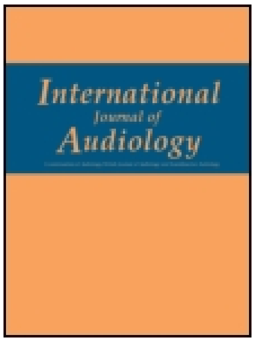

International Journal of Audiology

ISSN: 1499-2027 (Print) 1708-8186 (Online) Journal homepage: http://www.tandfonline.com/loi/iija20

The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise

Marre W. Kaandorp, Annette M.B. De Groot, Joost M. Festen, Cas Smits & S. Theo Goverts

Taylor & Francis
Taylor & Francis Group

International Journal of Audiology 2016; 55: 157–167

International Journal of Audiology

Original Article

The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise

Marre W. Kaandorp[1], Annette M.B. De Groot[2], Joost M. Festen[1], Cas Smits[1] & S. Theo Goverts[1]

[1]Department of Otolaryngology - Head and Neck Surgery, Section Ear & Hearing, and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands and [2]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

The British Society of Audiology

The International Society of Audiology

NAS NORDIC AUDIOLOGICAL SOCIETY

Abstract

Objective: The main objective was to investigate the effect of linguistic abilities (lexical-access ability and vocabulary size) on different measures of speech-in-noise recognition in normal-hearing listeners with various levels of language proficiency. *Design*: Speech reception thresholds (SRTs) were measured for sentences in steady-state (SRT_{stat}) and fluctuating noise (SRT_{fluc}), and for digit-triplets in steady-state noise (DIN). Lexical-access ability was measured with a lexical-decision test and a word-naming test. Vocabulary size was also measured. For the SRT, keyword scoring and sentence scoring were compared. *Study sample*: To introduce variation in linguistic abilities, three groups of 24 young normal-hearing listeners were included: higher-educated native, lower-educated native, and higher-educated non-native listeners. *Results*: Lexical-access ability was most accurately measured with combined results of lexical decision and word naming. Lexical-access ability explained 60% of the variance in SRT. The effect of linguistic abilities on SRTs was up to 5.6 dB for SRT_{stat} and 8 dB for SRT_{fluc}. Using keyword scoring reduced this effect by approximately 1.5 dB. For DIN the effect of linguistic ability was less than 1 dB. *Conclusions*: Lexical-access ability is an important predictor of SRTs in normal-hearing listeners. These results are important to consider in the interpretation of speech-in-noise scores of hearing-impaired listeners.

Key Words: Speech-in-noise recognition; lexical access; vocabulary size; normal hearing

Speech recognition abilities vary considerably among listeners with impaired hearing, and although many of the underlying factors have been subject of investigation, not all variance in performance can yet be explained. In clinical practice a better understanding of the variables involved is needed, for instance for selection of hearing-aid or cochlear-implant candidacy and in user-specific rehabilitation programs. Auditory factors play a major role in explaining speech recognition in noise performance in listeners with impaired hearing (Houtgast & Festen, 2008). Various studies have, however, also shown the importance of cognitive skills, like working memory capacity, information processing, and phonological skills, in speech recognition in noise for listeners with both impaired and normal hearing (e.g. Pichora-Fuller, 2003; Akeroyd, 2008; Houtgast & Festen, 2008; Rönnberg et al, 2008). The relation with linguistic skills, i.e. the level of language proficiency, has also been demonstrated in listeners with normal hearing (e.g. Van Rooij & Plomp, 1990; Van Wijngaarden et al, 2002; Bradlow & Alexander, 2007; Weiss & Dempsey, 2008; Goverts et al, 2011). For instance, Van Wijngaarden and colleagues (2002) showed that non-native listeners needed a 1- to 7-dB more favourable signal-to-noise ratio (SNR) than native listeners in a study on the effect of language proficiency on speech recognition in noise. In the current study we further investigate the relation between linguistic skills and speech recognition in noise performance for several standard speech-in-noise tests in young listeners with normal hearing and various levels of proficiency in the Dutch language. A better understanding of this relation in a diverse group of normal-hearing listeners is needed for the interpretation of speech-in-noise scores of hearing-impaired listeners, where both auditory and non-auditory factors play a role. These results will also be clinically relevant for understanding differences in speech-in-noise scores in listeners with a normal peripheral hearing function in, for instance, educational settings. For this purpose we will also explore test methods that could be applied in the clinic and measure specific linguistic skills which are relevant for speech recognition in noise. The results can serve as a stepping stone for future research of speech in noise performance in hearing-impaired listeners.

Correspondance: Marre W. Kaandorp, Department of Otolaryngology - Head and Neck Surgery, Section Ear & Hearing, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: mw.kaandorp@vumc.nl

(Received 20 November 2014; revised 22 September 2015; accepted 4 October 2015)
ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.3109/14992027.2015.1104735

Abbreviations

DIN	Digits-in-noise
HE-N	Higher-educated native listeners
HE-NN	Higher-educated non-native listeners
LDT	Lexical-decision test
LE-N	Lower-educated native listeners
RT	Response time
SD	Standard deviation
SNR	Signal-to-noise ratio
SRT	Speech reception threshold

Linguistic skills comprise different domains, such as phonology, vocabulary, and syntax. Because *words* (vocabulary) convey most of the meaning of a language utterance and are thus central to the process of language comprehension, we consider the process of a word making contact with its representation in the mental lexicon, in psycholinguistics often referred to as 'word recognition' or 'lexical access' (Grainger & Jacobs, 1996), to be an aspect of linguistic ability that is of major importance for speech recognition in noise. Hence, the current study on the relation between linguistic skills and speech recognition in noise focuses on word recognition. Word recognition concerns the retrieval of knowledge regarding a perceived word, which is stored in the mental lexicon (De Groot, 2011). In this process the number of words stored in this mental lexicon (vocabulary knowledge) could be important. Vocabulary knowledge is assumed to represent *crystallized* knowledge, a type of cognitive ability based on accumulated knowledge (Horn & Cattell, 1967). In addition, the efficiency (i.e. speed) of word retrieval, which we will refer to in this paper as *lexical-access ability*, could also be important. We consider lexical-access ability to represent a *fluid* ability, the capacity of processing information and reasoning (Horn & Cattell, 1967). Benichov and colleagues (2012) studied the role of verbal ability and cognitive function in the use of linguistic context in the recognition of spoken words. Their 53 participants were all native English speakers, aged 19 to 89 years, and varied in educational level and hearing acuity (normal hearing to mild hearing loss). The authors considered verbal ability, composed of vocabulary size and reading scores, to represent crystallized knowledge. They found that this measure of verbal ability was not a predictor of speech recognition performance of words presented in a linguistic context. However tests of episodic memory, working memory and speed of processing, representing fluid cognitive abilities, were significant predictors in their study. The results of a previous study in listeners with mild to severe hearing impairment (Kaandorp et al, 2015) showed that vocabulary size was, on average, smaller for listeners who were not able to perform a speech-in-noise test with full-sentence scoring than for those who were able to perform the test. However, in line with the results of Benichov and colleagues (2012), this linguistic measure did not contribute significantly to a regression model after speech recognition in quiet was used as a predictor of the speech reception threshold (SRT). Thus, in the current study we decided to also focus on more process-related measures of linguistic skills (i.e. fluid abilities).

