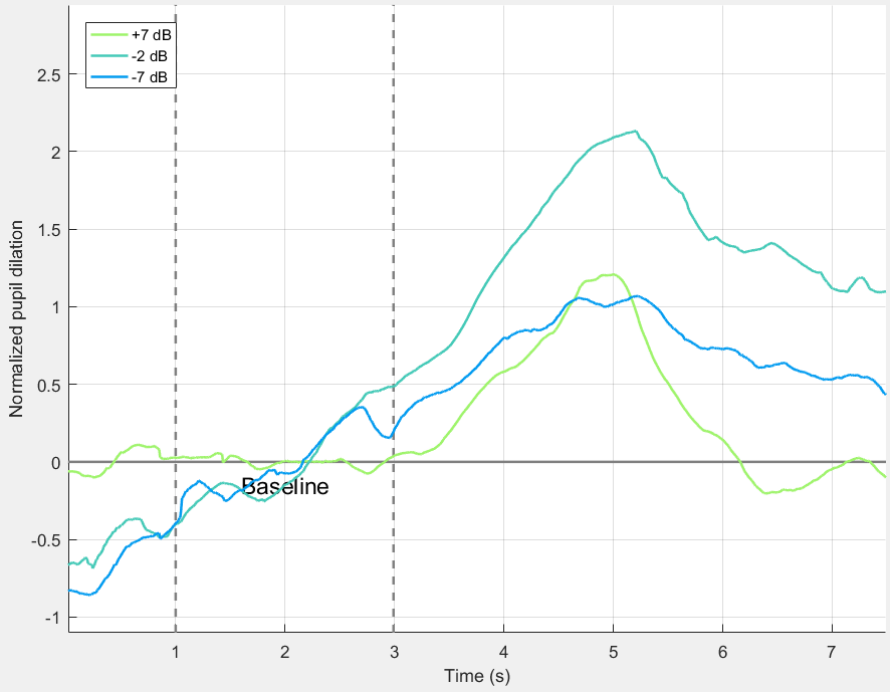


Figure 9. *The pupil dilation curve for each condition. +7 dB=easy condition, -2 dB=medium condition and -7 dB= hard condition. n=11*

5.4 Correlation Analysis

When comparing the PPD vs perceived effort as rated in question 1 by Spearman's rank correlation coefficient, no significant correlation was found in the easy condition (r= 0,21, p=0,47). For the medium condition, no significant correlation was found (r= -0,46, p=0,073). For the hard condition there were no significant difference either (r= 0,189, p=0,5).

Across all the conditions for PPD and Question 1, no significant correlation was found using Spearman's rank correlation coefficient (p=1). The values used in the correlation analysis are displayed in Figure 10.

