speech recognition in cochlear-implant users

outcome measures and the role of linguistic skills

Marre Kaandorp

Speech recognition in cochlear implant users

Outcome measures and the role of linguistic skills

Marre W. Kaandorp

This thesis was prepared within the Amsterdam Public Health research institute, Department of Otolaryngology – Head and Neck Surgery, section Ear & Hearing, VU University Medical Center, Amsterdam, the Netherlands

The study presented in the second chapter of this thesis was supported by Cochlear Ltd. Europe.

The author gratefully acknowledges the financial support for publication of this thesis by: Makker Hoortoestellen, Oticon Medical Benelux, Cochlear Benelux NV, Advanced Bionics Benelux, Phonak, Beter Horen B.V., Schoonenberg Hoorcomfort, Nederlandse Vereniging voor Audiologie.

MIX
Paper from responsible sources
FSC® C128610

Cover design: Marieke en Mark Delreux

Lay-out: Persoonlijkproefschrift.nl, Lara Leijtens

Printed by: Ipskamp Printing, Enschede

ISBN: 978-94-028-0915-2

VRIJE UNIVERSITEIT

Speech recognition in cochlear implant users

Outcome measures and the role of linguistic skills

ACADEMISCH PROEFSCHRIFT
ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 8 maart 2018 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Marre Willemijn Kaandorp
geboren te Amsterdam

promotoren:	prof.dr.ir. J.M. Festen
	prof.dr. A.M.B. de Groot
copromotoren:	dr. S.T. Goverts
	dr.ir. J.C.M. Smits

CONTENTS

LIST OF ABBREVIATIONS

aLDT	Auditory lexical-decision test
CVC	Consonant-vowel-consonant
DIN	Digits-in-noise
HA	Hearing aid
HE-N	Higher-educated native listeners
HE-NN	Higher-educated non-native listeners
ICC	Intraclass correlation coefficient
LDT	Lexical-decision test
LE-N	Lower-educated native listeners
NH	Normal hearing
RAU	Rationalized arcsine unit
RSpan	Reading-span test
RT	Reaction time
SD	Standard deviation
SEM	Standard error of measurement
SIN	Sentence-in-noise
SNR	Signal-to-noise ratio
SRT	Speech reception threshold
SRT_{diff}	Difference measure between SIN and DIN
TRT	Text reception threshold
vLDT	Visual lexical-decision test
VS	Vocabulary size
WN	Word-naming test

1

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a cochle t. A much better
that expla e range of perf
needed to impr reoperative co
rehabilitation pr s. This thesis aim
between clinic plicable linguis
performance wi chlear implant. T
skills tha n people w
of curren of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

General introduction

Cochlear implantation is a commonly used method to improve hearing abilities of severe to profoundly hearing-impaired persons. Because speech understanding is of major importance for participation in society, the main purpose of a cochlear implant (CI) in postlingually deafened adults is to restore auditory communication abilities. Although many of the CI users achieve reasonably good speech-recognition abilities in quiet, there is a large range of performance, especially in more complex listening situations. To improve pre-operative CI counselling and optimize user-specific rehabilitation programs after implantation, it is important to obtain a better understanding of the underlying factors that explain this wide range of speech-recognition performance. Several factors have already been the subject of investigation, but a major part of the variance in speech-recognition outcome in CI users can still not be explained (Blamey et al. 2013; Lazard et al. 2012). The studies described in this thesis aimed to further investigate possibly relevant factors, especially the role of linguistic abilities, in speech recognition of CI users.

Speech recognition is influenced by multiple factors that can be described in different categories: 1) auditory factors like the quality of the input signal and the listeners' auditory capacity, 2) linguistic/cognitive skills like working-memory capacity and verbal processing. Furthermore, biomedical factors like age at onset of hearing loss, are known to relate to speech-recognition performance. These biomedical factors likely affect the listeners' auditory capacity, but they might also influence central processing. Auditory factors (i.e., bottom-up factors such as reduced audibility and peripheral auditory processing) play a major role in explaining speech-recognition-in-noise performance in listeners with impaired hearing (Houtgast and Festen 2008). In the last two decades, however, an increasing interest has emerged in cognitive factors that influence speech understanding in challenging conditions. Various studies have shown the importance of top-down cognitive skills, like working-memory capacity, information processing, and phonological skills, in speech recognition in noise both in listeners with impaired and with normal hearing (e.g., Akeroyd 2008; Houtgast and Festen 2008; Pichora-Fuller 2003; Ronnberg et al. 2008; Besser et al. 2013). A review of Akeroyd (2008) showed that measures of working memory were most effective in these listeners (especially the reading span test; Daneman and Carpenter 1980). The relation of linguistic skills with speech recognition in noise, i.e. the level of language proficiency, has also been demonstrated in listeners with normal hearing (e.g., Bradlow and Alexander 2007; Goverts et al. 2011; Van Rooij and Plomp 1990; Van Wijngaarden, Steeneken, and Houtgast 2002; Benard, Mensink, and Baskent 2014).

