Reading Behind the Lines: The Factors Affecting the Text Reception Threshold in Hearing Aid Users.

Adriana A Zekveld, Marieke Pronk, Henrik Danielsson and Jerker Rönnberg

Reading behind the lines:

The factors affecting the text reception threshold in hearing-aid users

Adriana A. Zekveld[1,2,3]

Marieke Pronk[3], Henrik Danielsson[1,2] and Jerker Rönnberg[1,2]

[1]Dept. of Behavioural Sciences and Learning, Linköping University, Sweden

[2]Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Sweden

[3]Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health research institute, VU University Medical Center, the Netherlands

Correspondence:

Adriana A. Zekveld, PhD,

Section Ear & Hearing, Dept. of Otolaryngology-Head and Neck Surgery, Amsterdam Public Health research institute,

VU University medical center

P.O. Box 7057,

1007 MB Amsterdam,

The Netherlands.

aa.zekveld@vumc.nl

Running title: reading behind the lines

Keywords: Text reception threshold, Speech perception in noise, cognitive abilities

Abstract

Purpose: The visual Text Reception Threshold (TRT) test has been designed to assess modality-general factors relevant for speech perception in noise. In the last decade, the test has been adopted in audiology labs worldwide. The first aim of the present study was to examine which factors best predict inter-individual differences in the TRT. Secondly, we aimed to assess the relationships between the TRT and speech reception thresholds (SRTs) estimated in various conditions.

Method: First, we reviewed studies reporting relationships between the TRT and auditory and/or cognitive factors and formulated specific hypotheses regarding the TRT predictors. These hypotheses were tested using a prediction model applied to a rich dataset of 180 hearing-aid users. In separate association models, we tested the relationships between the TRT and various SRTs and subjective hearing difficulties, while taking into account potential confounding variables.

Results: The results of the prediction model indicate that the TRT is predicted by the ability to fill in missing words in incomplete sentences, by lexical access speed, and by working memory capacity. Furthermore, in line with previous studies, a moderate association between higher age, poorer pure-tone hearing acuity, and poorer TRTs was observed. Better TRTs were associated with better SRTs for the correct perception of 50% of Hagerman matrix sentences in 4-talker babble, as well as with better subjective ratings of speech perception. Age and pure-tone hearing thresholds significantly confounded these associations. The associations of the TRT with SRTs estimated in other conditions and with subjective qualities of hearing were not statistically significant when adjusting for age and PTA.

Conclusions: We conclude that the abilities tapped into by the TRT test include processes relevant for speeded lexical decision making when completing partly masked sentences, and that these processes require working memory capacity. Furthermore, the TRT is associated

with the SRT of hearing-aid users as estimated in a challenging condition that includes informational masking and with experienced difficulties with speech perception in daily-life conditions. The current results underline the value of using the TRT test in studies involving speech perception and aid in the interpretation of findings acquired using the test.

Speech comprehension in challenging conditions involves both auditory-specific functions (e.g., hearing acuity/audibility, the encoding of a signal in the auditory periphery; George, et al. 2007; Plomp, 2002) and modality-aspecific cognitive functions (e.g., linguistic ability, verbal working memory; Humes, 2007; Rönnberg, 2003; Pichora-Fuller, 2007; Zekveld, et al., 2007). In order to understand inter-individual differences in the ability to perceive speech in noise, independent measures of both types of abilities are required. A test that was designed to tap into the modality-aspecific factors relevant for speech perception in noise is the Text Reception Threshold (TRT) test (Zekveld et al., 2007). The TRT test measures the ability to perceive masked written sentences.

Regarding the nature of the abilities relevant for the TRT test, up till now, the TRT test has been assumed to measure "linguistic closure" or the ability to complete fragmentary verbal information (Besser, Zekveld, Kramer, Rönnberg, & Festen, 2012; George et al., 2007; Zekveld, Kramer, Kessens, Vlaming, & Houtgast, 2009). However, this is a relatively general term. Examination of the specific nature of the abilities assessed by the test will aid the interpretation of the shared variance in SRT and TRT performances. For this purpose, we first reviewed studies reporting relationships between the TRT and auditory, cognitive, or general (e.g., age) factors in order to formulate specific hypotheses regarding the nature of the abilities relevant for the TRT test. Subsequently, these hypotheses were tested using a comprehensive dataset of 215 participants of the Swedish n200 study (Rönnberg et al., 2016). Finally, we tested the hypothesis that the TRT is associated with the SRT within the n200 data set in order to confirm the relationship between these two tests in the current sample. Addressing both aims is relevant for future studies in which the TRT test is applied: the first one for studies examining the TRT and/or any relationships between the TRT and other factors, and the second one is relevant for studies that aim to test the association between subjectively and/or objectively assessed speech perception and cognitive abilities such as tapped into by the TRT test.

Candidate TRT predictors

The current literature indicates a relationship between the TRT and several cognitive and more general individual factors. For example, TRT has been shown to be related to the *updating ability* of young, normal-hearing (NH) listeners (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013a, 2013b), and to the *inhibition ability* of young NH and older hearing-impaired (HI) listeners (Mishra et al., 2013a, Mishra, Stenfelt, Lunner, Rönnberg, & Rudner, 2014). Both updating and inhibition ability are executive functions (Miyake, Emerson, & Friedman, 2000) relevant for working memory processing. Further, associations with TRTs have been observed for measures of *global cognitive processing* (Humes, Kidd, & Lentz, 2013) and verbal *working memory capacity* in young NH and elderly HI participants (Humes et al., 2013; Mishra et al., 2013a,b; 2014; Koelewijn, Zekveld, Festen, Rönnberg, & Kramer, 2012; Koelewijn, Zekveld, Festen, & Kramer, 2014). However, other studies failed to demonstrate these relationships (Besser, Koelewijn, Zekveld, Kramer, & Festen, 2013; Mishra et al., 2014; Zekveld et al., 2011a). Mixed results have also been observed regarding the association between TRT and *age*. Higher age is related to slightly worse TRTs (Kramer, Zekveld, & Houtgast, 2009; Krull, Humes & Kidd, 2012). However, most of the studies showing this relation compared TRTs of young NH with those of elderly HI participants (Besser, Festen, Goverts, Kramer, & Pichora-Fuller, 2015; Haumann et al., 2012; Krull et al., 2012; Mishra et al., 2013a, 2014; Zekveld, Kramer, & Festen, 2011b). Therefore, *hearing acuity* may have influenced these results to some extent, as was indicated by Rönnberg et al. (2011). Rönnberg et al. demonstrated that worse hearing thresholds are associated to poorer episodic and semantic long-term memory. This indicates that hearing acuity may affect cognitive abilities independent of age, and possibly, also the TRT and/or abilities associated with the TRT. Therefore, it is relevant to take into account hearing acuity when comparing inter-individual differences in linguistic and cognitive functions.

There are a number of candidate abilities that may be related to the TRT test that have not been assessed in combination with the TRT yet. Here, we list several factors that were assessed in the current study (for details, see below). Other potentially relevant factors are mentioned in the Discussion section. First, *context-bound verbal inference-making ability* is probably highly relevant for the TRT test. This ability is relevant for performing the sentence completion test. In this test, participants are asked to complete partly incomplete sentences that are visually presented (Lyxell & Rönnberg, 1987, 1989; Rönnberg, 1990; Rönnberg et al., 2016). The test is similar to the TRT test, except that the TRT test additionally includes the subtasks of target-masker segregation and the identification of word-boundaries behind the mask. Secondly, *lexical access* as assessed with a lexical decision making task could be associated with the TRT. In the lexical decision task, participants have to judge whether a string of three letters constitute a real word (Rönnberg, 1990; Rönnberg et al., 2016). We hypothesized that participants who are faster in making these decisions may have an advantage in the TRT test. Furthermore, we expected that *phonological processing* relevant for the rhyme judgment task may be related to the TRT test, as they are more generally relevant for (the acquisition of) reading skills (Nation & Snowling, 2004). In the rhyme judgment test, subjects have to decide whether or not two visually presented tests rhyme. Also, we hypothesized that *lexical inference making* (Hannon & Daneman, 2001; Rönnberg et al., 2016), as assessed by the logical inference-making test, would relate to the TRT. In both tests, inference-making ability is relevant, albeit at different levels as the lexical inference making tests assesses the ability to evaluate whether successive statements are correct or incorrect. Finally, *visual acuity* may be associated with the TRT in the sense that lower visual acuity may hinder the reading of the text and/or affect the contrast of the TRT stimuli.