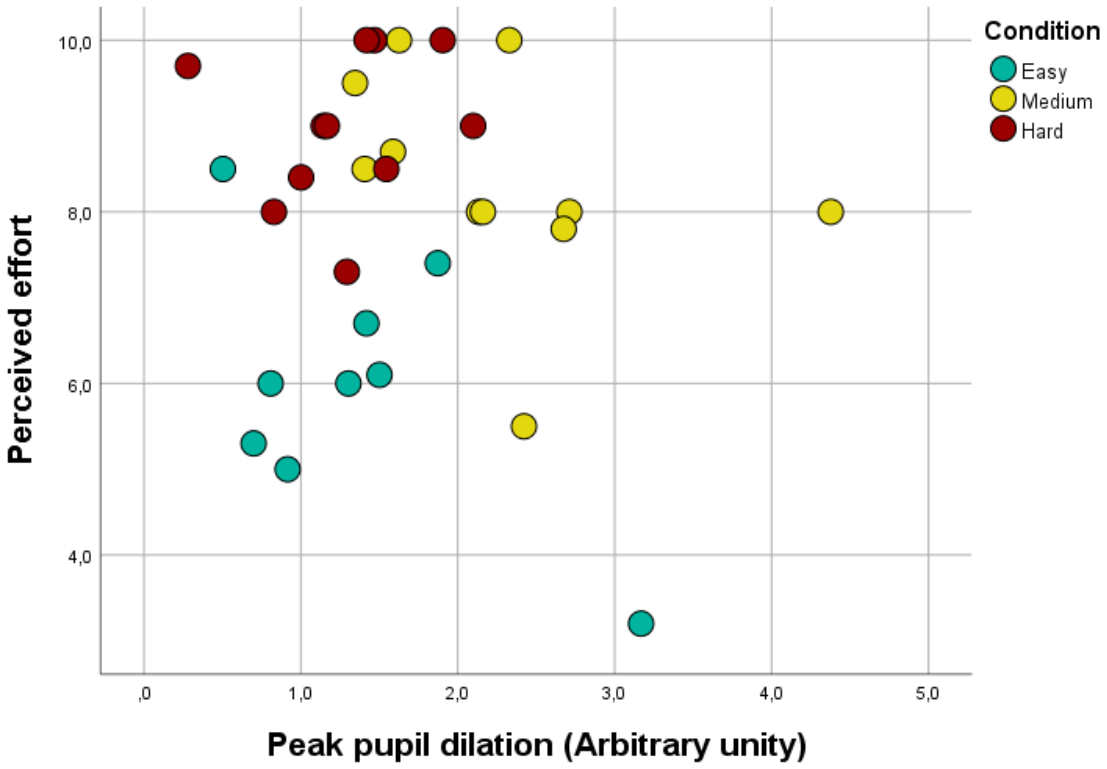


Figure 10. *Perceived effort and peak pupil dilation across the different conditions. n=31.*

6. DISCUSSION

6.1 Sustainable development discussion

According to United Nations sustainable development is defined as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs”. Furthermore, United Nations mentions three aspects especially important to a person's and a society’s well-being: economic growth, social inclusion and environmental protection (United Nations, 2018).

Hearing impairment is shown to associate with difficulties in communication and reduced quality of life. It can also be the cause of a reduced function of everyday-activities such as shopping, cooking dinner and talking on the phone. There is also a possibility that hearing loss has an effect on the hearing impaired listener’s economy as it reduces the possibility to work and make money (Dalton, 2003). This would, as we interpret it, be an obstacle for the societies and the hearings impaired listener’s well-being and therefore not a sustainable development.

By using hearing aids it has been shown that these negative results of quality of life is reversible (Mulrow et al., 1990). Therefor it is of great importance to constantly develop and evaluate hearing aids objectively as well as subjectively to reverse the down going results of quality of life in hearing impaired listeners. By evaluating methods of rating hearing aid-users listening effort and cognitive load we get a better picture of how much the hearing aids are of use.

6.2 Method discussion

Previously made studies, for example Zekveld et al. (2014), have used similar methods to research the motivations impact on pupillometry. To use a similar method allows us to compare the data we acquire with previous studies.

6.2.1 Hearing in noise-test

Danish HINT is a well-known and used method for measuring speech intelligibility. By using a 4-talker babble noise we acquire a sound environment closer to the real-life scenarios for hearing aid users, as compared to a stationary noise. But as an effect of the 4 talker babbles fluctuating sound level the SNR is constantly shifting during the measurement, which makes it harder to get the calibration precise. By measuring the sound level over a longer period of time you can control the overall SNR of the test. As the intelligibility level of the different HINT-lists are slightly different, we can minimize the influence of this difference by using the HINT-lists equal amount of times. The HINT-lists were also balanced equally across the three different conditions.

6.2.2 Hearing aids

For this study, the participants wore hearing aids fitted to their hearing loss. The noise reduction was turned off and the microphones were set to omni. These settings were selected to reduce the influence of the hearing aid on the test. The purpose of the hearing aids were only to provide audibility during the test. As the participants usually wear hearing aids with the noise reduction turned on and with directional microphones, the settings used in this study may not replicate the everyday listening situations for the participant. However to avoid branding the results of this test with the noise reduction and microphones used in Oticon Opn1, it is appropriate to deactivate these functions. We believe that the hearing aid settings used in this study is more relevant for further research. The VAC+ amplification method was selected because the participants had previously worn hearing aids with this method. We

chose to use this method instead of a non-brand specific method because we did not want to change the amplification of the hearing aids too much from what the participant was used to.