For CI users it is well recognized that several biomedical factors play a role in speech-recognition performance: e.g., duration of severe to profound hearing loss before implantation (Blamey et al. 2013; Budenz et al. 2011; Holden et al. 2013; Mosnier et al. 2014; Roditi et al. 2009), the position of the electrodes, (e.g., Finley et al. 2008; Lazard et al. 2012), residual hearing and preoperative speech recognition (Lazard et al. 2012; Leung et al. 2005). Besides these biomedical CI-related factors, some studies have shown the importance of cognitive skills in speech recognition tasks in CI users (Heydebrand et al. 2007; Lyxell et al. 2003; Pisoni 2000). For instance, verbal learning (Heydebrand et al. 2007) and the ability to identify sentences that are visually masked in the text-reception-threshold test (TRT; Zekveld et al. 2007) were found to explain individual differences in speech recognition with a CI (Haumann et al. 2012). Recently, Rosemann et al. (2017) also suggested to consider the TRT test and a working-memory test, among other cognitive tests, when aiming to predict CI outcome. These tests significantly predicted recognition of vocoded speech in their study with 21 young adult normal-hearing (NH) listeners and 18 older NH listeners. Moreover, it has been demonstrated that linguistic factors play a role in speech recognition of CI-users as well (Heydebrand et al. 2007; Lyxell et al. 1996).

Although several studies support the assumption that linguistic factors and cognitive factors in general are relevant in CI speech-recognition performance, results vary among studies and do not point to tests that can be easily used in the clinic for pre-operative CI counselling or for evaluation of rehabilitation progress. Some of these studies only examined the influence of cognitive and linguistic factors on recognition of words and sentences in quiet, and that may be less cognitively and linguistically demanding than speech-in-noise recognition, which is closer to daily life speech understanding. The number of studies on the effect of cognitive and linguistic skills on CI performance, although increasing, is still limited. Therefore, the main question that is addressed in this thesis is: Is speech-recognition ability with a CI influenced by linguistic abilities and how strong are the effects compared to effects of other cognitive measures? To investigate these factors, we consider it first of all important to have insight in how CI outcome could be measured. Therefore we examine clinically relevant measures of speech recognition in quiet and in noise to evaluate outcome measures of CI performance. Secondly, clinically feasible measures of linguistic abilities are needed for pre-operative testing in CI candidates. To avoid interaction with auditory capacity, these tests are conducted in the visual domain. We first measure these abilities in NH listeners to provide a basis for studying the effect of linguistic skills in persons with hearing loss or CI's, where there might be an interaction between auditory and linguistic/cognitive factors. With this basis we explore the effects of these linguistic measures compared to the influence of two cognitive measures on speech recognition in CI users. Finally we examine measures of linguistic skills in the auditory domain and the relation with speech recognition in CI users, because they can provide more insight in the interaction between these linguistic skills and speech understanding as it occurs in daily life.

OUTCOME MEASURES OF SPEECH RECOGNITION IN CI USERS

Speech-recognition ability can be measured in several ways and it is important to understand advantages and disadvantages of those measures. Often phoneme scores or word scores of monosyllable words in quiet are measured. In addition, also sentence recognition in quiet is often used. Where word recognition with phoneme scores mainly represents how well a listener perceives language specific details of the input signal (bottom up), sentence recognition is assumed to demand more top-down processing and the use of linguistic knowledge. When sentences are presented in noise, these top-down demands increase even more (see e.g., Mattys et al. 2012).

The diversity in measures of CI outcome makes it difficult to compare results from different studies or centers. Multi-center studies frequently used percentile ranks for participants per center to compare participants because centers used different speech recognition tests at different presentation levels (e.g., Blamey et al. 2013). Recognition of phonemes or words in quiet is often used to define CI outcome instead of the more relevant speech-recognition abilities in noise, to determine auditory functioning in daily life listening situations.