Abilities that are likely *not* related to the TRT include peripheral auditory measures such as frequency and temporal auditory resolution (George et al., 2007), modulation detection and auditory stream segregation (Haumann et al., 2012; Humes et al., 2013).

Relationship between TRT and SRT

Regarding the relationship between the TRT and SRT in HI listeners, Humes et al. (2013) observed that better TRTs are associated with better scores on a “general speech understanding” factor in a large sample of 98 older adults with typical age-related high-frequency hearing loss (see also Zekveld, George, Houtgast & Kramer, 2013). Similar results were observed for HI individuals tested by George et al. (2007) when controlling for auditory temporal resolution. Furthermore, the relationships between speech perception performances and TRT seem to be relatively strong for NH as compared to HI participants when these groups are tested in the same study (e.g., George et al., 2007; Krull et al., 2012; Zekveld et al., 2011b). However, Besser et al. (2013) reviewed the evidence regarding the relationship between the TRT and SRT and concluded that speech recognition by young NH participants seems relatively less prone to influences by abilities assessed with the TRT. Obviously, the sample size and other study aspects (e.g. the characteristics of the SRT task applied) should be taken into account when comparing such results. In several studies, no statistically significant relationship has been observed between the TRT and speech perception scores of NH listeners for speech in stationary noise (Schoof & Rosen, 2014) or in interfering speech (Zekveld et al., 2014). Furthermore, an opposite relationship (i.e. better TRTs associated with worse SRTs in amplitude modulated noise) has even been observed in NH listeners (Schoof & Rosen, 2014). Finally, Zekveld et al. (2013) observed that better TRTs were associated with less *subjective* difficulty with sound detection, sound discrimination, and speech intelligibility in quiet and noisy conditions (see also Haumann et al., 2012). For each of these factors, it is important to distinguish meaningful signals and noise or irrelevant sounds. No association between the TRT and subjective sound localization was observed by Zekveld et al. (2013).

Present study

In the present study, we addressed the following research questions: 1) What set of factors best predicts inter-individual differences in the TRT? 2) What are the relationships between the TRT and SRTs estimated in various conditions (i.e., for several masker types and target stimuli) and subjective hearing ability? To answer these questions, we analysed the data of a comprehensive dataset of participants who use hearing aids bilaterally (Rönnberg et al., 2016). These participants performed a test battery that included several measures that are possibly associated with the TRT.

In Analysis 1 we addressed the first research question by building a prediction model in which we distinguished between *conventional* and *novel* predictors. The conventional predictors were predictors that were likely to be associated with the TRT based on the reviewed literature. These included verbal inference making, verbal working memory capacity, updating, and inhibition, with possible (confounding) effects of age, PTA, gender and visual acuity. The novel predictors were variables that were likely to be associated with the TRT based on the assumed shared abilities tapped into by the tests but for which the relationships with the TRT were currently unknown. These novel predictors included lexical inference making, phonological and lexical processing.

Regarding research question 2, we tested the hypothesis that better TRTs would be associated with better SRTs and better subjective ratings of speech perception and hearing quality. As the SRT (and to some extent, the TRT) generally declines with increasing age and reduced hearing acuity, we took into account the possible confounding effect of these variables. The current literature suggests that the association between speech perception and cognitive processing may furthermore depend on the difficulty of the speech perception test as determined by the intelligibility level (e.g., Davis & Johnsrude, 2003; Koelewijn et al., 2014; Rönnberg, 2003; Zekveld, Kramer, & Festen, 2010), and masker type (George et al., 2007; Koelewijn et al., 2012, 2014; Lunner & Sundewall-Thorén, 2007). Also, the amount of contextual information in the sentences may influence this relationship (Moradi et al., 2013; Rönnberg et al., 2016; Schoof & Rosen, 2015). Two types of sentence sets were applied in the SRT tests used in the n200 study (Rönnberg et al., 2016): the hearing in noise (HINT; Hällgren, Larsby & Arlinger, 2006; Nilsson, Soli, & Sullivan, 1994) and the Hagerman matrix sentences (Hagerman, 1982; Hagerman & Kinnefors, 1995). In the natural HINT sentences, the semantic context aids the perception of the sentence words, whereas this is not the case for the Hagerman sentences. Winn (2016) showed that the pupil dilation response reflecting cognitive processing load was smaller during the processing of words that were predicted by the sentence context as compared to unpredictable words. In line with this, we expected that the availability of contextual information would reduce the top-down processes required during sentence perception (Krull et al., 2012). This may be reflected by a weaker association between the TRT and the SRTs estimated with the HINT as compared to the Hagerman sentences.

Method

Sample

The sample analysed in the current study originates from the n200 study that includes data of 215 participants who were bilaterally fitted with hearing aids and had been using the hearing aids for at least one year (Rönnberg et al., 2016). All participants were recruited among patients of the Audiology Clinic of the University Hospital of Linköping. The study was approved by a regional Ethics Committee. Eighty-seven percent of the participants had normal or corrected to normal vision (at least 0.9 on the Snellen chart) and the remaining 13% had at least 0.5 on the Snellen chart. All participants were native speakers of Swedish and each of them provided written informed consent.

In total, 185 participants had a complete dataset for each of the variables included in the analysis (see below). From this dataset, the data of five participants were omitted from the analysis due to extreme (> 4 SDs from the average score) values on one or more variables. The uncorrected average pure-tone hearing thresholds (PTA) of the remaining sample of 180 participants (74 females, mean age = 61.0 years, SE = .63 years) averaged over 500, 1000, 2000, and 4000 Hz of the best ear was 37.5 dB HL (SE = .78 dB HL) and ranged between 10 dB HL and 75 dB HL. The average PTA of the worst ear was 40.6 dB HL (SE = .81 dB HL) and ranged between 12.5 dB HL and 82.5 dB HL. For the majority of the participants, the hearing loss was relatively symmetrical; for 25 participants, the difference in PTA between the left and right ear exceeded 5 dB, and this difference was 12 dB at most for one of the participants. We used the PTA of the best ear in the analyses, as these thresholds are likely most relevant in daily-life listening conditions. See Figure 1.

General procedure

Participants were tested in three test sessions of 2-3 hours each. They were tested by clinical audiologists in the University Hospital of Linköping, Sweden. Participants were allowed to wear their own hearing aids during the tests with visual stimuli.

TRT test

In the TRT test, partly masked written sentences from the Swedish HINT sentence-set (Hällgren et al., 2006) were presented on a PC screen at 40 cm distance from the participant (Zekveld et al., 2007). Text was masked with a bar pattern. First, a bitmap image was created that was proportionally filled with black bars, depending on the required percentage of unmasked text. This percentage was based on the number of pixels the bars consisted of in relation to the number of pixels of the intermediate fields where the text was visible. The bitmap image was then stretched to the fixed dimensions of the field in which the words appeared. For each trial, the bar pattern consisted of bars of equal thickness. Between trials, we varied the percentage of unmasked text by changing the thickness of the presented bars. The text color was red, the color of the mask was black and the text was presented on a white background. At the start of each trial, the mask became visible for 3000 msec. Then, the text appeared 'behind' it in a word-by-word fashion, until the sentence was entirely presented. The entire sentence remained on the screen for 3500 msec (Zekveld et al., 2007). The duration between the onset of displaying each word and the onset of the next word (or the start of the 3500 ms presentation interval of the entire sentence after the appearance of the last word) was equal to the duration of the utterance of each word in the corresponding audio file (Hällgren et al., 2006). An adaptive procedure was applied that estimated the TRT required for the correct perception of 50% of the sentences. After each correct response, the masking of the sentence was increased by 6%, and it was reduced by 6% after each incorrect response. In total, two lists of 20 sentences were presented. The first list was a practice list. The TRT was the mean percentage of unmasked text of sentence 5 to 21 of the second list. Note that sentence 21 was not actually presented, but that the percentage unmasked text of that trial was determined by the previous sentence. Sentences did not overlap with the sentences used in the SRT test with HINT sentences (see below). Lower TRTs indicate better performances.