6.2.3 Signal-to-noise ratio

The SNRs for each condition were selected partly based of previous research and partly on experience from Eriksholm research centre. For the easy and the hard condition, we strived for a low PPD, whereas in the medium condition we strived for a high PPD. SNR +7 for the easy condition and -7 for the hard should give us low PPD, and SNR -2 for the medium condition should give us a high PPD according to Ohlenforst et al. (2017). We adjusted these slightly because of the fact that we are using a different noise than what is used in Ohlenforst et al. (2017), which is slightly harder than stationary noise. Previous studies indicate that the highest PPD is reached at speech intelligibility levels around 50%. For the medium condition of our study, the speech intelligibility was 30,3% when including all the 20 participants. This shows that we might have used slightly too low SNR for this condition. The fact that the PPD decreases in hard conditions, while the difficulty of the listening task increases, is what this study is trying to investigate. Therefore, excluding the six subjects that scored 15% or less in the medium condition can make up for the fact that the medium condition was slightly too hard.

6.2.4 Questionnaire

We used a total of five questions for evaluating listening effort and motivation (Table 1). Q1-Q3 have been used before in similar studies, for example Zekveld et al. (2014), and has been shown to measure what is asked for in a good way. The results obtained from these questions were overall what we expected. By using these three questions we get results that are easy to compare with other studies and has a high reliability. Q4 and Q5 were composed for this study. Q4 intended to see how the subjects' motivation changed during the listening task for each condition. The subjects' ratings did not follow any pattern and we suspect that the subjects did not understand the question completely. Q5 intended to see how much the subjects felt they could affect the results by putting in effort. We could suspect a pattern that the harder the conditions got, the less subjects felt they could affect the results by putting in effort. But there were no significant results. A theory is that it would show bigger differences if the question was asked in a slightly different way, asking how much they could affect the results by putting in effort, instead of putting in more effort. Our believes is that using

"putting in more effort" makes the subjects rate lower scores no matter the difficulty because they feel like they did everything they could.

6.2.5 Eye-tracker

In this study we used Pupil Labs eye-tracking headset, with which we got good recordings of the PPD. We believe that one of the reasons to use this eye-tracker is that it is head mounted, which makes it less sensitive to head movements during the measurement. We also believe that this can mirror in the subjects' behaviour, making them less aware of keeping the head still and focusing more on the listening task. Even though it is preferred to keep the head still and be fully focused on the listening task. One more thing to highlight is that Pupil Labs eye-tracking headset is in a lower price class compared to many other cameras used to measure the PPD. This makes it available for those with a lower budget, while still producing good recordings.

As far as we know, light is the most common thing that influences the pupil dilation the most. In this study, the lights were set to steady level for all of the participants during the testing. Other things that could influence the pupil dilation was not taken into account during our testing. These could for example be caffeine or other drugs. Age of the test subject could also be a factor that could affect the pupil dilation. These could have been taken into account and monitored for further reliability of our results.

6.3 Results discussion

The purpose of this study was to investigate if the motivation of the test subject influences the PPD in different listening conditions. The goal is to assess effort involved in speech recognition with both pupillometry (as an indicator of listening effort) and rating scales (as an indicator of subjective/perceived effort). To achieve a good picture of this we formulated three research questions that we wanted to answer.

The first question is *How does perceived effort (as measured by the questionnaire) and objective listening effort (as measured by the PPD) change with increasing task demands?* When looking at the results (See Table 3) for perceived effort, we can see an indication that perceived effort gets higher the harder the task demands get. For the easy condition, the subjects rated a mean average of 5,0 on an eleven-point scale. On the medium condition the mean average raised to 8,6 and furthermore on the hard condition it raised to 9,1. When doing

the Wilcoxon signed rank test we could see a significant difference between the perceived effort for the easy condition versus both the medium and hard condition. However, we did not see a significant difference between the perceived effort for the medium condition and the hard condition. Previous studies shows that there is a linear increase of perceived effort when task demands get higher (Zekveld et al., 2010, 2014; Rudner et al., 2012). See also Figure 2. The reason that we can not see any significant difference between the perceived effort for the medium and hard condition might be due to the medium condition being slightly too hard for the subjects. This could lead to a ceiling effect that caused the perceived effort in the medium condition to already be at very high levels, which prevents the perceived effort to increase enough for a significant difference from the medium to hard condition.