As a crystallized knowledge-related task we used a vocabulary size test (GIT2, Luteijn & Barelds, 2004) that measures knowledge of low frequency words without a time limit. As a fluid ability-related task we measured lexical-access ability, which is widely examined using different tasks (e.g. lexical decision, Rubenstein et al, 1971; word naming, Balota & Chumbley, 1985; and perceptual identification, Grainger & Segui, 1990). In the current study we used a lexical-decision test (LDT, e.g. De Groot et al, 2002). The lexical-decision test is essentially a discrimination task in which participants have to decide quickly whether a test item is a word or a pseudoword. Pseudowords are letter strings that obey the orthography and phonology of the test language but lack meaning. Using well-formed pseudowords is important to ensure that the response is based on lexical access and not purely on a perceptual process that assesses whether the stimulus looks normal (De Groot, 2011). Most studies in the field of hearing research used only one task at the time to assess lexical-access ability (e.g. Van Rooij & Plomp, 1990; Lyxell et al, 1998; Rönnberg et al, 2000). However, besides lexical-access ability per se, these tasks typically also measure task-specific processes. In the current study we therefore used a second task of lexical-access ability: a word-naming test, which we kept simple, for potential use in the clinic. In lexical decision, both the time duration of the lexical-access component and the discrimination component (time to decide between yes or no) are influenced by experimental variables such as word frequency and familiarity of the stimulus (Balota & Chumbley, 1984). The word-naming task requires pronouncing the word after recognizing it and does also not provide a pure measure of lexical-access ability. For instance, McRae et al. (1990) showed that also the response stage of word naming is sensitive to word frequency. For determining the role of lexical-access skills in speech recognition it is therefore essential to carefully construct these measures and to obtain converging evidence from multiple tests (cf. Grainger & Jacobs, 1996). Combining the results of the lexical-decision (lexical access + discrimination + response) and word-naming (lexical access + speech production) test into a single composite measure will provide a more accurate estimate of lexical-access ability than using either test on its own. Response times (RTs) in both tests have shown to be related to proficiency in the test language (De Groot et al, 2002). In addition, de Groot and colleagues found a stronger correlation between word frequency and lexical-decision RTs for persons with relatively low proficiency than for persons with relatively high proficiency in the test language. We therefore included lists containing words of different usage frequency in the lexical-decision task to evaluate the predictive value of the frequency effect within individual listeners on speech-recognition performance. We used visual stimuli in all tests of linguistic abilities to obtain results that are independent of hearing acuity.

To cover a wide range of linguistic abilities among our participants, we included native listeners with higher and lower education levels, and to obtain a lower boundary of linguistic skills we also included non-native listeners, i.e. students who learned Dutch as a second language. We hypothesized that linguistic abilities of these groups would overlap, resulting in a continuous range of linguistic skills. We assumed that the higher-educated non-native listeners have non-verbal fluid abilities (e.g. processing speed, working memory) comparable to those of the higher-educated native listeners, but perform less well in verbal processing tasks. Non-native listeners have had limited exposure to the Dutch spoken language. This also applies to listeners with congenital hearing impairment and may affect representations in the mental lexicon to an extent comparable to that of non-native listeners. More detailed analyses of language proficiency were, however, beyond the scope of the current study. To avoid any effect of age-related declines on processing speed, only young participants were included.

We measured speech recognition in noise with a sentence test as well as the digits in noise test (DIN test, Smits et al, 2013), as the latter is intended to be less influenced by linguistic abilities. For the sentence test we used steady-state long-term average speech spectrum (LTASS) masking noise and fluctuating noise (two-band speech modulated noise; Festen & Plomp, 1990). In the Netherlands performance of sentence recognition in noise is commonly evaluated with the SRT test (Plomp & Mimpen, 1979; Versfeld et al, 2000). In this test the entire sentence has to be repeated correctly for the response to be marked as correct. To facilitate clinical testing in the future and enlarge the group that is feasible for testing, we decided to adapt the scoring method to allow for small mistakes that have a minimal effect on the meaning of the sentence. This is essentially the method that is used in the hearing in noise test (HINT test, Nilsson et al, 1994), which is widely used outside the Netherlands. In the HINT test deviations in verb tense (e.g. 'is' and 'was') and articles ('a' and 'the') are allowed. For the Dutch test, Versfeld and colleagues (2000) found a 0.7-dB better SRT when small, but not systematically defined, mistakes were allowed in the responses for a group of listeners with normal hearing. In the current study we used a keyword-scoring method, in which the sentence is considered correct when all pre-defined keywords are repeated correctly. In a sub-study, we compared this keyword-scoring method with the standard-scoring procedure for the three groups of listeners to evaluate the effect of this choice on the SRT score.

To summarize, the objectives of the current study are: (1) to investigate the relation of visual lexical-access ability, measured with two tasks, and vocabulary size with speech recognition in noise in normal-hearing listeners with a wide range of proficiency levels in the Dutch language; and (2) to select the most suitable linguistic outcome measures for use in clinic and in research. In addition, the effect of keyword-scoring method in the SRT is examined.

Method

Participants

Three groups of participants with normal hearing participated: 24 native Dutch listeners with higher education (HE-N), 24 native Dutch listeners with lower education (LE-N), and 24 non-native listeners with higher education (HE-NN). Mean ages were 24.0 years ($SD = 5.0$) for HE-N, 27.7 years ($SD = 7.7$) for LE-N, and 28.9 years ($SD = 5.8$) for HE-NN. Participants reported no dyslexia or reading problems in an interview prior to participation. Normal-hearing was defined as pure-tone thresholds equal to or better than 20 dB HL at octave frequencies 500 to 4000 Hz in both ears. Mean thresholds were 3.9 dB HL ($SD = 3.3$) for HE-N, 4.9 dB HL ($SD = 3.3$) for LE-N, and 5.5 dB HL ($SD = 2.7$) for HE-NN. HE-N were all students from the VU University, Amsterdam. LE-N were participants who had their highest degree in intermediate vocational education from various vocational schools. HE-NN were all research university students (at VU University) or had a university degree from other research universities. Their level of Dutch language proficiency was at least level B1 (independent user, CEFR, common European Framework of Reference for languages, Council of Europe). Their mother tongues were English ($n = 4$), German ($n = 7$), Spanish ($n = 3$), French ($n = 2$), Italian, Estonian, Croatian, Tagalog, Somali, Swahili, Arabic, or Moroccan. Their reported experience with the Dutch language ranged from a few months to 22 years ($M = 6.9$ years). For all participants their vision, with corrective eyewear if needed, was checked with a near vision screening test (Bailey & Lovie, 1980). All participants were able to read the words of the chart down to a size of 16 points or lower at approximately 50 cm from the screen. The study was approved by the Medical Ethics Committee of VU University Medical Center.

Tests

Speech recognition measures

A sentences-in-noise test was used as well as a digits-in-noise test. Recognition of sentences in noise was measured with sentence lists (VU98, Versfeld et al, 2000) consisting of 13 short meaningful sentences, pronounced by a female speaker, that were eight or nine syllables in length. They were presented in LTASS masking noise or in fluctuating noise (Festen & Plomp, 1990). The SRT in noise was defined as the SNR where on average 50% of the sentences was repeated correctly as was measured by the adaptive procedure described by Plomp and Mimpen (1979). The speech level was kept constant at 55 dB A and the noise level was varied. The first sentence of each trial was presented at −10 dB SNR for sentences in steady-state noise (SRT_{stat}) and −15 dB SNR for sentences in fluctuating noise (SRT_{fluc}), and was repeatedly presented with a 4-dB increase of SNR until the participant responded correctly. All subsequent sentences were presented only once with SNRs depending on the response to the previous sentence.

Contrary to the procedure described by Plomp and Mimpen we used a keyword-scoring method. For each sentence we defined a set of keywords: all content words and function words that were needed to understand the content of the sentence, for example prepositions that indicate a place or direction (e.g. next to, above, towards), and numerals. A response was considered correct if all keywords in the sentence were repeated correctly in the presented order. After a correct response the SNR was lowered by 2 dB, and after an incorrect response the SNR was raised by 2 dB. The SRT was calculated by taking the average SNR for Sentences 5 to 14 (where Sentence 14 does not exist, but its SNR can be calculated from the response to Sentence 13; Plomp & Mimpen, 1979). For the sub-study that compared keyword and sentence scoring, where responses were considered correct only if the entire sentence was repeated correctly, the same test procedure was used.

Recognition of digits in noise was measured using the DIN test (Smits et al, 2013) that uses digit-triplet lists containing 24 broadband, homogeneous digit-triplets to test digit recognition. The digits were pronounced by a male speaker and were presented in LTASS masking noise. The same adaptive procedure was used for the DIN test as for the SRT tests. Here, the overall intensity level was kept constant at 65 dB A and the first digit-triplet was presented at 0 dB SNR. The SRT was calculated by taking the average SNRs of triplets 5 to 25. All three digits had to be repeated correctly for the response to be considered correct.

Linguistic measures

Vocabulary size was measured with a subtest of the Groningen Intelligence Test - II (Luteijn & Barelds, 2004), which uses a list of 20 items. Participants had to choose the correct synonym out of five words for each visually presented test word. We were interested in differences between participants, not how participants score relative to people with the same age. Therefore, we used raw scores instead of the normally-used age related scores.