In Dutch clinical practice speech-recognition performance is primarily assessed with monosyllables of the consonant-vowel-consonant type (CVC) in quiet, expressed in the percentage of correctly reproduced phonemes. In addition, speech-recognition performance in adverse conditions is usually assessed with the standard Dutch speech-in-noise test that uses short meaningful sentences (sentence-in-noise test, SIN test, Plomp and Mimpen 1979). This test involves an adaptive procedure to estimate the speech reception threshold (SRT) defined as the signal-to-noise ratio (SNR) where on average 50% of the sentences is repeated correctly. Similar sentence-in-noise tests exist in other languages (e.g., the HINT test, Nilsson, Soli, and Sullivan 1994; for a review see Theunissen, Swanepoel, and Hanekom 2009). Although the SIN test has been widely used in research and clinic, severe hearing impairment or limited linguistic skills (as in children or second language users) limit the applicability of this test. The Dutch digits-in-noise (DIN) test was developed as an alternative for the SIN test that measures the SRT for digit-triplets instead of sentences. It was intended to be applicable to a very wide range of listeners because it was developed to measure primarily bottom-up (auditory) speech recognition in noise. The validity of the DIN test was determined by Smits, Goverts, and Festen (2013) in NH listeners. We hypothesized that the DIN test is more applicable to CI users than the SIN test. Van Wieringen and Wouters (2008) demonstrated that a large range of hearing abilities, including CI users, could be mapped with a test that uses numbers in noise, the LINT (Leuven intelligibility number test). Recognition tests of monosyllable phonemes, sentences in noise and digits in noise are, thus, assumed to test different parts of speech recognition, each demanding a different amount of top-down processing.

MEASURES OF LINGUISTIC SKILLS

Linguistic skills comprise different domains, such as phonology, vocabulary, and syntax. To study the interaction between linguistic skills and speech-recognition ability it is essential to select clinically suitable and relevant measures of linguistic ability that can be applied in both the visual and auditory modality. Words, representing vocabulary, convey most of the meaning of a language utterance and are thus central to the process of language comprehension. For this reason, we consider the process of a word making contact with its representation in the mental lexicon, in psycholinguistics often referred to as 'lexical access' (e.g., Grainger and Jacobs 1996; Marslen-Wilson and Welsh 1978), to be an aspect of linguistic ability that is of major importance for speech recognition. In this process, the speed of word retrieval, which we will refer to in this thesis as *lexical-access ability*, could be important. We consider lexical-access ability to represent a *fluid* ability, the capacity of processing information and reasoning (Horn and Cattell 1967). There is evidence to suggest that efficient lexical access is related to better speech-recognition performance (e.g., Lyxell et al. 1996; Van Rooij and Plomp 1990; Larsby, Hällgren, and Lyxell 2012). In addition, the number of words stored in the mental lexicon (vocabulary knowledge) could also be important. Vocabulary knowledge is assumed to represent *crystallized* knowledge, a type of cognitive ability based on accumulated knowledge (Horn and Cattell 1967). The results of some studies suggest that better receptive vocabulary knowledge is related to better speech recognition of NH listeners in adverse conditions (e.g., Benard, Mensink, and Baskent 2014). However, a recently published study by Benichov et al. (2012) showed that verbal ability (crystallized knowledge) was not a predictor of word recognition in noise under conditions involving different levels of contextual probability, but hearing acuity and cognitive function (fluid abilities) were significant contributors to the variance obtained.

Lexical decision and word naming. In this thesis lexical-access ability is assessed by lexical-decision tests (LDT, because these are easy to use and can be applied both in the auditory (aLDT) and visual modality (vLDT). The LDT is essentially a discrimination task, in which participants have to decide quickly whether a test item is a word or a pseudoword by pressing a button that corresponds to the choice made. Reaction times (RTs) and accuracy are recorded for both word- and pseudoword responses. 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). Besides the time needed for the process of interest, lexical access, the LDT typically also encompasses the time needed for task-specific processes like pressing the button. For determining the role of lexical-access skills in speech recognition it is therefore preferable to obtain converging evidence from multiple tests (cf., Grainger & Jacobs, 1996), so that a more accurate estimate of lexical access can be obtained than using either test on its own. We therefore use a second lexical-access task, word-naming, which we keep simple, for potential use in the clinic. In this task, participants have to read words aloud as quickly as possible, thus, it requires pronouncing the word after recognizing it and does also not provide a pure measure of lexical-access ability. Combining the results of the lexical-decision (lexical access + discrimination + response) and word-naming (lexical access + speech production) test into a single composite measure might provide a more accurate estimate of lexical-access ability.

In order to describe some of our hypotheses and interpret the results, we use a simple modality-independent model for lexical decision. This model was based on established models of lexical access in spoken-word recognition (the Cohort Model by Marslen-Wilson 1987; Shortlist B by Norris and McQueen 2008; the TRACE model by McClelland and Elman 1986; e.g., the Neighbourhood Activation Model by Luce and Pisoni 1998) and of visual lexical access like the Multiple Read-out Model (Grainger and Jacobs 1996). It consists of the following stages:

1. *Activation*: With the onset of input, orthographically similar lexical elements (in visual lexical decision; e.g., mouse, house, mouth, muse, morse, …) or phonologically similar lexical elements (in auditory lexical decision; e.g., like, light, lime, …) are simultaneously activated and form a cohort.