When looking at objective listening effort as measured by the PPD we could see a significant difference between the easy condition and the medium condition. With the PPD increasing from 1,35 AU to 2,25 AU indicating that the subjects did have to put in more effort when the task demands got higher. However, the PPD did decrease when comparing the medium condition to the hard condition, where subjects PPD was 1,29 AU on the hard condition. So in total we did see a significant difference in the PPD between the easy and medium condition, the medium and hard condition, but not between the easy and hard condition. The PPD did increase from the easy to medium condition, indicating that the listening was more effortful, but decreased to the hard condition, indicating that the listening was less effortful. If looking at the PPD for hearing impaired listeners in figure 3 we can see a similar curve, where the PPD increases with decreasing SNR up to a certain turning point, then decreases with decreasing SNR.

Taken together, going from medium to hard condition, there was a discrepancy between the perceived effort and the objective effort, see figure 11. The perceived effort indicates that the hard condition is at least as effortful as the medium condition, while the objective effort indicates that the listener spend less effort in the hard condition compared to the medium condition. That is, in the objective effort it seems as if the listener ‘give up’ in line with what Pichora-Fuller et al. (2016) predicts.

How can this discrepancy be explained? One way to interpret the data is that listeners are not so good at monitoring their own effort. So when they are asked, as in the perceived effort rating, they infer that the hard condition must be effortful since they cannot hear what is said

(their objective speech intelligibility was low, Table 1, as well as the subjective intelligibility, Table 3). But the objective effort tells a different story, where the physiological pupil response reveals that it is plausible that the listener has ‘given up’, and thus spend less effort in the hardest condition. So the perceived effort rating seems to indicate the listener’s expectation on the effort in the hardest condition rather than the factual effort spent in the condition.

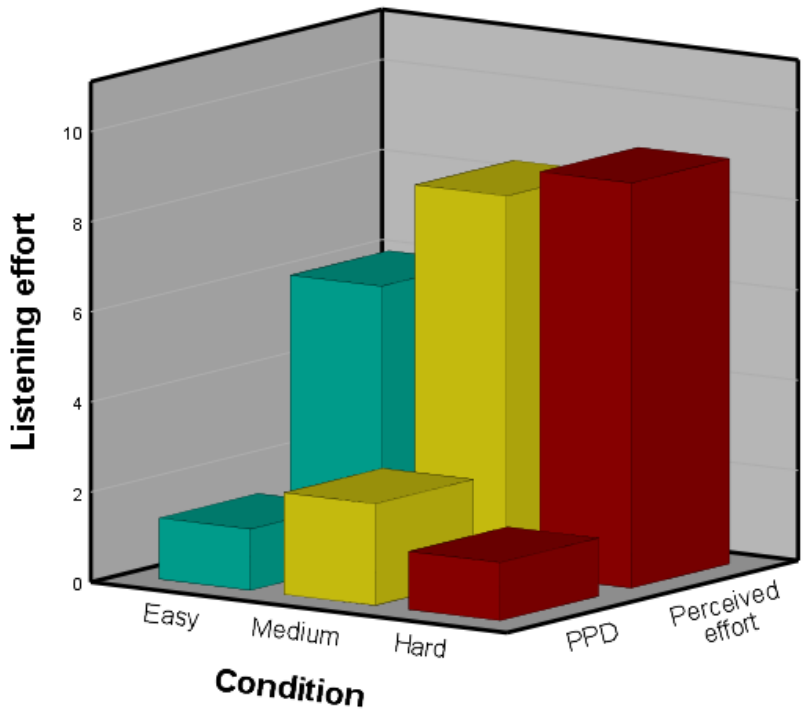


Figure 11. *Peak pupil dilation (PPD) and perceived effort in the different conditions. n=14*