For the LDT test (Rubenstein et al, 1971) we used words from the study of De Groot et al. (2002). For these words information was available on the word frequency of occurrence in text, word length, and subjective word concreteness rated on a 7-point scale. We constructed three lists of 60 words each, with different ranges of word frequency for each list. The logarithm of word frequency ranged from 1.89–2.76, 3.10–3.42, and 3.82–4.69, for respectively the low-frequent (LDT_{LF}), middle-frequent (LDT_{MF}), and high-frequent (LDT_{HF}) lists. Lists were matched on subjective word concreteness. All words consisted of 4–7 letters. To each word list forty different pseudowords of 4–7 letters were added. We chose to include more words than pseudowords in each list, because pseudowords could be more salient than normal words, which might cause a bias to 'no' responses. The pseudowords were constructed from words of the CELEX database (Baayen et al, 1993) that were altered by changing at least one letter in such a way that they represent orthographically correct, but meaningless letter strings. Pilot testing showed that these lists of 100 stimuli (words and pseudowords) were too tiring to respond to in one go. Therefore we split the lists in half with two equal word-frequency ranges. The lists were then divided over two test blocks of three lists (in the order: LDT_{MF}, LDT_{HF}, and LDT_{LF}), each containing 30 words and 20 pseudowords. Participants were instructed to respond as quickly and accurately as possible. They were asked to press a green button with their right hand for each word and a red button with their left hand for each pseudoword. The intertrial interval was 1500 ms. RTs for correct responses to words and pseudowords were recorded as well as the number of errors. RTs under 300 ms or above 1500 ms (e.g. Carreiras et al, 1997), as well as RTs that deviated more than 2.5 *SD*s from the participants' resulting list average, were omitted. Several variables, based on RTs for words, that can be derived from this LDT setup were examined to evaluate whether they were suitable to measure lexical-access ability. Pseudowords are generally seen as 'fillers', therefore their RTs were not examined.

We measured word naming with a simplified, short, and easy test. We showed 30 words from the study of De Groot et al. (2002) simultaneously on the screen. For these words the logarithm of word frequency ranged from 2.80–2.95. To stimulate the use of a lexical route in the word-naming task, and avoid just applying script-to-sound conversion rules, we included only words (Tabossi & Laghi, 1992). Participants were instructed to read the words out loud, as quickly as possible. As soon as the words appeared a timer was started. The timer was stopped by the experimenter at the offset of the last word. In all word-naming results we assumed that the experimenter's reaction time is roughly the same (in the order of ms) on a result of seconds of time needed to read the text. The total time needed to read all the words was used as test score. Unfortunately, word-naming scores were accidentally left out for three HE-N and one LE-N participants. A combined variable of lexical decision and speeded word naming was used as measure of lexical-access ability. The group that word naming and lexical access are reported and analysed for comprised 68 participants (21 HE-N, 23 LE-N, and 24 HE-NN).

Questionnaire

The LDT test setup was evaluated by comparing RTs with a subjective measure of how much people read. For experienced readers words are more familiar, thus frequent, which should result in shorter RTs. Participants filled out a questionnaire with seven multiple choice questions concerning their reading behavior. Questions focused on the frequency of reading Dutch reading materials: books, magazines, newspapers, information on the internet, and texts at work, the number of books read, and number of work hours reading. Answers were rated on a 4-point scale ranging from 0 to 3, for 'almost never' to 'multiple times a week', or 'no books' to 'more than 50 books a year', or 'never' to 'more than 20 hours a week'. Sub-scores were summed to a final 'Reading behavior' score.

Table 1. Measurement protocol for the main study (Part 1) and the sub-study (Part 2).

Protocol Part 1 Relation between linguistic abilities and speech recognition in noise (fixed order)

Test	*Test details*
Lexical-decision test	6 lists: Block 1 (MF, HF, LF), Block 2 (MF, HF, LF)
SRT_{stat} test	3 lists: keyword scoring
SRT_{fluc} test	3 lists: keyword scoring
DIN test	3 lists: triplet scoring
Vocabulary size test	1 list
Wordnaming test	1 list

Protocol Part 2 Comparison of scoring methods (Latin square balanced)

SRT_{stat} keywords	3 lists: keyword scoring
SRT_{fluc} keywords	3 lists: keyword scoring
SRT_{stat} sentence	3 lists: sentence scoring
SRT_{fluc} sentence	3 lists: sentence scoring

Note. Lists of the lexical-decision test consisted of high-frequent (HF), middle-frequent (MF), or low-frequent words (LF). SRT_{stat} = speech-reception threshold for sentences in stationary noise; SRT_{fluc} = speech-reception threshold for sentences in fluctuating noise; DIN = digits-in-noise test.

Procedure

Table 1 shows the measurement protocol. In Part 1 of the test session, the main study, tests were presented in a fixed order. All SRTs were measured with keyword scoring in this part. Part 2 concerns the sub-study on the evaluation of keyword scoring. In Part 2 test conditions were presented in a counterbalanced order, using a 4×4 Latin square design. Each condition was measured three times. Tests were performed in a sound-treated booth by a trained experimenter. Pure-tone audiograms were measured with the aid of a clinical audiometer (Decos Audiology Workstation, Decos Systems, Noordwijk, The Netherlands) and a Dell computer with Windows XP. Speech-in-noise tests were measured with a Soundblaster Audigy soundcard and a Soundblaster T20 loudspeaker. Participants were seated either facing the loudspeaker at a distance of approximately 70 cm, or at a comfortable distance from the monitor.

Results

Outcome measures: Speech recognition in noise

The results for SRT_{stat}, SRT_{fluc}, and DIN (Part 1, all keyword scoring method) are presented in Figure 1. The figure shows that HE-NN scored more poorly than the native groups, especially on the sentence SRTs. For the separate participant groups, SRT_{stat}, SRT_{fluc}, and DIN were approximately normally distributed (checked by visual inspection of variable histograms), except for SRT_{stat} in group HE-NN. As a log-transformation (as used by e.g. Smits et al, 2013) did not lead to a normal distribution, we decided to use untransformed variables. Data were analysed with a mixed model with fixed effects for Group (HE-N, LE-N, and HE-NN) and SRT-type (SRT_{stat}, SRT_{fluc}, and DIN), and two-way interaction effects. The analysis showed main effects for Group (Wald $\chi^2 = 51.8$, df = 2, $p < 0.001$) and for SRT-type (Wald $\chi^2 = 1042.0$, df = 2, $p < 0.001$), and an interaction for Group × SRT-type (Wald

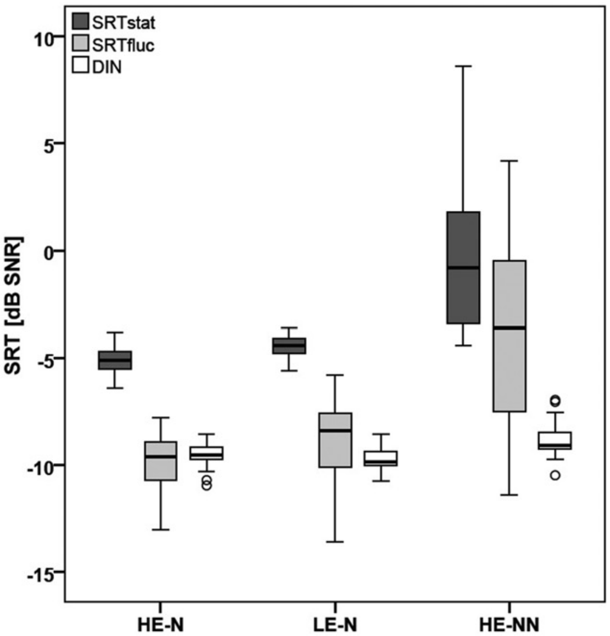

Figure 1. Boxplots of speech-reception thresholds for sentences in stationary noise (SRT_{stat}), sentences in fluctuating noise (SRT_{fluc}), and digits in noise (DIN) for three participant groups: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN).