Reading span test

The RSpan test (Baddeley, Logie, Nimmosmith, & Brereton, 1985; Daneman and Carpenter, 1980; Rönnberg, Arlinger, Lyxell, & Kinnefors, 1989) measures verbal working memory capacity. In this test, Swedish sentences were presented visually. Half of the sentences were semantically incoherent (e.g., “The car drinks milk.”) and half were coherent (e.g., “The farmer builds a house.”). First, two sets of two sentences were presented, followed by two sets of three sentences, two sets of four sentences, and two sets of five sentences. The sentences were presented in a word-by-word fashion on a PC screen at a rate of one word per 800 msec. After each sentence, participants verbally indicated whether the sentence made sense or not during a fixed 5-sec interval. After each set of sentences, participants recalled all first or all last nouns of the sentences in the set. The experimenter recorded the total number of words correctly recalled regardless of order. The maximum total score was 28.

Sentence completion test

The Sentence Completion Test taps into context-bound verbal inference-making ability (Lyxell & Rönnberg, 1987; Rönnberg et al., 2016). In the test, 28 sentences of 4 to 13 words long were visually presented on a PC screen for 7 seconds. Each sentence misses 2-4 words. Participants were asked to fill in the missing words such that a grammatically and semantically correct sentence was created. The disappearance of the sentence from the screen acted as response-prompt for the participants, who had 30 seconds to answer before the next sentence was presented. The verbal response was scored as correct if the missing words were grammatically and semantically correct in the context of the sentence. The number of correctly produced words was divided by the number of deleted words per sentence, and averaged over all 28 sentences (Andersson, Lyxell, Rönnberg, & Spens, 2001).

Rhyme judgment

Phonological processing speed was assessed by the rhyme judgment task (Classon, Rudner, & Rönnberg, 2013; Lyxell et al., 1994). Two words were presented simultaneously on a PC screen and participants had to indicate whether the words rhymed or not by pressing two predefined buttons for “yes” and “no” (Lyxell, Rönnberg, & Samuelsson, 1994). In total, 32 word-pairs were presented and in half of the trials, the words rhymed. Four conditions of eight trials each were presented: in the matching conditions, the words were orthographically similar and rhyme or they were orthographically similar and did not rhyme. In the non-matching conditions, the words were orthographically dissimilar but rhymed, or they were orthographically similar but did not rhyme. We analyzed the reaction time (RT)of the correct responses, averaged over the four conditions.

Inhibition

The inhibition test (Miyake et al, 2000) applied was a stop-signal task that assessed the ability to inhibit automatic responses (Logan, 1994; Rönnberg et al., 2016). A series of digits was visually presented on a PC screen. One digit [0-9] was presented at the time and participants were asked to press the space bar on a computer keyboard every time they saw a digit other than 3, but to omit pressing the space bar when they saw the digit ”3”. The instruction was to respond as quickly and as correctly as possible. Digits were presented until a response was given or for 1 sec in case of a “3”. We analysed the number of times that the response to ”3” was correctly inhibited. The maximum number of errors equaled 22.

Updating

The “keep-track” updating task (Yntema, 1963) assessed the ability to monitor incoming information for relevant items that were used to update no-longer relevant items in working memory (Morris & Jones, 1990). The material consisted of 36 Swedish mono-, bi-, and tri-syllabic words. Each word belonged to one of six semantic categories: metals, colors, animals, relatives, countries, and fruits. The participants were instructed to attend to only four out of six categories. The names of these target categories were visually presented on a PC screen. Below these category names, the words were sequentially presented in random order at a rate of 3 seconds per word with an inter-stimulus interval of 0.5 second. The participants had to retain the last words presented for each of the four target categories and to verbally recall them at the end of the series. Four trials of nine words each were presented. The performance measure was the number of words correctly recalled (max = 16).

Lexical decision

Lexical access speed was measures with a lexical decision task (e.g., Rönnberg, 1990) in which 40 strings of 3 letters were presented on a PC screen. Half of these letter strings constituted common Swedish three-letter words. Participants had press one of two buttons as accurately and quickly as possible: “yes” (for a real word) or “no” (for a non-word/pseudoword) during a 5000 msec interval before the presentation of the next word. The dependent variable was the average RT of the correct responses.

Logical inference making

Logical inference-making was assessed using a paradigm adopted from Hannon and Daneman (2001; Rönnberg et al., 2016). In the test, a question and two successive statements were simultaneously presented on a PC screen. For example, the question “Is a JAL larger than a PONY?” was presented together with “A JAL is larger than a TOC” and “A TOC is larger than a PONY”. The participants answered by means of pressing one of two buttons (yes or no). We analysed the number of correct responses (max = 16).

Hearing in Noise Test (HINT)

An adaptive Swedish Hearing-in-Noise Test (HINT, Hällgren et al., 2006; Nilsson et al., 1994) was used to estimate SRTs for female target speech presented in stationary noise with the long-term average spectrum of the target speech. The speech-to-noise ratio (SNR) required for the correct perception of 50% of the sentences (i.e., SRT50) was estimated using a list of 20 HINT sentences. In the first trial, both the target speech and the noise were presented at 65 dB SPL (i.e., 0 dB SNR). A one-up-one-down adaptive procedure was applied. The noise level of the subsequent sentences was increased with 2 dB for each correct response, and reduced with 2 dB after each incorrect response. One practice list and one test list was presented. The SRT was the average SNR of sentences 5 to 21 of the test list. The twenty-first sentence was not actually presented, but the SNR of that trial was determined by the response to the twentieth sentence.

Hagerman Matrix SRT test

In the Hagerman test (Hagerman, 1982; Hagerman & Kinnefors, 1995), an adaptive procedure was used to estimate the SNR required for the correct perception of 50% or 80% of the words (Brand, 2000). The sentences were spoken by a female and consist of five words each with a proper noun, verb, numeral, adjective, and noun (Example: “Ann had five red boxes”). Two interleaved tracks of 15 sentences each were presented, one for each threshold. The level of the speech was fixed at 65 dB SPL and the SNR was adaptively varied by adapting the level of the masker with a variable step size that was based on the number of words correctly repeated in the previous trial (Hagerman & Kinnefors, 1995). A decrease in the SNR of about 1 dB was made for each correct word exceeding the targeted threshold (for details, see Hagerman & Kinnerfors). The SRT was the average SNR of the final 10 sentences in each track. The target speech was masked with either unmodulated speech-weighted noise or with a multi-talker masker (4 talkers, 2 males and 2 females). In total, four SRTs as estimated with the Hagerman stimuli were analysed in the present study: the SRT for 50% (SRT50) or 80% (SRT80) correct sentence perception in either the stationary noise masker or the 4-talker babble.

Speech, Spatial and Qualities of Hearing Scale

The self-report Speech, Spatial and Qualities of Hearing (SSQ) scale (Gatehouse & Noble, 2004) consists of three subscales relating to speech perception, quality of sound and spatial hearing. Each subscale comprises of a number of listening situations. The participant scored his/her hearing ability for each situation by marking a scale between 0 and 10. A score of 10 indicates complete ability and a score of zero indicates minimal ability (Noble & Gatehouse, 2006). A total of 50 listening situations were rated. In the present study, only the subscale scores relating to speech perception (max score = 140) and quality of sounds (max score = 190) were analysed.

Auditory stimulus presentation

The presentation level of the stimuli was adapted according the hearing thresholds of the individual listeners. For both the HINT and Hagerman matrix test (see below), audibility of the stimuli was enhanced by linear amplification by an experimental hearing aid (Oticon Epoq XW) positioned in an anechoic box (Brüel & Kjær type 4232). The output of the hearing aid was routed via an ear simulating coupler (Brüel & Kjær type 4157) and NAD Electronics 2400 THX amplifier and presented to the participants through a pair of ER3A insert earphones. This set-up provided linear amplification based on a voice aligned compression rationale (Le Goff, 2015; Ng, Rudner, Lunner, Pedersen, & Rönnberg, 2013) with a linear gain (1:1 compression ratio) for pure-tone input levels in the range from 30 to 90 dB SPL (see Rönnberg et al., 2016, Figure 2 for insertion gain response curves). The gain was based on the average pure-tone hearing thresholds of the octave frequencies between 250 and 8000 Hz of both ears.