Examining the relation between the objective listening effort as measured by the PPD and the perceived listening effort as measured by the questionnaire, no significant correlation was found within or across all three conditions. When looking at the PPD and perceived effort in the easy versus the hard condition, both the PPD and the perceived effort increases. However, it is for the medium and hard condition where these two measures of listening effort differs. For the PPD, the effort decreases from the medium to hard condition. For the perceived effort, the results from this study and previous studies such as Zekveld et al. (2010) are aiming towards an increase in perceived effort. This increase for the perceived effort are not significant, but we can see a trend towards significance (p=0,12). These findings when comparing PPD and perceived effort are in line with previous research that found no

connection between the physiological indications of effort and the perceived effort (Gosselin et al., 2011; Mackersie et al., 2011; Koelewijn et al., 2012).

Our third research question was *How does motivation change with increasing task demands?* When looking at the results from question three in Table 5 we can see that subjects reported to give up more often when task demands got higher. There was a significant difference between the results for the easy condition versus the medium condition and the hard condition, as well as between the medium condition and the hard condition. This indicates that the subjects were not equally motivated to perceive the sentences during higher task demands. It is speculated that the subjects do not feel that it is worth putting in effort during the harder task demands, as their rewards would be small. In this case this would save them from experiencing more mental fatigue trying to perceive sentences not possible for them to perceive. According to Pichora-Fuller et al. (2016) listeners may stop putting in effort when exposed to increased listening effort, in order to avoid mental fatigue. The results from this study indicates this as well, as the motivation for the subjects drop the harder the task demands get.

These results are indicating that the eye tracker and the questionnaire are not measuring the same aspect of listening effort. In the hard condition, where the test subjects reported the lowest motivation, the PPD and perceived effort differed the most. As we believe the PPD corresponds better to the actual resources spent in these hard conditions, motivation of the test subject could be a big influence of the PPD. The perceived effort might be more related to how the subjects rated their effort based of how much effort they think they should be spending. As the motivation was low, and therefore the participation in the task was also low, the actual resources spent during the task should be fairly low. This can be reflected in the figure by Pichora-Fuller et al. (2016), see Figure 1. When the task demands gets too high, the motivation drops but the subject still rates their perceived effort as very high. The listening effort as reflected by the PPD then drops because the participation in the task is low and therefore the subject doesn't spend any resources. Further, as the motivation significantly dropped from the easy to the medium condition, this contradicts our statement that the motivation influences the PPD. We believe an explanation for this could be that the motivation in the medium condition were not low enough for the test subject to give up completely and stop putting in effort.

6.3.1 Research suggestions

We suggest future studies that measures the mental fatigue in medium and hard speech intelligibility levels and compares these. Which condition produces the highest mental fatigue? The medium condition where PPD is higher, or in the hard condition where subjects rate their perceived effort higher? It would be interesting to see how the state where the subjects rate their perceived effort as high relates to mental fatigue compared to when the subjects' objective effort is high.

6.3.2 Clinical application

In the future pupillometry might be a good way to further develop the evaluation of hearing aids clinically. One common way to subjectively evaluate the use of hearing aids is subjective ratings before and after fitting. You can also measure the use of hearing aids objectively by measuring the performance in speech intelligibility tasks with and without hearing aids. By adding pupillometry to the evaluation of hearing aids there would be an objective measure showing the eventual decrease of physiological effort. This would complement the evaluation adding further reliability.

7. CONCLUSION

This study shows that when task demands get higher, the subjects' motivation decreased. There was a difference between the perceived effort (as measured by the questionnaire) and the objective effort (as measured by the pupillometry). The perceived effort indicate that the hard condition is at least as effortful as the medium condition, while the objective effort indicate that the subjects spend less effort in the hard condition compared to the medium condition.

REFERENCES

Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *The gerontologist*, *43*(5), 661-668.

Downs, D. W. (1982). Effects of hearing aid use on speech discrimination and listening effort. *Journal of Speech and Hearing Disorders*, *47*(2), 189-193.