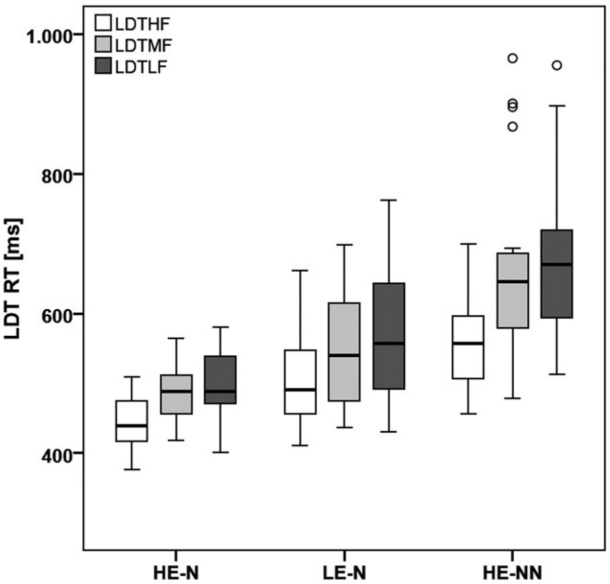

Figure 2. Boxplots for Lexical-Decision test (LDT) results for three lists with different average word frequencies: high frequent (LDT_{HF}), middle frequent (LDT_{MF}), and low frequent (LDT_{LF}). For each frequency the mean of two lists was used. Results are shown for three groups of participants: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN).

$\chi^2 = 45.2$, df = 4, $p < 0.001$). Hence, the difference in SRT between conditions is not equal for the three groups. Post-hoc comparisons with Bonferroni correction for multiple comparisons showed that, compared to HE-N, SRT_{stat} was 0.6 dB poorer for LE-N ($p = 0.01$), and 5.1 dB poorer for HE-NN ($p < 0.01$). SRT_{fluc} was not statistically different for HE-N and LE-N ($p = 1.00$), but was better for both HE-N and LE-N (both $p < 0.01$) than for HE-NN (5.9 and 4.9 dB, respectively). DIN was not statistically different for HE-N and LE-N ($p = 1.00$), but for HE-N and LE-N it was better than for HE-NN, respectively 0.7 dB ($p = 0.01$) and 0.9 dB ($p < 0.01$). It can be concluded that the three study groups have different speech-in-noise recognition abilities. This holds especially for the non-native group compared to the native groups.

Linguistic measures

Lexical-decision test setup

First we evaluated whether the current LDT setup was appropriate to measure lexical-access ability, by evaluating the word-frequency effect and the difference in LDT RT for words between participant groups. Boxplots of the average LDT RTs for words per frequency for each participant group are shown in Figure 2. It shows longer RTs for lower frequent words and longer RTs for LE-N, than for HE-N, and the longest RTs for HE-NN. Lexical-decision RTs were normally distributed per participant group (checked by visual inspection of variable histograms). A mixed model with fixed effects for Group (HE-N, LE-N, HE-NN), Frequency (HF, MF, LF), and Block (1 and 2) and two-way and three-way interaction effects showed a main effect for Block (Wald $\chi^2 = 8.1$, df = 1, $p = 0.005$), but no interaction for Block × Group (Wald $\chi^2 = 0.052$, df = 2, $p = 0.97$). Hence, for all participant groups there was a learning effect (about 14 ms) between Block 1 and Block 2. A main effect was also found for Frequency (Wald $\chi^2 = 206.1$, df = 2, $p < 0.001$). Post-hoc analyses showed that LDT_{HF} was 65 ms shorter than LDT_{MF} ($p < 0.01$), which was 13 ms shorter than LDT_{LF} ($p = 0.03$). This indicates that the test setup is sensitive to word frequency, which is known to influence lexical access. There were interaction effects for Frequency × Group (Wald $\chi^2 = 23.7$, df = 4, $p < 0.001$) and for Frequency × Block (Wald $\chi^2 = 49.8$, df = 2, $p < 0.001$).

Speed-accuracy trade-off occurs when participants with shorter RTs make more errors (Ratcliff et al, 2004). In our data, however, RTs and the number of errors were positively correlated (Spearman's rho $r_s = 0.43$, $p < 0.01$), indicating that no speed accuracy trade-off had occurred. In addition, the questionnaire sum-score Reading behaviour was significantly correlated to the overall LDT mean RT ($r = -0.35$, $p < 0.01$), indicating that participants who read more Dutch reading materials had shorter RTs. This, together with the frequency effect and the absence of a speed-accuracy trade-off, leads to the conclusion that our test-setup is appropriate for measuring lexical-access ability.

Lexical-decision variables

Various lexical-decision variables were examined to evaluate whether they were equally suitable to index differences in lexical-access ability. We evaluated firstly the correlation with wordnaming (our second test for lexical-access ability) and, secondly, how the variables differentiated between the three groups which were chosen to have various levels of proficiency in Dutch. We examined the following LDT variables based on RTs for words: the average RT of all six lists' mean RTs (LDT_6); the average RTs of two equal-frequency lists for LDT_{HF}, LDT_{MF}, and LDT_{LF}; the frequency effect in Block 1 ($LDT_{diff1} = LDT_{LF1} - LDT_{HF1}$); the average frequency effect ($LDT_{diff} = ((LDT_{LF1} - LDT_{HF1}) + (LDT_{LF2} - LDT_{HF2}))/2$); and the

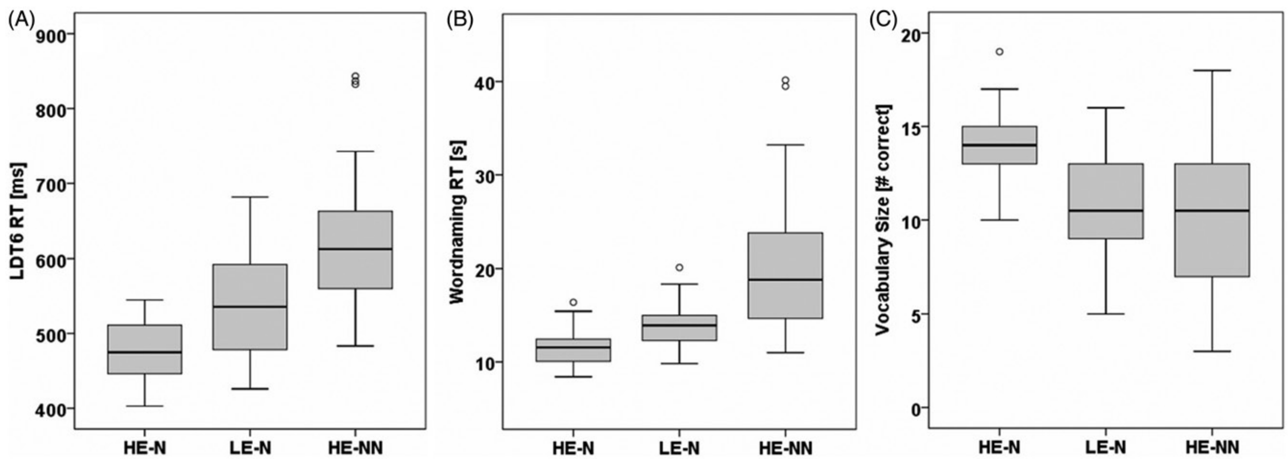

Figure 3. Boxplots for (A) Lexical-decision RTs (LDT_6), (B) Wordnaming RTs, and (C) Vocabulary size scores for the three groups of participants: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN).

mean RT of the very first list (LDT_{MF1}). Spearman's rho correlations were significant ($p < 0.01$) between all LDT variables and wordnaming, with r varying between approximately 0.40 for LDT_{diff1} and LDT_{diff}, 0.59 for the single list LDT_{MF1}, and values around 0.67 for the mean LDT RTs, with the highest for LDT6: $r = 0.69$. A multivariate ANOVA with Bonferroni correction for post-hoc comparisons showed that LDT_6, LDT_{HF}, LDT_{MF}, LDT_{LF}, and LDT_{MF1} discriminated between all three groups (all $p < 0.05$). LDT_{diff} and LDT_{diff1} only discriminated between non-native and native listeners.

Hence, LDT_6, LDT_{HF}, LDT_{MF}, and LDT_{LF} are approximately equally suitable for measuring lexical-access ability. LDT_{MF1}, LDT_{diff}, and LDT_{diff1} are less suitable. We used LDT_6 in further analyses. However, the results indicate that an average of two lists (of the same frequency) should also suffice to measure lexical-decision ability, which can be relevant for use in clinical evaluations.