Statistical analyses

Two statistical analyses were performed using linear regression analyses. In Analysis 1, we first assessed the univariate relationships between 1) the TRT and each of the potential predictor (i.e., updating, inhibition, verbal working memory, verbal inference-making ability, phonological and lexical processing) and confounder variables (i.e., age, PTA, gender, visual acuity). Subsequently, we applied a backward linear regression analysis to examine which set of variables best predicted inter-individual differences in the TRT. The starting model included all conventional predictor and confounder variables (see above). Multicollinearity was assessed by calculating Pearson correlation coefficients between the potential predictors and confounders and by assessing the variance inflation factor (VIF) of the starting model. A VIF < 5 criterion was applied. All VIFs were below 1.5, indicating no multicollinearity. Subsequently, a variable was excluded from the model if the *p*-value of the relationship with the TRT was >.157, starting with the variable with the highest *p*-value. This exclusion criterion for individual predictors can be used as a substitute for all-subset procedures such as the Aikaike information criterion (AIC; Moons et al., 2015; Sauerbrei, 1999). This relatively lenient exclusion criterion is recommended to increase the predictive power of the regression model (Steyerberg, 2008; Sauerbrei, 1999). This analysis resulted in “Model 1”. In a second step, novel predictors were sequentially added to Model 1. Novel predictors were included based on a relatively stringent significance value of $p < .05$, in order to reduce the possibility of an overly complex final model (Sauerbrei, 1999). In the case of multiple significant novel predictors, the strongest (i.e., most significant) novel predictor was added first, after which the second strongest predictor was added, and so on, until no more significant predictors could be added. We evaluated the AIC values of the resulting candidate models. The AIC value indicates the balance between goodness of fit and overfitting, with lower AIC values indicating better model quality. We tested the constant error variance assumption of linear regression by testing heteroscedasticity of the final model. Finally, robust regression with Huber weighting was performed in the MASS package (Venables & Ripley, 2002) in statistical software R (R Core Team, 2017) in order to validate the robustness of the results of the conventional linear regression analyses.

In Analysis 2, seven separate association models were built. These assessed the relationships between the TRT (independent variable) and the following dependent variables: 1) the HINT SRT50% (stationary noise), the Hagerman SRTs (SRT50, SRT80, both in stationary noise and in 4-talker babble) and 3) the SSQ *speech perception* and *qualities of hearing* subscales. A critical α level of $p < .05$ was applied. For each model, we checked the assumptions of linear regression analysis as described for Analysis 1. Next, for each model, we tested for any confounding by age and PTA. A variable was considered to be a significant confounder if it was 1) univariately associated with TRT ($p < .20$), 2) univariately associated with the dependent variable ($p < .20$), and if 3) the regression coefficient of TRT changed with at least 10% after adding the variable to the model (Grayson, 1987). If both variables met these criteria, the strongest confounder was included in the model first. The additional confounder was only added to that model if criterion 3) was met.

Results

Analysis 1

The means and standard errors of the test performances are presented in Table 1. The mean TRT is similar to the mean TRT observed in previous studies (e.g., Zekveld et al., 2007; Zekveld et al., 2009). The scores on the other tests are also similar to that observed in previous studies. See Rönnberg et al. (2016) for a detailed comparison. Table 1 also shows the Pearson correlation coefficients between the TRT and the potential predictor and confounder variables (except for gender). Higher age, poorer updating ability, smaller working memory capacity, poorer context-bound verbal inference-making ability (sentence completion) and logical inference making ability and slower phonological processing (rhyme judgment) and lexical access (lexical decision making) were all significantly (Bonferroni corrected $p < .05$; correction applied for 55 tests) associated with higher (poorer) TRTs.

Table 2 shows the results of the backward regression analysis (i.e., Model 1). The variables that were excluded from the model during the backward selection process were visual acuity ($p = .88$), inhibition ($p =.75$), and gender ($p =.36$). Lower age, lower (better) PTA, larger working memory capacity, better updating ability and higher context-bound verbal inference-making ability (sentence completion) were all associated with lower (better) TRTs. In total, around 39% of the inter-individual variance in TRT was explained by this model. The AIC value of this model equaled 509.1.

In the next step, we sequentially included the novel predictors in order to assess their contribution to predicting the TRT. These results are presented in Table 3. Model 2 indicated that lexical decision significantly predicted the TRT, with shorter RTs being associated with better TRTs. Note that by including lexical decision making in the model, the prediction by updating ability was no longer significant. The AIC value of this model was 492.2. In Model 3, rhyme judgment was added to Model 1 as a significant predictor of TRT, resulting in an AIC value of 501.2. In this model, the contribution of updating again was no longer significant and was therefore excluded from the model. Thirdly, we examined the contribution of logical inference making ability relative to Model 1 (not shown in Table 3). The prediction by this variable was not statistically significant (p = .25). Finally, in a next model, we assessed whether rhyme judgment added to the prediction of TRT relative to a model in which lexical decision was included and updating was not. The results demonstrated that rhyme judgment (p = .61) did not significantly add to the prediction of the TRT. Therefore, the final set of predictors consisted of age, PTA, reading span, sentence completion ability and lexical decision making, explaining 45% of the variance in the TRT (see Table 3, Model 4).

We tested the assumptions of linear regression analysis and the generalizability of the results. The residual plots of standardized residuals versus the predicted values indicated heteroscedasticity or unequal variance of the residuals (larger variance for higher predicted TRT values). Therefore, we assessed whether robust regression resulted in the same model. This was the case, supporting the robustness of the current findings. Furthermore, AIC values (Sauerbrei, 1999) were calculated. Model 4 had an AIC value of 491.2. The AIC values of Model 2 and 4 likely result in less estimated information loss as compared to Models 1 and 3 (Burnham & Anderson, 2002). However, the values of Model 2 and 4 are so close (492.2 and 491.2, respectively) that they do not allow to select between these two Models. However, Model 4 has fewer predictors, which may be preferred.

Analysis 2

Table 4 presents the descriptive statistics of the variables analysed in Analysis 2. It also presents the Pearson correlation coefficients between TRT, age, PTA, the five SRTs estimated for the correct perception of either 50% or 80% of HINT or Hagerman sentences in stationary noise or 4-talker babble, and the ratings on the two SSQ scales. The results of the seven separate regression analyses of the associations between the TRT and the SRT and SSQ variables are shown in Table 5.

Univariately, better TRTs were significantly associated with better SRT thresholds, except for the SRT80 for Hagerman sentences presented in speech-shaped noise, and the SSQ speech rating (see Table 5). For each of these variables, age and/or PTA turned out to be significant confounders. The relation between the SRT50 in 4-talker babble as assessed with Hagerman sentences remained significant with the inclusion of both age and PTA in the model (model 11). Also, the association between TRT and the SSQ speech rating remained significant with the inclusion of the only confounder PTA (model 19). In the remaining models, the association between TRT and the outcome measure was no longer significant (p > .09 for all associations) when including one or both confounder variables in de model. For details, see Table 5.

Discussion

TRT predictors

The first aim of the current study was to examine the nature of the abilities relevant for performing the TRT test. The TRT test is widely applied for examination of the non-auditory part of the variance in inter-individual differences in speech perception performances (Zekveld et al., 2007). The present study with data from a large sample of hearing-aid users demonstrates that the test taps into 1) abilities relevant for filling missing words in incomplete sentences (i.e., context-bound verbal inference-making ability; Lyxell & Rönnberg, 1987; Rönnberg et al., 2016), 2) lexical access, and 3) working memory capacity. Furthermore, a higher age and poorer hearing acuity were associated with slightly poorer TRTs - but note that the association between TRT and PTA was relatively weak, see Table 3, Model 4. When considering the nature of the TRT test, the reliance on these abilities is not surprising.