Gosselin, P. A., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, *54*(3), 944-958.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291-300.

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Audiology*, *36*(3), 155-164.

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work: Desempeño laboral: Comparación de empleados con audición normal o alterada usando el Listado Amsterdam para Audición y Trabajo. *International Journal of Audiology*, *45*(9), 503-512.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects Desempeño cognitivo y percepción del esfuerzo en tareas de procesamiento del lenguaje: Efectos de las diferentes condiciones de fondo en sujetos normales e hipoacúsicos. *International Journal of Audiology*, *44*(3), 131-143.

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and hearing*, *37*, 77S-84S.

Mackersie, C. L., & Cones, H. (2011). Subjective and psychophysiological indexes of listening effort in a competing-talker task. *Journal of the American Academy of Audiology*, *22*(2), 113-122.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 22-33.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*(2), 193-203.

Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., ... & DeNino, L. A. (1990). Quality-of-life changes and hearing impairment: a randomized trial. *Annals of Internal Medicine*, *113*(3), 188-194.

Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *International journal of audiology*, *50*(3), 202-208.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085-1099.

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., ... & Kramer, S. E. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hearing research*, *351*, 68-79.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... & Naylor, G. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S.

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*(8), 577-589.

Schepker, H., Haeder, K., Rennies, J., & Holube, I. (2016). Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners. *International journal of audiology*, *55*(12), 738-747.

United Nations (2018). *United Nations Sustainable Development*. Retrieved 29 March, 2018, from http://www.un.org/sustainabledevelopment/development-agenda/

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and hearing*, *38*(6), 690-700.

Wu, Y. H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort. *Ear and hearing*, *37*(6), 660-670.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and hearing*, *31*(4), 480-490.

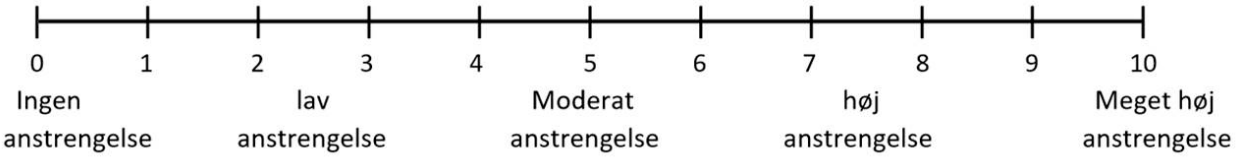
Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277-284.

APPENDIX 1 - Questionnaire

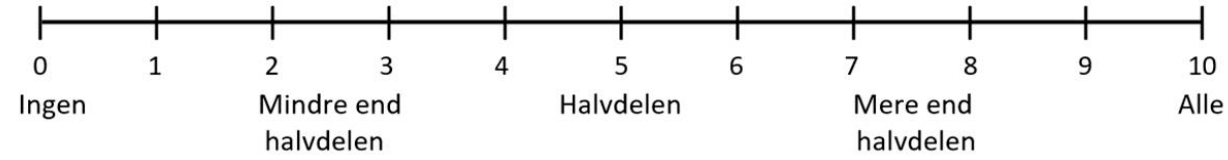
Spørgeskema nr:_____

Test subject:_____
Date:_____

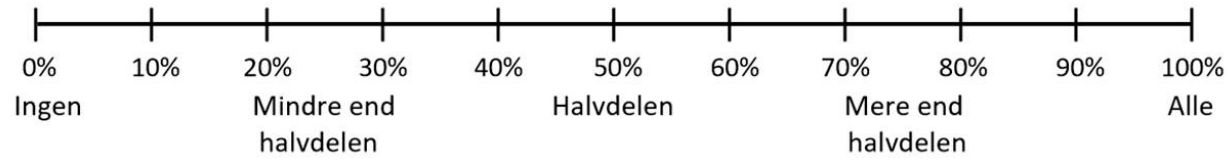
Hvor meget anstrengte du dig for at høre sætningerne?



Hvor mange af ordene tror du, at du forstod korrekt?