Group results

Lexical-decision scores (LDT_6) are presented in Figure 3, A. It shows that HE-N were on average faster than LE-N. HE-NN had the poorest performance. These observations in the data were confirmed by the post-hoc analysis with Bonferroni correction, which showed that, compared to HE-N, LDT_6 was 63 ms longer for LE-N ($p < 0.01$), and 153 ms longer for HE-NN ($p < 0.01$). The mean values were 478 ms (*SD* 39), 541 ms (*SD* 74), and 631 ms (*SD* 101) for HE-N, LE-N, and HE-NN, respectively.

Word-naming scores (total reading time for the list of 30 words) are presented in Figure 3, B. It shows that HE-N were slightly faster than LE-N. HE-NN again had the poorest performance. The scores were approximately normally distributed for participant groups HE-N and LE-N. For HE-NN the response times were right skewed, mostly due to two high scores in this group. We decided not to use a transformation on the outcome measure. Mean response times were 11.6 s (*SD* 2.0), 13.9 s (*SD* 2.5), and 20.7 s (*SD* 2.5) for HE-N, LE-N, and HE-NN, respectively. A one way ANOVA showed a main effect for Group, $F(2, 65) = 19.71$, $p < 0.01$. Post-hoc comparisons with Bonferroni correction for multiple comparisons showed that this test differentiates between native and non-native groups ($p < 0.01$), but not between native groups with higher and lower education ($p = 0.43$).

Vocabulary size scores, shown in Figure 3, C, seem better for HE-N than for the two other groups, but equal for LE-N and HE-NN. Mean scores were 14.1 (*SD* 1.9), 11.0 (*SD* 2.7), and 10.1 (*SD* 4.1) correct items for HE-N, LE-N, and HE-NN, respectively. A one-way ANOVA showed a main effect for Group, $F(2, 69) = 11.91$, $p < 0.001$. Post-hoc comparisons with Bonferroni correction for multiple comparisons confirmed a difference in vocabulary size score between HE-N and LE-N ($p = 0.01$), and also between HE-N and HE-NN ($p < 0.01$), but not between LE-N and HE-NN ($p = 0.99$).

Combined variable for lexical access

Word naming and lexical decision are assumed to both measure lexical-access ability in addition to other, task-specific, processes. Previous experiments indicate that lexical-access ability is more precisely measured when looking at the word-naming and lexical-decision measures combined (cf. Jacobs & Grainger, 1996). In this study the Spearman's rho correlation between LDT_6 and wordnaming was $r_s = 0.69$. A composite variable lexical access was calculated by converting the LDT_6 RTs and the wordnaming RTs to z scores and averaging them. In the remaining analyses we used the resulting new variable lexical access.

The Spearman's rho correlation between lexical access and vocabulary size was $r_s = -0.58$ (Figure 4), indicating a fairly high correlation between the different linguistic abilities that are tapped with these tests.

Relation between linguistic measures and speech recognition in noise

To first examine the predictive value of the LDT test, word-naming test and vocabulary size test on speech-in-noise recognition abilities separately, linear regression analyses were performed for SRT_{stat}, SRT_{fluc}, and DIN with these variables. The output of these models is given in Table 2. All variables separately were to some extent significant predictors of each speech-in-noise test.

The relation of both lexical-access ability and vocabulary size with speech-in-noise recognition was analysed with stepwise linear regression analyses for SRT_{stat}, SRT_{fluc}, and DIN separately, including both lexical access and vocabulary size as independent variables. These models showed that vocabulary size did not

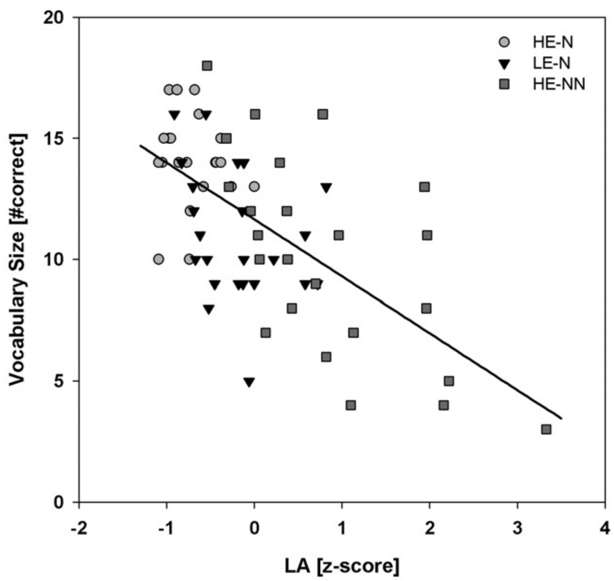

Figure 4. Relation between lexical access (LA) and vocabulary size for three groups of participants: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN).

contribute significantly in combination with lexical access (bottom line Table 2 and Figure 5), which explained 59% of the variance in SRT_{stat}, 60% of the variance in SRT_{fluc}, and 17% of the variance in DIN. Hence, lexical access is an important predictor of speech recognition of sentences in noise.

Keyword scoring compared to sentence scoring

In the main study, the recognition of sentences in noise was measured using keyword scoring. With the data of the sub-study (Part 2), the effect of keyword scoring versus sentence scoring was analysed. The SRTs (Figure 6) show slightly better results for keyword scoring than for sentence scoring. SRTs were approximately normally distributed per participant group (checked by visual inspection of variable histograms). Data were analysed with a mixed model with fixed effects for Group (HE-N, LE-N, or HE-NN), Scoring method (keywords or sentences), and Noise (stationary or fluctuating) and their two-way and three-way interactions and random intercepts for subjects. We found main effects for Group (Wald $\chi^2 = 51.7$, df = 2, $p < 0.001$), Noise (Wald $\chi^2 = 1630.8$, df = 1, $p < 0.001$), and Scoring method (Wald $\chi^2 = 39.0$, df = 1, $p < 0.001$); and interactions for Group × Noise (Wald $\chi^2 = 45.6$, df = 2, $p < 0.001$) and for Group × Scoring (Wald $\chi^2 = 10.5$, df = 2, $p = 0.005$). There was no interaction for Noise × Scoring (Wald $\chi^2 = 3.7$, df = 1, $p = 0.05$) and for Group × Noise × Scoring (Wald $\chi^2 = 0.2$, df = 2, $p = 0.92$). Overall, keyword scoring resulted in a 1.0-dB lower (more favourable) SRT than sentence scoring. Post-hoc comparisons with Bonferroni correction for multiple comparisons showed that scoring method had no significant effect ($p = 0.61$) for HE-N. For LE-N and HE-NN the SRT with keyword scoring was respectively 0.8 dB ($p < 0.01$) and 1.8 dB ($p < 0.01$) better than the SRT with sentence scoring. Since there was no interaction between Noise and Scoring method, the effect of scoring method is assumed to be equal for SRT_{stat} and SRT_{fluc} in our study population.

We also examined the effect on discriminative power between groups for keyword scoring compared to sentence scoring. Post-hoc comparisons with Bonferroni correction for multiple comparisons show that SRTs with sentence scoring were, compared to HE-N, 0.7 dB poorer for LE-N ($p < 0.01$), and − 6.1 dB poorer for HE-NN ($p < 0.01$). For keyword scoring the difference between HE-N and LE-N was not significant ($p = 1.00$). However, the differences between HE-NN and both HE-N and LE-N remained significant and were 5.4 dB ($p < 0.01$) and 5.1 dB ($p < 0.01$), respectively.

To conclude, keyword scoring resulted in a significantly better SRT than sentence scoring, making the test suitable for a wider range of participants. However, application of keyword scoring results in less discriminative power between the SRTs of higher and lower-educated native listeners, but the SRT difference between native and non-native participants remains significant.

Discussion

The main objective of this study was to investigate the relation between linguistic abilities (lexical access and vocabulary size) and speech recognition in noise in young listeners with normal hearing and various levels of language proficiency. To introduce these levels of proficiency we included, in addition to the group of higher-educated native listeners that are often studied, two groups with lower levels of proficiency in Dutch: lower-educated native listeners and higher-educated non-native listeners. This successfully resulted in a wide distribution of scores on the linguistic tests (see e.g. Figure 4).