Firstly, as expected, completing incomplete sentences is relevant for perceiving partially masked sentences. It is likely that context-bound verbal inference-making ability partly depends on semantic processes, as previous studies showed better TRTs when the sentence material used in the test contained more contextual information (Besser et al., 2015; Krull et al., 2012; Schoof & Rosen, 2015; Goverts, Huysmans, Kramer, De Groot, & Houtgast, 2011). However, syntactical and lexical processes also play a role in the TRT test (Goverts et al., 2011 and current results). In the sentence-completion test, entire words are missing, whereas the bars in the TRT test mask parts of the individual letter and word-boundaries. As discussed in the Introduction section, this may affect the relative importance of the linguistic processes relevant for performing these two tests. For example, the TRT likely depends to a greater extent on phonological processing and phonotactical constraints. It would be interesting to assess this hypothesis in more detail in future research.

Furthermore, lexical access speed predicted the TRT. Deciding whether a string of characters constitute a word or not likely contributes to the identification of the (partly masked) word-boundaries in the TRT test. Lexical access has been shown to be relevant for speech perception as well, and is a crucial factor in the Ease of Language Understanding Model (Rönnberg, 2003). In the TRT test, the participants have limited time to read the sentence and generate a response, which may have influenced the association with the lexical decision RT measure (Besser et al., 2012; Zekveld et al., 2011b). This may be especially true when another version of the TRT test is applied in which the text is only presented for 500 msec instead of 3500 msec (see Besser et al., 2012).

Furthermore, larger working memory capacity as assessed with the reading span test was associated with better TRTs. This association was similar to the association observed in HI participants in previous studies (Humes et al. 2013 and Koelewijn et al., 2014; but see Mishra et al., 2014). The commonalities of the TRT and reading span tests include that they both require sentence processing and invoke memory processes. For the TRT test, the memory processing aspect is rather limited in the sense that a single sentence has to be constructed and maintained until the participant responds. We conclude that the TRT taps into processes relevant for the making of speeded lexical decisions when completing partly masked sentences and that these processes require working memory capacity. Or, in short: the TRT tests taps into *controlled sentence completion ability and lexical access*.

The word-by-word presentation style of the stimuli presented in the TRT test may have influenced the reading of the text (e.g., by making it relatively unnatural) such that it influenced the association between the TRT and the cognitive abilities tested in the current study. Although the words were presented one by one, they remained on the screen until the entire sentence was completely presented for another 3500 msec. Besser et al. (2012) showed that other versions of the TRT test (e.g., versions in which only one word is presented at the time) tend to be more strongly associated with working memory. Also, the current word-by-word presentation of the sentences aided the reader to identify the partly masked word-boundaries, which may have reduced the difficulty of the test and may have made it more similar to normal reading. As far as we are aware, there is currently no study available that examines the relationship between speech perception and a TRT test version in which the sentences are entirely presented at once. If such a presentation style would reduce the reliance on visual processes, the relationship between the TRT test and auditory tests like the SRT test may increase.

Finally, the age-related decline in the TRT as observed in the current sample is similar to the association observed in previous studies (e.g., Kramer et al., 2009; Humes et al., 2013). Generally, the univariate correlation coefficient between age and TRT is between -.20 and -.30 (George et al., 2007; Zekveld et al., 2013) in samples of HI listeners. The present results indicate that over and above this age-related decline in TRT, poorer pure-tone hearing thresholds are also relatively weakly (univariate correlation coefficient around .20) associated with poorer TRT performance (c.f., Rönnberg et al., 2011). Hence, both differences in age and differences in hearing acuity between participants may contribute to any differences in TRTs observed (e.g., Besser et al., 2015; Haumann et al., 2012; Krull et al., 2012; Mishra et al., 2013a, 2014; Zekveld et al., 2011b).

Interestingly, several variables did *not* significantly predict the TRT. First of all, the data confirmed the assumption that the TRT is not strongly related to peripheral visual acuity, except for a prerequisite visual acuity required for reading sentences relatively quickly. More informatively, the executive functions updating and inhibition ability did not contribute to explaining the TRT over and above the predictors included in final Model 4. This contrasts with previous findings of Mishra et al. (2013a, 2013b; 2014). Mishra et al. (2013a, 2014; 2013b) applied different tests of updating and inhibition, which may relate to the differences between the current and their results. Also, lexical inference making did not predict the TRT, indicating that higher-order deductive decision making processes are not highly relevant for performing the TRT test.

Please note that over half of the variance in TRT performance was not accounted for in the final model. This indicates that the contributing variables described above only provide a rough outline of the abilities tapped into by the TRT test. Other potentially relevant variables that were not assessed in the present study may include a more general processing speed factor. Namely, rhyme judgement (a RT measure that was included in Model 3) was not significant anymore when lexical access speed was included as well. Therefore, the relationship between these variables and TRT may be partly based on a general processing speed factor. Also, the current results likely depend on the characteristics of the sentences used in the TRT test. For example, if sentences with lower semantical constraints (e.g. the Hagerman sentence set) would be used, the relevance of context-bound verbal inference-making ability would likely be lower, and this may in turn result in a greater importance of lower-level perceptual processes.

Relationship between TRT and SRT

The second aim of the present study was to assess the association between the TRT and the SRT estimated in different conditions, and between the TRT and subjective ratings of hearing difficulties as assessed with the SSQ scale. In line with our expectations, the TRT was significantly associated with each of the SRTs, except the SRT80 for Hagerman sentences presented in speech-shaped noise (see Table 5).

In two models, the TRT remained significantly associated with the outcome measure when controlling for PTA and/or age. Namely, better TRTs were significantly related to better SRT50s estimated with Hagerman sentences in 4-talker babble and with better subjective hearing as assessed with the SSQ speech scale. In both association models, PTA was a significant confounder. Age was additionally included as confounder in the model explaining the Hagerman SRT50 in 4-talker babble.

The results suggest that individuals who have better controlled sentence completion abilities and lexical access (i.e., TRTs) have better SRTs in a relatively challenging listening condition. The task in this SRT condition was relatively difficult as the intelligibility level was around 50%, the matrix sentences lack natural semantic constraints, and the 4-talker babble masker had informational masking effects. This corresponds to models of speech perception such as the Ease of Language Understanding model (Rönnberg, 2003; Rönnberg et. al., 2013) and previous experimental studies that indicated that informational masking increases top-down processing during listening (e.g., Koelewijn et al., 2012). The association between better TRTs and fewer difficulties with speech perception as examined using the SSQ speech scale indicates that better sentence completion abilities and lexical access are associated with better subjective ability to perceive speech in a range of listening situations. This result is in line with Zekveld et al. (2013) and Haumann et al. (2012) who also observed that better TRTs were related to better subjective hearing ability. In the final association model, PTA was included as confounder, and hence, the influence of PTA on subjective hearing ability was accounted for. The current relationships between the TRT and the two outcome measures support that verbal cognitive functions as assessed with visual stimuli are relevant for speech comprehension in challenging laboratory-based tests and as investigated using questionnaires concerning daily-life listening conditions (see also e.g. Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009; Akeroyd, 2008; Rönnberg, 2003). The present results are not in line with the lack of an association between cognitive abilities and subjective hearing ability as was observed by Heinrich et al. (2016).

The TRT was associated with the SRTs assessed with HINT sentences, with the Hagerman SRT50 assessed with speech shaped noise, and with the Hagerman SRT80 estimated in 4-talker babble. However, the inclusion of the confounder variables age and/or PTA in these association models weakened the associations between TRT and the SRT such that they were no longer significant. This indicates that the relationship between TRT and these SRTs is partly based on shared variance in inter-individual differences in age and PTA. The relevance of PTA and age for speech perception tests (SRTs) and PTA for subjective hearing ability is not surprising. Note that the relationship between PTA and the SRTs were observed even though the audibility of the SRT stimuli was enhanced by individualized linear amplification.

Relative to previous studies that applied the TRT test (e.g., Besser et al., 2013, Mishra et al., 2013a,b, 2014; Zekveld et al., 2007), the sample size assessed in the current present study was relatively large. Humes et al. (2013) have described a comparable sample of 98 hearing-impaired participants. In their study, TRTs explained around 10% of aided speech perception thresholds, which is comparable to the current results for the Hagerman SRT50 in 4-talker babble. The replication of the association between the TRT and speech perception performances in different laboratories, languages, and with different SRT and TRT stimuli demonstrates the value of including the TRT test in experimental and clinical settings.