Hvor ofte måtte du opgive at forstå sætningen?



Hvor motiveret var du for at svare rigtigt på de sidste par sætninger fra listen sammenlignet med de første sætninger fra listen?

Mindre motiveret mod slutningen

Lige motiveret hele vejen igennem

Mere motiveret mod slutningen

I hvilken grad kunne du påvirke resultaterne ved at anstrenge dig mere?

Slet ikke

Lidt

I nogen grad

Meget

Rigtig meget

APPENDIX 2 - Individual results from the questionnaire and PPD

Easy condition

Test subject nr	SNR	HINT result	Q1	Q2	Q3	Q4	Q5	PPD
504	7	55,2	8,5	4,5	7,5	5,0	2,5	1,61
1049	7	94,5	6,1	9,1	1,0	8,0	5,1	1,50
1076	7	74,4	5,1	7,6	1,2	5,1	5,3	0,82
1178	7	96,8	2,4	9,4	0,6	5,0	0,8	-
1217	7	96,0	6,7	9,7	1,2	4,6	7,1	1,42
1341	7	99,2	2,0	10,0	0,0	10,0	0,0	-
1463	7	82,4	5,3	7,5	2,5	5,5	7,4	1,03
1504	7	97,5	5,0	9,0	0,0	5,0	3,0	3,38
1544	7	96,8	5,0	10,0	0,0	5,0	5,0	-
1589	7	99,2	8,0	10,0	0,1	5,0	5,0	1,98
1601	7	97,6	3,2	9,8	0,2	5,0	5,0	3,17
1624	7	99,2	7,4	9,9	0,2	5,0	7,0	1,87
1630	7	99,2	8,5	9,7	0,0	8,5	8,5	0,50
1632	7	91,2	8,6	8,7	0,7	7,9	3,4	1,69
1637	7	100,0	6,0	9,0	0,0	5,0	3,0	1,30
1639	7	96,8	5,3	9,2	0,5	5,4	2,4	0,70
1641	7	98,4	6,0	8,0	1,0	5,1	1,0	0,81
1647	7	38,4	10	2	8	5	1	-
1654	7	100,0	5,0	10,0	0,0	5,0	3,0	0,92
1658	7	100,0	0,9	9,0	0,0	5,0	0,0	3,19
Average:		90,6	5,8	8,6	1,2	5,8	3,8	1,6
Standard dev.:		16,5	2,4	2,0	2,3	1,5	2,6	0,9

Medium condition

Test subject nr	SNR	HINT result	Q1	Q2	Q3	Q4	Q5	PPD
504	-2	4,8	8,5	0,6	9,5	2,6	1,5	1,97
1049	-2	36,0	10,0	2,1	7,8	9,1	5,1	1,63
1076	-2	7,2	7,7	0,7	0,3	5,1	0,3	0,92
1178	-2	44,0	9,0	1,8	8,4	5,0	0,6	-
1217	-2	33,6	5,5	7,4	6,8	4,8	5,3	2,42
1341	-2	67,2	8,0	7,0	2,0	10,0	0,0	2,13
1463	-2	2,4	9,5	0,5	9,5	9,4	9,3	-0,11
1504	-2	48,0	10,0	2,0	7,0	5,0	3,0	2,44
1544	-2	24,0	10,0	1,0	5,0	5,0	1,0	2,33
1589	-2	9,6	10,0	2,0	7,0	5,0	1,1	0,70
1601	-2	39,2	8,0	2,5	7,5	5,0	5,0	2,71
1624	-2	66,4	8,0	7,0	2,0	5,0	5,0	2,16
1630	-2	19,2	8,5	1,5	8,8	8,5	5,5	1,41
1632	-2	1,6	9,6	0,5	9,4	4,6	9,6	1,45
1637	-2	63,2	8,0	7,0	2,0	5,0	2,0	4,38
1639	-2	15,2	8,7	1,2	8,8	8,6	0,6	1,59
1641	-2	18,4	9,5	2,5	8,3	3,6	0,5	1,34
1647	-2	0,8	10	2	9,5	5	1	0,85
1654	-2	64,2	7,8	7,0	2,0	5,0	6,0	2,67
1658	-2	41,6	9,0	1,0	9,5	5,0	5,0	5,84
Average:		30,3	8,8	2,9	6,6	5,8	3,4	2,0
Standard dev.:		23,2	1,1	2,6	3,1	2,1	3,0	1,3