As expected, SRT_{stat} for non-native listeners was poorer than for native listeners. The difference of 5.1 dB is in line with the literature (Goverts et al, 2011; Van Wijngaarden et al, 2002). In fluctuating noise this difference was larger: HE-NN participants needed a 5.9-dB higher SNR to recognize 50% of the sentences correctly, which can be largely explained by the higher SRT_{stat} (Smits & Festen, 2013). In the DIN test non-native participants needed only a 0.8-dB higher SNR to recognize 50% of the triplets correctly. This supports the idea that digit recognition puts a much lower demand on linguistic abilities than the recognition of sentences (Smits et al, 2013). We found that the limited linguistic skills of our LE-N participants only had a small effect on the recognition of sentences in noise. There was only a small increase in SRT_{stat} for the LE-N participants compared to the HE-N participants, and we found no significant difference in SRT_{fluc} between these participant groups. The latter seems counter-intuitive, because people with higher linguistic skills might be expected to make better use of information from short time frames of better audibility than people with lower linguistic skills. The non-native listeners were, however, less able to use this information. Wang and Humes (2010) showed that linguistic factors affect temporal integration and recognition of interrupted words. In their study, to reach the same degree of word-recognition performance, less acoustic information was required for lexically easy words (with high word frequency, low neighbourhood density, and low neighbourhood frequency) than for lexically hard words (with low word frequency, high neighbourhood density, and high neighbourhood frequency), so it might be concluded that persons with higher linguistic skills also need less acoustic information. For these listeners more words are relatively high frequent. Furthermore, in a correlation analysis, Besser and colleagues (2012) showed that a higher level of education was associated with better performance on both SRT_{stat} and SRT_{fluc}. Though this effect was not found for SRT_{fluc} in the present study,

Table 2. Prediction of SRTs by mean lexical-decision score (LDT_6), wordnaming, vocabulary size, and the combined variable lexical access.

Linear regression						
	SRT_{stat}		SRT_{fluc}		DIN	
Predictor	Adj. R^2	p	Adj. R^2	p	Adj. R^2	p
LDT_6	0.40	<0.001	0.43	<0.001	0.12	<0.001
Wordnaming	0.57	<0.001	0.58	<0.001	0.13	<0.001
Vocabulary size	0.27	<0.001	0.29	<0.001	0.05	0.028
Lexical access	0.59	<0.001	0.60	<0.001	0.17	<0.001
Stepwise linear regression *(variables included: Lexical access and Vocabulary size)*						
Lexical access	0.59	<0.001	0.60	<0.001	0.17	<0.001

Note. Adj. R^2 = Adjusted R^2 values. Predictor variables LDT_6, wordnaming, lexical access, and vocabulary size were first analysed separately with linear regression analyses. Additionally, in a stepwise regression analysis, both lexical access and vocabulary size were included. Only lexical access contributed significantly to the model.

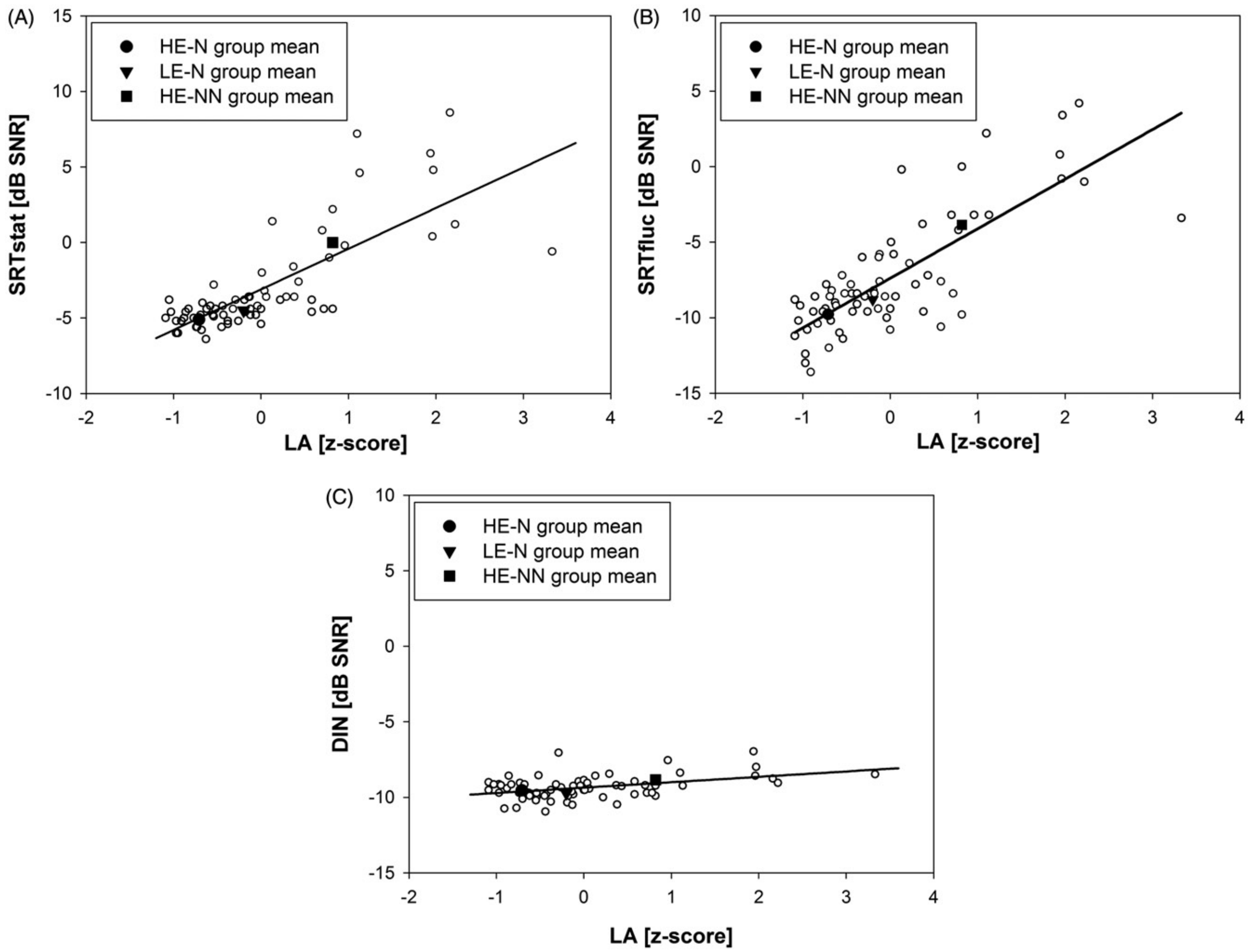

Figure 5. Relation between lexical-access ability (LA) and speech recognition of sentences in (A) stationary noise (SRT_{stat}), (B) fluctuating noise (SRT_{fluc}), and (C) digits in noise (DIN). Mean scores for three groups of normal-hearing participants are shown: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN).

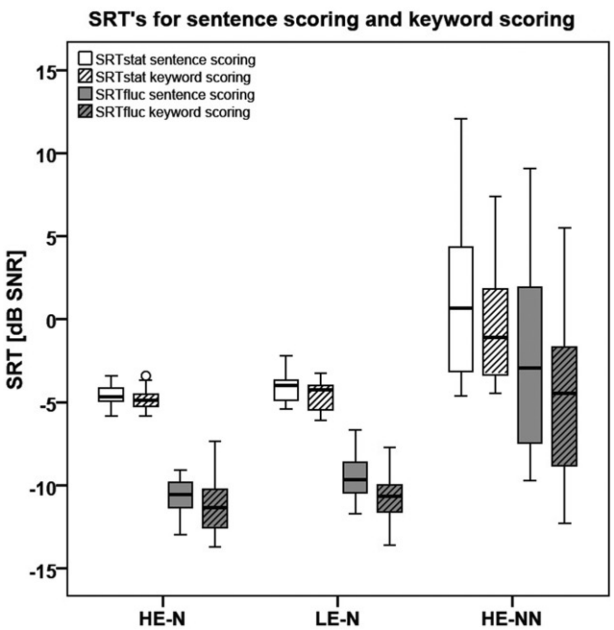

Figure 6. Comparison of SRTs obtained with different scoring methods. Boxplots for SRTs with sentence scoring (plain boxes) and for SRTs with keyword scoring (striped boxes) are given for three groups of participants: Higher Educated Native (HE-N), Lower Educated Native (LE-N), and Higher Educated Non-native (HE-NN). SRT_{stat} boxes are shown in white, SRT_{fluc} boxes are shown in grey.

the median and range of SRT scores (shown in Figure 1) demonstrate that there was more within-group variability for LE-N (with no more than intermediate vocational education level) than for HE-N. In a group with even lower educated listeners, as might be expected in the general society, the effect of educational level on the SRT will probably be more pronounced. Furthermore, it should be noted that we measured the SRT using the (eight to nine syllable) sentence SRT. In communication settings with more complex linguistic content and longer utterances, the effect of educational level on speech recognition will most likely be more relevant.