Limitations of the current study

Limitations of the current study include the limited selection of factors included in the regression analyses performed in Analysis 1, and the potential influence of the specific characteristics of the currently adopted TRT test (e.g. the word-by-word presentation of the sentences). Another limitation of the current work is that only hearing-aid users were included. Previous studies suggest that the association between the TRT and speech perception performances may be stronger in other (normal-hearing) samples (see Besser et al., 2013 for a review, but see also George et al., 2007). Furthermore, in the current study, an experimental hearing aid was applied. As a result, the experimental amplification may have differed from the amplification used by the participants in their daily life. Novel amplification may increase the involvement of capacity as compared to amplification settings that participants are used to (Ng et al., 2014). This would imply an overestimation of the relationship between the TRT and the SRT for Hagerman sentences in 4-talker babble, but likely has not affected the relationship between the TRT and the subjectively rated speech perception difficulties (SSQ speech scale).

Conclusions

In conclusion, the present study provides more insight into the nature of the functions relevant for speech perception when assessed with the TRT test. The TRT test taps into *controlled sentence completion processes* and *lexical access.* The TRT is associated with speech perception by hearing-aid users in an informational masker (50% SRT in 4-talker babble) and with subjective speech perception ability in daily life, when controlling for the effects of hearing peripheral hearing sensitivity (PTA) and age.

Acknowledgements

This research was supported by a Linnaeus Centre HEAD excellence center grant (349-2007-8654) from the Swedish Research Council and by a program grant from FORTE (2012-1693), awarded to Jerker Rönnberg.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception threshold related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47, Suppl. 2,* S53-S71. doi:10.1080/14992020802301142.

Andersson, U., Lyxell, B., Rönnberg, J., & Spens, K. E. (2001). Cognitive correlates of visual speech understanding in hearing-impaired individuals. *Journal of Deaf Studies and Deaf Education*, *6*(2), 103-116. doi:10.1093/deafed/6.2.103.

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology, 50, 371-384.* doi:10.1111/j.1467-9450.2009.00753.x.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423. doi:10.1016/s1364-6613(00)01538-2.

Baddeley, A., Logie, R., Nimmosmith, I. & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24*, 119–131. doi:10.1016/0749-596x(85)90019-1.

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R. (2013). Spectrotemporal Modulation Sensitivity as a Predictor of Speech Intelligibility for Hearing-Impaired Listeners. *Journal of the American Academy of Audiology, 24(4),* 293–306. doi:10.3766/jaaa.24.4.5.

Bernstein, J. G. W., Danielsson, H., Ha llgren, M., Stenfelt, S., Ro nnberg, J., & Lunner, T. (2016). Spectrotemporal Modulation Sensitivity as a Predictor of Speech-Reception Performance in Noise With Hearing Aids. *Trends in Hearing, 20(0).* doi:10.1177/2331216516670387.

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., & Pichora-Fuller, M. K. (2015). Speech-in-Speech Listening on the LiSN-S Test by Older Adults With Good Audiograms Depends on Cognition and Hearing Acuity at High Frequencies. *Ear and Hearing, 36(1),* 24–41. doi:10.1097/aud.0000000000000096.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise--a review. *Trends in Amplification, 17,* 75-93. doi:10.1177/1084713813495459.

Besser, J., Zekveld, A. A.**,** Kramer, S. E., Rönnberg, J, & Festen, J. M. (2012). New measures of masked text recognition in relation to speech-in-noise perception and their associations with age and cognitive abilities. *Journal of Speech, Language, and Hearing Research*, *55,* 194-209. doi:10.1044/1092-4388(2011/11-0008)

Brand, T. (2000). *Analysis and optimization of psychophysical procedures in audiology*. BIS Verlag.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261-304.

Classon, E., Rudner, M. & Rönnberg, J. (2013). Working memory compensates for hearing related phonological processing deficit. *Journal of Communication Disorders, 46*, 17–29. oi:10.1016/j.jcomdis.2012.10.001.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 561–584. doi:10.1037/0278-7393.9.4.561.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423-3431.

Gatehouse, S. & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology, 43*, 85–99. doi:10.1080/14992020400050014.

George, E.L.J., Zekveld, A.A., Kramer, S.E., Goverts, S.T., Festen, J.M. & Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America, 121*, 2362-2375. doi:10.1121/1.2642072

Goverts S. T., Huysmans E., Kramer S. E., de Groot A. M., Houtgast T. (2011). On the use of the distortion-sensitivity approach in examining the role of linguistic abilities in speech understanding in noise. *Journal of Speech, Language, and Hearing Research, 54*, 1702-1708. doi:10.1044/1092-4388(2011/09-0268).

Grayson D. A. (1987). Confounding confounding. *American Journal of Epidemiology, 126,* 546-553. doi: 10.1093/oxfordjournals.aje.a114687.

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology*, *11*(2), 79-87. doi:10.3109/01050398209076203.

Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scandinavian Audiology*, *24*(1), 71-77. doi:10.3109/14992029509042213.

Hällgren M., Larsby B., & Arlinger S. (2006). A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*, 227–237. doi:10.1080/14992020500429583.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*, 103-128. doi:10.1037/0022-0663.93.1.103.

Haumann, S., Hohmann, V., Meis, M., Herzke, T., Lenarz, T., & Büchner, A. (2012). Indication criteria for cochlear implants and hearing aids: impact of audiological and non-audiological findings. *Audiology Research*, *2(e12),* 55-64. doi:10.4081/audiores.2012.e12.

Humes, L.E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology, 18,* 590-603 doi: 10.3766/jaaa.18.7.6.

Humes, L. E., Kidd, G. R., & Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Frontiers in Systems Neuroscience, 7,* doi: 10.3389/fnsys.2013.00055.

International Organization for Standardization (2000). *Acoustics Statistical Distribution of Hearing Thresholds as a Function of Age.* ISO 7029. Geneva: ISO.

Koelewijn, T., Zekveld, A.A., Festen, J.M., Rö nnberg, J., Kramer, S.E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology,* Article ID 865731, 11 pages. doi:10.1155/2012/865731.

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S.E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *The Journal of the Acoustical Society of America, 153,* 1596-1606. doi:10.1121/1.4863198.

Kramer, S.E., Zekveld, A.A., Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold test as a visual equivalent of the SRT test. *The Scandinavian Journal of Psychology, 50*, 507-515. doi:10.1111/j.1467-9450.2009.00747.x.

Krull, V., Humes, L. E., & Kidd, G. R. (2012). Reconstructing wholes from parts: effects of modality, age, and hearing loss on word recognition. *Ear and Hearing*, *34*(2), e14-23. doi:10.1097/aud.0b013e31826d0c27.

Le Goff, N. (2015). Amplifying soft sounds—A personal matter [White paper]. Retrieved from Oticon A/S http://www.oticon.global/professionals/evidence

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In: Dagenbach, D. & Carr, T.H. (eds.) *Inhibitory Processes in Attention, Memory, and Language*. San Diego, CA, US: Academic Press, p. 189–239.

Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology, 18*(7), 604-617. https://doi.org/10.3766/jaaa.18.7.7

Lyxell, B., & Rönnberg. J. (1987). Guessing and speechreading. *British Journal of Audiology, 21*, 13-20. doi:10.3109/03005368709077769.

Lyxell, B., & Rönnberg, J. (1989). Information-processing skills and speechreading. *British Journal of Audiology, 23*, 339-347.

Lyxell, B., Rönnberg, J. & Samuelsson, S. (1994). Internal speech functioning and speechreading in deafened and normal hearing adults. *Scandinavian Journal of Audiology, 23*, 179–185. doi:10.3109/01050399409047505.

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*, 169–183. doi:10.1055/s-2000-7563.

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013a). Seeing the talker's face supports executive processing of speech in steady state noise. *Frontiers in Aging Neuroscience, 7.* doi: 10.3389/fnagi.2013.00096.