Hard condition

Test subject nr	SNR	HINT result	Q1	Q2	Q3	Q4	Q5	PPD
504	-7	0,0	8,5	0,5	9,5	5,0	0,7	1,78
1049	-7	0,0	8,0	0,2	10,0	10,0	0,0	0,83
1076	-7	0,0	7,3	0,1	9,8	5,1	0,0	0,19
1178	-7	7,2	9,3	1,0	0,5	5,0	0,0	-
1217	-7	0,8	7,3	3,7	6,7	4,8	6,5	1,29
1341	-7	6,4	10,0	1,0	9,0	10,0	0,0	1,47
1463	-7	0,0	9,5	0,5	9,4	9,4	9,4	-0,02
1504	-7	0,8	9,0	1,0	9,0	5,0	3,0	-
1544	-7	0,8	10,0	1,0	5,0	7,0	1,0	1,42
1589	-7	0,0	10,0	0,1	10,0	5,1	0,0	0,59
1601	-7	1,6	10,0	0,0	10,0	5,0	0,0	1,90
1624	-7	27,2	9,0	1,6	8,5	5,0	6,5	1,15
1630	-7	0,8	8,5	0,1	9,8	8,5	0,5	1,54
1632	-7	0,0	7,6	2,5	7,4	7,6	3,6	0,84
1637	-7	4,8	9,0	1,0	10,0	3,0	1,0	2,10
1639	-7	0,0	8,4	2,5	8,4	7,5	2,6	1,00
1641	-7	0,0	9,7	0,3	9,8	5,0	0,4	0,28
1647	-7	0	10	0	10	5	0	0,44
1654	-7	4,8	9,0	2,0	8,0	5,0	3,0	1,17
1658	-7	1,2	10,0	0,0	10,0	5,0	5,0	3,17
Average:		2,8	9,0	1,0	8,5	6,2	2,2	1,2
Standard dev.:		6,2	0,9	1,0	2,3	2,0	2,8	0,8

APPENDIX 3 – Written test instructions

Du vil høre sætninger i baggrundsstøj. Støjen består af flere talere. Støjen starter 3 sekunder før hver sætning og ender 3 sekunder efter hver sætning. Sætningerne er indtalt af en mand.

I pausen efter støjen skal du forsøge at gentage så mange ord som muligt. Hvis du ikke er sikker, er det okay at gætte dig frem.

Det er din opgave at fokusere på et aftalt punkt, mens du lytter til støjen og sætningen. Det er vigtigt, at du undgår at blinke med øjnene eller bevæger hovedet mens du lytter. Hvis du har brug for at blinke eller bevæge dig lidt, er det muligt i pauserne efter støjen.

Efter hver testrunde skal du udfylde et spørgeskema, hvor du bedømmer forskellige aspekter af anstrengelsen i forbindelse med at lytte. Papirerne er navngivet med numrene 1, 2 og 3. Papiret med nummer 1 er for den første testrunde, nummer 2 er for den næste testrunde og nummer 3 for den sidste runde.

Hvert papir består af 5 spørgsmål. Du skal markere dit svar med et X på en skala for hvert spørgsmål. Vær opmærksom på, at skalaerne varierer lidt fra spørgsmål til spørgsmål. Når du er færdig med at udfylde papiret bedes du sige *"Jeg er klar"*. Så vil næste testrunde starte. Hver testrunde har forskellig sværhedsgrad, der varierer fra meget svær til let.

Testen starter med en træningsrunde – så du bliver fortrolig med testen. Efter øvelseslisten starter den første testrunde. Hver liste tager ca. 6 minutter.

Læs venligst det første spørgeskema på bordet igennem før testen begynder.

Spørg endelig, hvis du har spørgsmål!