In this study we combined the results of lexical decision and word naming to diminish the influence of task-specific processes involved in both tasks and obtain a stronger measure of lexical access. The results of the LDT test differed between all three groups of listeners, whereas the time needed to read 30 words in the word-naming test only differentiated between the native and non-native listeners. The composite variable lexical access differed between all three groups. The slower word-naming performance of the non-native group is not necessarily purely caused by slower processing during the lexical-access stage of word naming, but can also partly be caused by slower processing during the speech-production stage of this task (Balota & Chumbley, 1985). A more commonly used test-procedure for word naming with presentation of one word at the time and recording of responses by a voice-operated switch would provide more precise results than our simple test procedure and would most likely reduce some of the interindividual variance in naming latencies.

Vocabulary size scores were different between HE-N and both groups with lower linguistic abilities. LE-N and HE-NN participants scored equally. The difference in discriminative power between lexical access and vocabulary size suggests that these tasks measure different aspects of linguistic ability. It is possible that a difference in word usage frequency of the test items between tasks plays a role. The current LDT test measures the recognition of relatively frequent words, whereas the vocabulary size test measures knowledge of the meaning of infrequent words. In the present study we found that vocabulary size and lexical access were fairly highly correlated ($r_s = -0.58$). Crystallized abilities are thought to develop by the interaction of fluid intelligence and cultural experience (Van Der Maas et al, 2006), and a correlation between crystallized and fluid abilities, in which we can categorize our measures, is often found. It was therefore expected that listeners with reduced linguistic abilities have reduced skills on both measures, despite the differences between tests. In the current study, the HE-NN participants were assumed to largely have the same non-verbal fluid intelligence (e.g. working memory) as the HE-N participants, but a limited exposure to the Dutch language will have caused reduced linguistic fluid abilities. The LE-N participants were assumed to differ only slightly in exposure to the Dutch language from the HE-N participants but have in general somewhat lower fluid abilities. Thus, lexical access and vocabulary size were related, but measure different aspects of linguistic ability and, as a result, discriminate differently between groups.

The main result of this study was that lexical access, composed of lexical-decision and word-naming scores, explained approximately 60% of the variance in SRT outcome in young normal-hearing listeners with a wide range of levels in language proficiency. Lexical access, a measure of fluid ability, was a stronger predictor of SRT than vocabulary size, a measure of crystallized ability. This is in line with results of Benichov et al. (2012), who found that their measure of fluid cognitive ability contributed significantly to the recognition of spoken words, in contrast to their measure of crystallized verbal ability. For recognition of the sentence material used in this study (i.e. meaningful sentences, eight to nine syllables long, composed of relatively frequent words), presented at a normal speech rate, it is perhaps not surprising that fluid processing ability plays a larger role in real-time speech processing, than crystallized knowledge.

These findings can be of interest when considering performance of individuals with normal hearing in speech-in-noise situations, e.g. classrooms and public intercom systems. They are also important for the interpretation of speech-in-noise scores of hearing-impaired listeners. Furthermore, they can be relevant for considerations in rehabilitation of people with hearing loss. Knowledge about the role of fluid and crystallized linguistic abilities in speech recognition will not only help to predict rehabilitation outcome, but also guides the construction of personalized training programs. Here, it is relevant to know which aspects can be trained, and in what modalities, or which aspects influence training results. For instance, the correlation found in this study between subjective reading behaviour and lexical access suggests that lexical access might improve with reading practice. This is, of course, a topic of further research, both fundamental and clinical. For a good understanding of the test results it is also important to know that fluid abilities decline after a certain age, whereas crystallized abilities are largely preserved during adult aging or even improve with aging (Pichora-Fuller, 2008; Horn & Cattell, 1967). Given the speech material used in this study, this effect of aging might lead to poorer performance in elderly patients. In this view, Pichora-Fuller and Levitt (2012) suggest to provide training for older adults in how to use top-down processing based on knowledge to compensate for problems in bottom-up processing of the signal.

The results showed that the DIN test is relatively immune to variations in linguistic abilities, as was intended in its development (Smits et al, 2013). In assessments of speech-in-noise recognition that only concern auditory capacity, the DIN test might therefore be preferred.

One of the basic assumptions of the current study was that lexical-access ability is most accurately measured by combining the two measures of this construct, lexical decision and word naming, thus diminishing the influence of task-specific processes. To evaluate this hypothesis we compared the predictive value of the separate tests to that of the combined variable lexical access. We were particularly interested in the predictive value of wordnaming alone, because this 20-second test is easily implemented in standard test protocols in the clinic, and no specific hardware and software is needed. Regression analyses for the entire study group showed an almost equally strong predictive value for wordnaming compared to the predictive power of the combined lexical-access variable (see Table 2, $R^2 = 0.57$ for SRT_{stat} and 0.58 for SRT_{fluc}). However, we also found that word naming did not discriminate between the two groups of native listeners (higher and lower educated). Therefore, we also performed the regression analyses for the native listeners alone ($n = 44$). Wordnaming separately explained no more than 8–10% of the variance, whereas the combined variable lexical access explained 13–14% of the variance in SRT_{stat} and SRT_{fluc} outcome for only the native groups. Based on these results and given our assumption that the production stage in the word-naming test influences test performance, we assume that indeed lexical-access ability is more accurately tested by the combination of word naming and lexical decision. We therefore recommend the use of both tests for a more accurate measure of lexical-access ability in future research and also for use in the audiology clinic. These tests require simple software and it only takes approximately 5–10 minutes to measure both. Hence, they are feasible in clinical practice. Contrary to the common practice of applying just one test (e.g. Lyxell et al, 1998; Rönnberg et al, 2000; Van Rooij & Plomp, 1990), the use of both tests will enhance the accuracy of identifying the role of lexical-access ability in speech recognition. This will be advantageous for research and clinic.

Lexical decision variables

Evaluation of the various variables from the LDT test showed that RT variables were the best predictors of lexical-access ability. The RT variable LDT_6 discriminated most between the participant groups and correlated most highly with wordnaming. The other variables, reflecting the frequency effect in lexical decision (LDT_{diff}), were less suitable, although in these variables several task-specific factors are excluded. Apparently the total duration of the lexical-access process is a stronger predictor of lexical-access ability than the individual word-frequency effect, in this set of all relatively frequent words. Additionally, LDT_{diff} only represents the difference in lexical-access time between high-frequent and low-frequent words. Part of the lexical-access time is thus excluded in this variable. Compared to LDT_6 the other examined LDT RT variables, that are averages of two lists of equal word frequency, appeared almost equally suitable. Regression analyses repeated with the lexical access variable composed of wordnaming with LDT_{MF} instead of LDT_6 showed similar results. For clinical purposes or future research, the use of two LDT lists (with comparable word-frequency) should thus be appropriate.

Keyword scoring

We used keyword scoring in this study, since we hypothesized that a more tolerant scoring procedure for the SRT would improve SRT scores and thus make the SRT applicable for a wider range of listeners. As mentioned in the Introduction section, Versfeld et al. (2000) found a 0.7-dB better SRT when small mistakes were allowed in the responses for a group of mostly university students with normal hearing, than when whole sentences had to be repeated correctly. Since the slope of the psychometric curve was not affected by this difference in scoring procedure, they concluded that precision of the judgment criteria hardly affects measurement efficiency. In the present sub-study, where keyword scoring was compared to entire-sentence-scoring, keyword scoring resulted in an, on average, 1-dB better SRT than sentence scoring. However, keyword scoring also implies less discriminative power between the groups of participants. The small difference in SRT_{stat} between the two groups of native participants with different levels of education was significant for sentence scoring, but not for keyword scoring in the sub-study. The discriminative power between native and non-native groups for both SRT_{stat} and SRT_{fluc} stayed intact. Therefore, we conclude that the adapted SRT test still involves a significant demand on someone's linguistic skills. The relations found in this study between linguistic variables and SRTs will be even stronger when SRTs are measured by means of sentence scoring.