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013b). Visual information can hinder working memory processing of speech. *Journal of Speech, Language, and Hearing Research, 56,* 1120-1132. doi:10.1044/1092-4388(2012/12-0033).

Mishra, S., Stenfelt, S., Lunner, T., Rönnberg, J., & Rudner, M. (2014). Cognitive spare capacity in adults with hearing loss. *Frontiers in Aging Neuroscience, 6,* doi: 10.3389/fnagi.2014.00096.

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., ... & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): Explanation and ElaborationThe TRIPOD Statement: explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1-W73.doi: 10.7326/M14-0698

Moradi, S., Lidestam, B., & Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: effects on time and accuracy. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00359

Morris, N. & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*, 111–121. doi:10.1111/j.2044-8295.1990.tb02349.x.

Nation, K ., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*, 342–356. doi:10.1111/j.1467-9817.2004.00238.x

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*, 1085-1099. doi:10.1121/1.408469.

Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, *52*(7), 433-441. doi:10.3109/14992027.2013.776181.

Noble, W. & Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ). *International Journal of Audiology, 45,* 172–181. doi:10.1080/14992020500376933

Pichora-Fuller, M.K. (2007). Audition and cognition: What audiologists need to know about listening (pp. 71-85). In C. Palmer, & R. Seewald (Eds.). *Hearing care for adults*. Phonak: Stäfa, Switzerland.

Pichora-Fuller, M.K. (2008). Use of supportive context by older adult listeners: balancing bottom-up and top-down information processing. *International Journal of Audiology, 47, S72-S82.* doi:10.1080/14992020802307404.

Plomp, R. (2001). The Intelligent Ear: On the Nature of Sound Perception. Mahwah, NJ: Lawrence Erlbaum Associates.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rönnberg, J. (1990). Cognitive and communicative function: the effects of chronological age and "handicap age". *European Journal of Cognitive Psychology, 2*(3), 253-273. doi:10.1080/09541449008406207.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology, 42,* S68-S76. doi:10.3109/14992020309074626.

Rönnberg, J., Arlinger, S., Lyxell, B. & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *Journal of Speech, Language, and Hearing Research, 32,* 725–735.

Ronnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternang, O., Wahlin, A., & Nilsson, L. G. (2011). Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory. *Journal of Speech, Language, and Hearing Research*, *54*(2), 705-726. doi:10.1044/1092-4388(2010/09-0088).

Rönnberg, J., Lunner, T., Hoi Ning Ng, E., Lidestam, B., Zekveld, A.A., Sörqvist, P., … Stenfelt, S. (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *International Journal of Audiology.* doi: http://dx.doi.org/10.1080/14992027.2016.1219775.

Sauerbrei, W. (1999). The use of resampling methods to simplify regression models in medical statistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(3), 313-329. doi:10.1111/1467-9876.00155.

Schoof, T. & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing listeners. *Frontiers in Aging Neuroscience, 6,* doi:10.3389/fnagi.2014.00307.

Schoof, T., & Rosen, S. (2015). High sentence predictability increases the fluctuating masker benefit. *The Journal of the Acoustical Society of America*, *138*(3), EL181-EL186. doi:10.1121/1.4929627.

Steyerberg, E. W. (2008). Clinical prediction models: A practical approach to development, validation, and updating. New York, NY: Springer Science + Business Media.

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S. Fourth Edition*. Springer, New York. ISBN 0-387-95457-0.

Winn, M. (2016). Rapid release from listening effort resulting from sematic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing, 20,* 1-17. doi:10.1177/2331216516669723.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors*, *5*, 7–17.

Zekveld, A. A., George, E. L. J., Kramer, S.E., Goverts, S. T., Festen, J. M., & Houtgast, T. (2007). The development of the text reception threshold test: A visual analogue of the speech reception threshold test. *Journal of Speech, Language and Hearing Research, 50***,** 576-584**.** doi:10.1044/1092-4388(2007/040).

Zekveld, A. A., George, E. L. J., Houtgast, T., & Kramer, S. E. (2013). Cognitive abilities relate to self-reported hearing disability. *Journal of Speech, Language and Hearing Research, 56,* 1364-1372. doi:10.1044/1092-4388(2013/12-0268).

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, 480-490.* doi: 10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011b). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing, 32,* 498-510. doi:10.1097/aud.0b013e31820512bb.

Zekveld, A. A., Kramer, S. E., Kessens, J. M., Vlaming, M. S. M. G., & Houtgast, T. (2009). The influence of age, hearing, and working memory on the speech comprehension benefit derived from an automatic speech recognizer. *Ear and Hearing, 30,* 262-272. doi:10.1097/aud.0b013e3181987063.

Zekveld, A.A., Kramer, S.E., Rönnberg, J., Rudner, M. (2017). In a Concurrent Memory and Auditory Perception Task, the Pupil Dilation Response is more Sensitive to Central Memory Load than to Auditory Stimulus Characteristics. Under revision (*Ear and Hearing*).

Zekveld, A.A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. M. H., Rönnberg, J. (2011a). The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear and Hearing, 32*, e16-e25. doi:10.1097/aud.0b013e318228036a.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J. & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience, 8,* 88. doi:10.3389/fnins.2014.00088.

Figure captions

Figure 1. Median and interquartile range of the pure-tone hearing thresholds of the best ear. The ISO/FDIX 7029: 2000 provides descriptive statistics of the hearing threshold deviation for populations of otologically normal persons of various ages under monaural earphone listening conditions.

Table 1. Descriptive statistics and Pearson correlation coefficients between the Text Reception Threshold (TRT) and potential predictor and confounder variables.

	M (SE)	range	1	2	3	4	5	6	7	8	9	10
1. TRT (% unmasked text) †	55.1 (.38)	39.7-73.2										
2. Age (year)	61 (.63)	33-80	.26*									
3. Visual acuity	.94 (.01)	.4-1.0	-.11	-.19								
4. PTA (dB HL) †	37.5 (.78)	10.0-75.0	.22	.18	-.07							
5. Reading Span (# correct)	15.9 (.29)	5-26	-.38**	-.36**	.11	.02						
6. Inhibition (# errors) †	1.7 (.13)	0-9	-.04	-.19	.08	-.03	.12					
7. Updating (# correct)	10.3 (.20)	2-16	-.29**	-.15	-.08	-.06	.34**	.05				
8. Sentence completion (% correct)	83.2 (1.1)	29.0-100	-.54**	-.11	.18	-.10	.32**	.01	.23			
9. Rhyme judgment RT (msec) †	1683 (30.4)	1004-3204	.40**	.03	-.11	.15	-.25*	-.06	-.31**	-.32**		
10. Lexical decision RT (msec) †	971 (14.4)	652-1675	.49**	.06	-.06	.24	-.27*	-.11	-.27*	-.35**	.71**	
11. Logical inference making (% correct)	63.7 (1.3)	12-88	-.28**	-.34**	.09	-.17	.33**	.21	.18	.20	-.14	-.14

PTA = pure tone average of the best ear (.5, 1, 2, 4 kHz), RT = reaction time. †= Lower scores mean better/faster performances; * Bonferroni corrected $p < .05$; ** Bonferroni corrected $p < .01$.

Table 2. Results backward regression analysis: Prediction Model 1

Predictor	β (SE)	β_{stand}	*t*	*p*
Age	.07 (.04)	.11	1.73	.085
PTA	.08 (.03)	.16	2.55	.012
Reading Span	-.23 (.09)	-.18	-2.54	.012
Sentence completion accuracy	-.15 (.02)	-.43	-6.88	<.001
Updating	-.19 (.12)	-.10	-1.61	.110

PTA = pure tone average of the best ear (.5, 1, 2, 4 kHz)

Table 3. Results of the regression analyses (TRT prediction models) assessing the novel predictors.