These results indicate that the keyword-scoring method facilitates the use of sentence-in-noise tests in hearing-impaired listeners in future studies, but a drawback may be the reduction in the effect of linguistic skills on the outcome of the sentence-in-noise test. A wider range of people with lower linguistic abilities or more severe hearing losses (e.g. children, cochlear implant users) will be able to do the test with keyword scoring than with full-sentence scoring. For these listeners the effect of keyword scoring may even be larger since small mistakes in for instance 'he/she/we' are allowed. The Dutch sentence-in-noise test (Versfeld et al, 2000) can therefore be used in a similar way as the HINT test (Nilsson et al, 1994).

To conclude, for evaluation of auditory capacity the DIN test is preferred. Recognition of sentences measured by the Plomp and Mimpen procedure is more suitable for evaluation of auditory functioning in everyday listening situations, but for listeners who are not able to perform the standard sentence test, keyword scoring might be a valuable alternative.

Conclusion

Our lexical-access measure, composed of lexical-decision and word-naming scores, explained approximately 60% of the variance in SRT outcome in listeners with normal hearing and a wide range of language proficiency levels. Performance on the DIN test was much less influenced by linguistic abilities. Hence, lexical access is, more than vocabulary size, an important predictor of speech recognition using sentences in noise. The effect of linguistic abilities on speech recognition in noise can be up to 5.6 dB for recognition of sentences in stationary noise and 8 dB for sentences in fluctuating noise. Using keyword scoring reduces this effect by approximately 1.5 dB. For recognition of digits in noise this effect is less than 1 dB. These results are important for the interpretation of speech-in-noise scores of hearing-impaired listeners. Lexical-access ability is best measured with a lexical-decision and word-naming task, that are clinically feasible, and should ideally be combined to measure lexical-access ability most accurately in relation to speech-in-noise recognition.

Acknowledgements

The authors thank all participants who participated in the study, Hans van Beek for technical support, Jaap Peter Kronemeijer, Katrien Devroye, and Sietske Temminghoff for data collection. We also thank Elske Bolk and Andriana Zekveld for their help in defining the keywords, and Birgit Witte for her statistical guidance.

Declaration of interests: The authors report no conflicts of interest.

References

Akeroyd M.A. 2008. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal-hearing and hearing-impaired adults. *Int J Audiol*, 47(Suppl. 2), S53–S71.

Baayen H., Piepenbrock R. & Van Rijn H. 1993. *The CELEX Lexical Database (CD-ROM)*. Philadelphia, USA: University of Pennsylvania, Linguistic Data Consortium.

Bailey I.L. & Lovie J.E. 1980. The design and use of a new near-vision chart. *Am J Optom Physiol Opt*, 57, 378–387.

Balota D.A. & Chumbley J.I. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J Exp Psychol Hum Percept Perform*, 10(3), 340–357.

Balota D.A. & Chumbley J.I. 1985. The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *J Mem Lang*, 24, 89–106.

Benichov J., Cox L.C., Tun P.A. & Wingfield A. 2012. Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability, and cognitive function. *Ear Hear*, 33(2), 250–256.

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. *J Speech Lang Hear Res*, 55, 194–209.

Bradlow A.R. & Alexander J.A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am*, 121(4), 2339–2349.

Carreiras M., Perea M. & Grainger J. 1997. Effects of orthographic neighbourhood in visual word recognition: Cross-task comparisons. *J Exp Psychol Learn Mem Cogn*, 23(4), 857–871.

De Groot A.M.B., Borgwaldt S., Bos M. & Van den Eijnden E. 2002. Lexical decision and word naming in bilinguals: Language effects and task effects. *J Mem Lang*, 47, 91–124.

De Groot A.M.B. 2011. *Language and Cognition in Bilinguals and Multilinguals: An Introduction*, New York, Hove: Psychology Press.

Festen J.M. & Plomp R. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am*, 88, 1725–1736.

Goverts S.T., Huysmans E., Kramer S.E., De Groot A.M.B. & Houtgast T. 2011. On the use of the text reception threshold test and the distortion-sensitivity approach in examining the role of linguistic ability in speech understanding in noise. *J Speech Lang Hear Res*, 54(6), 1702–1708.

Grainger J. & Segui J. 1990. Neighbourhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Percept Psychophys*, 471, 191–198.

Grainger J. & Jacobs A.M. 1996. Orthographic processing in visual word recognition: A multiple read-out model. *Psychol Rev*, 103, 518–565.

Horn J.L. & Cattell R.B. 1967. Age differences in fluid and crystallized intelligence. *Acta Psychol*, 26, 107–129.

Houtgast T. & Festen J.M. 2008. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *Int J Audiol*, 47, 287–295.

Kaandorp M.W., Smits C., Merkus P., Goverts S.T. & Festen J.M. 2015. Assessing speech recognition abilities with digits in noise in cochlear implant and hearing-aid users. *Int J Audiol*, 54, 48–57

Luteijn F. & Barelds D.P.H. 2004. *GIT2 Groninger Intelligentie Test 2. Handleiding*. Amsterdam: Harcourt Test Publishers.

Lyxell B., Andersson J., Andersson U., Arlinger S., Bredberg G. et al. 1998. Phonological representation and speech understanding with cochlear implants in deafened adults. *Scand J Psychol*, 39, 175–179.

McRae K., Jared D. & Seidenberg S. 1990. On the roles of frequency and lexical access in word naming. *J Mem Lang*, 29, 43–65.

Nilsson M.J., 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. *J Acoust Soc Am*, 95, 1085–1099.

Pichora-Fuller M.K. 2003. Cognitive aging and auditory information processing. *Int J Audiol*, 42, 2S26–2S32.

Pichora-Fuller M.K. 2008. Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *Int J Audiol*, 48 (Suppl. 2), S72–S82.

Pichora-Fuller M.K. & Levitt H. 2012. Speech comprehension training and auditory and cognitive processing in older adults. *Am J Audiol*, 21, 351–357.

Plomp R. & Mimpen A.M. 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18, 43–52.

Ratcliff R., Comez P. & McKoon G. 2004. A diffusion model account of the lexical decision task. *Psychol Rev*, 111(1), 159–182.

Rönnberg J., Samuelsson E. & Borg E. 2000. Visual cognitive tests, central auditory function, and auditory communication. *Scand Audiol*, 29, 196–206.

Rönnberg J., Rudner M., Foo C. & Lunner T. 2008, Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol*, 47 (Suppl. 2), S99–S105.

Rubenstein H., Lewis S.S. & Rubenstein M.A. 1971. Evidence for phonemic recoding in visual word recognition. *J Verbal Learning Verbal Behav*, 10, 645–657.

Smits C., Goverts S.T. & Festen J.M. 2013. The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *J Acoust Soc Am*, 133(3), 1693–1706.

Smits C. & Festen J.M. 2013. The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. *J Acoust Soc Am*, 133(5), 3004–3015.

Tabossi P. & Laghi L. 1992. Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Mem Cogn*, 20, 315–328.

Van Der Maas H.L.J., Dolan C.V., Grasman R.P.P.P., Wicherts J.M., Huizenga H.M. et al. 2006. A dynamic model of general intelligence: The positive manifold of intelligence by Mutualism. *Psychol Rev*, 113, 842–861.

Versfeld N.J., Daalder L., Festen J.M. & Houtgast T. 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107, 1671–1684.

Van Rooij J.C.G.M. & Plomp R. 1990. Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *J Acoust Soc Am*, 88(6), 2611–2624.

Van Wijngaarden S.J., Steeneken H.J.M. & Houtgast T. 2002. Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am*, 111, 1906–1916.

Weiss D. & Dempsey J.J. 2008. Performance of bilingual speakers on the English and Spanish versions of the hearing in noise test (HINT). *J Am Acad Audiol*, 19, 5–17.

Wang X. & Humes L.E. 2010. Factors influencing recognition of interrupted speech. *J Acoust Soc Am*, 128(4), 2100–2111.