Model 2	Predictor	β (SE)	β_{stand}	t	p	R^2	R^2change
	Age	.08 (.04)	.14	2.22	.028		
	PTA	.05 (.03)	.09	1.54	.127		
	Reading Span	-.17 (.09)	-.13	-1.90	.059		
	Sentence completion	-.13 (.02)	-.37	-5.91	<.001		
	Updating	-.11 (.12)	-.06	-.98	.33		
	Lexical decision	.01 (.002)	.28	4.52	.000	.455	.061
Model 3	Age	.08 (.04)	.13	2.09	.038		
	PTA	.06 (.03)	.13	2.15	.056		
	Reading Span	-.19 (.09)	-.15	-2.13	.027		
	Sentence completion	-.13 (.02)	-.39	-6.20	.000		
	Updating	-.11 (.12)	-.06	-.98	.327		
	Rhyme judgement	.003 (.001)	.20	3.34	.001	.427	.030
Model 4	Age	.08 (.04)	.14	2.25	<.001		
	PTA	.05 (.03)	.09	1.54	.124		
	Reading Span	-.19 (.09)	-.14	-2.19	.030		
	Sentence completion	-.13 (.02)	-.37	-6.02	.000		
	Lexical decision	.008 (.002)	.29	4.59	<.001	.452	.067

PTA = pure tone average of the best ear (.5, 1, 2, 4 kHz)

Table 4. Mean and standard error (SE; between parentheses) of the test performances analysed in Analysis 2

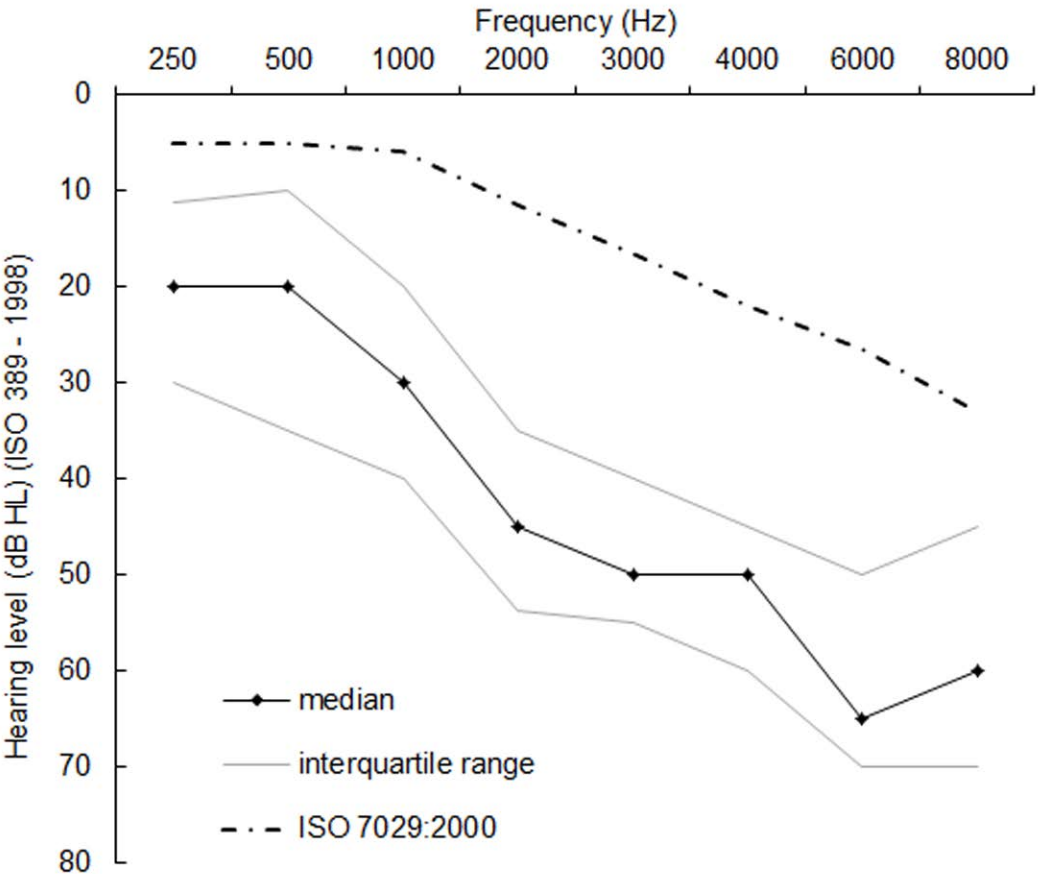
			Hagerman SRT test					
			Stationary noise		4-talker babble masker		SSQ ratings	
	TRT	HINT SRT50 test	SRT50	SRT80	SRT50	SRT80	Speech rating	Quality rating
M (SE)	55.1 (.38)	-1.51 (.12)	-6.2 (.12)	-1.8 (.22)	-.88 (.12)	3.8 (.21)	79.4 (1.8)	132.9 (2.2)
Pearson correlation coefficients between TRT, Age, PTA, the SRTs, and Speech, Spatial and Qualities (SSQ) ratings.								
TRT	-	.17	.18	.13	.29**	.23*	-.22	-.14
Age	.26**	.26*	.32**	.23*	.26*	.28**	-.10	-.05
PTA	.22*	.37**	.40**	.31**	.42**	.39**	-.26*	-.30**

TRT = text reception threshold; PTA = pure tone average of the best ear (.5, 1, 2, 4 kHz); STM = spectro-temporal modulation; SNR = speech-to-noise ratio; HINT = hearing in noise test; SRT = speech reception threshold. * = Bonferroni corrected *p* <.05; ** = Bonferroni corrected *p* < .01.

Table 5. Results of the regression analyses (association models) assessing the associations between the TRT and speech perception outcome measures.

Dependent variable		Model #	Indep. variable	β (SE)	$β_{stand}$	*t*	*p*	R^2	R^2change
HINT SRT50		1	TRT	.05 (.02)	.17	2.3	.001	.03	
		2	TRT	.03 (.02)	.11	1.6	.09		
			Age	.04 (.01)	.23	3.0	.003	.08	.05
		3	TRT	.03 (.02)	.10	1.3	.19		
			PTA	.05 (.01)	.35	4.9	<.001	.14	.11
Hagerman-SSN	SRT50	4	TRT	.06 (.02)	.18	2.5	.02	.03	
	SRT80	7	TRT	.09 (.04)	.13	1.7	.09	.02	
Hagerman-4-talker babble	SRT50	8	TRT	.09 (.02)	.29	4.0	<.001	.08	
		9	TRT	.08 (.02)	.24	3.2	.001		
			Age	.04 (.01)	.19	2.6	.01	.12	.03
		10	TRT	.07 (.02)	.22	3.0	.003		
			PTA	.06 (.01)	.38	5.5	<.001	.22	.13
		11	TRT	.06 (.02)	.17	2.4	.02		
			PTA	.06 (.01)	.36	5.2	<.001		
			Age	.03 (.01)	.15	2.2	.03	.24	.02
	SRT80	12	TRT	.13 (.04)	.23	3.2	.002	.05	
		13	TRT	.09 (.04)	.17	2.3	.02		
			Age	.08 (.02)	.25	3.3	.001	.11	.05
		14	TRT	.09 (.04)	.15	2.2	.03		
			PTA	.10 (.02)	.35	5.0	<.001	.17	.12
		15	TRT	.06 (.04)	.10	1.5	.15		
			PTA	.09 (.02)	.33	4.7	<.001		
			Age	.07 (.02)	.20	2.8	.005	.21	.04
SSQ speech		16	TRT	-1.0 (.34)	-.22	-2.9	.004	.05	
		17	TRT	-.95 (.35)	-.20	-2.7	.008		
			Age	-.13 (.21)	-.05	-.61	.54	.05	.00
		19	TRT	-.78 (.34)	-.17	-2.3	.02		
			PTA	-50 (.17)	-.22	-3.0	.003	.09	.03
SSQ quality		20	TRT	-.76 (.44)	-.13	-1.7	.09	.02	.02

TRT = text reception threshold; PTA = pure tone average of the best ear (.5, 1, 2, 4 kHz); HINT = hearing in noise test; SSN = speech shaped noise; SRT = speech reception threshold; SSQ = speech, spatial and quality of hearing scale.



Frequency (Hz)
250
500
1000
2000
3000
4000
6000
8000
0
10
20
30
40
50
60
70
80
Hearing level (dB HL) (ISO 389 - 1998)
median
interquartile range
ISO 7029:2000