2. *Cohort evaluation*: When the stimulus unfolds and more information becomes available, competitor words in the cohort are deactivated (inhibition).

3. *Decision*: When a word's level of activation reaches a critical difference from that of its competitors (Marslen-Wilson 2013) a word-response follows, or if a time deadline is passed before a match is found a pseudoword-response follows (e.g., Grainger and Jacobs 1996; Ratcliff, Gomez, and McKoon 2004).

Models of auditory and visual lexical access largely overlap. An important difference to keep in mind, however, is that in the auditory domain the stimulus is not presented and interpreted instantaneously, as in the visual modality, but sequentially.

Reaction times are influenced by several variables in aLDT and vLDT. For instance, word frequency (how often a word occurs during language use; in the present studies we use its frequency of occurrence per million words) is considered to influence activation levels (e.g., Luce and Pisoni 1998). Highly-frequent words are more strongly activated, have lower recognition thresholds, or higher baseline levels of activation, than words with a lower frequency of occurrence, resulting in shorter LDT reaction times. It was found that CI users use word-frequency information to identify spoken words similarly to the way NH listeners recognize spoken words (Meyer et al. 2003).

Reaction times are not only determined by test settings and word characteristics such as word frequency, but also by individual differences between the participants. For instance, different levels of proficiency in a language are likely associated with different baseline levels of activation in lexical representations. Furthermore, LDT reaction times (and accuracy) depend on the amount of so called 'spurious activation' in the mental lexicon, that is, the number of lexical candidates that, in addition to the element representing the input stimulus, are initially activated (i.e., the size of the cohort). It has been shown that the amount of spurious activation during auditory word recognition can be excessively large in nonnative speakers (Broersma 2007; Broersma and Cutler 2011). This finding is of particular relevance in the context of the present studies because it may be expected that for listeners with hearing loss relatively many phonemes are confusable, even in the native language. In other words, the amount of spurious activation during lexical access may be excessively large in people with hearing loss, resulting in relatively long response times and low accuracy.

EFFECTS OF LINGUISTIC AND COGNITIVE ABILITIES ON SPEECH RECOGNITION

Relations between visual measures of linguistic and cognitive abilities and speech recognition. We hypothesize that in vLDT longer word RTs are related to poorer language proficiency and that longer RTs in general are related to poorer speech recognition in noise. Accuracy is expected to be near ceiling in vLDT, when only highly familiar words are used. vLDT could be measured preoperatively to predict speech-recognition performance with a CI. For instance, Lyxell et al. (1996) found in a group of 11 CI candidates, that preoperatively measured verbal information processing speed and working-memory capacity were predictors of CI users' subjective levels of speech understanding. A better understanding of the relation between vLDT and speech recognition in groups of NH listeners that differ in language proficiency is needed for the interpretation of these results for listeners with hearing loss. The latter plausibly differ between them in linguistic abilities, where both auditory and non-auditory factors play a role.

Relations between auditory measures of linguistic abilities and recognition of degraded speech. Also in the auditory modality, lexical access has been subject of investigation for many years, however mostly with clear input signals. Efficient visual lexical access, however, does not necessarily imply efficient auditory lexical access. This holds especially for situations with degraded input, like CI use in everyday communication. A few recent studies used the visual world paradigm to investigate lexical access with degraded auditory stimuli by tracking eye-movements to pictures of target words and competitor words. They found that NH listeners adjust their strategy when there is more uncertainty in the signal (McQueen and Huettig 2012) and there are indications that CI users adapt their auditory lexical access to deal with degraded input (e.g., Farris-Trimble et al. 2014; McMurray et al. 2016). Knowing in what way they might have done so could be important for speech-recognition outcomes of an individual CI user.

Based on the model, we hypothesize that for speech stimuli that are degraded by a noise vocoder, compared to original stimuli, the RT for real words will be longer in NH listeners, because the process of activation of candidates and evaluation of competitors takes longer (e.g., Farris-Trimble et al. 2014) due to uncertainty in the signal and probably more spurious activation of lexical items in the cohort. We also expect the RT for pseudowords to be longer in the vocoded condition because the response deadline is usually set higher for more difficult conditions (e.g., Ratcliff, Gomez, and McKoon 2004). For CI users we expect a larger bias toward word responses, because they are used to guessing. These hypotheses are tested in Chapter 5.

OUTLINE OF THE THESIS

This thesis describes four studies that were conducted to achieve a better understanding of the role of linguistic skills, particularly of lexical-access ability, in speech-recognition performance of cochlear-implant users.

Chapter 2 describes a study that examines the suitability of clinical outcome measures of speech recognition in CI users. The feasibility, reliability and validity of the DIN test for measuring speech recognition in CI users and hearing-aid users is examined in relation to commonly used clinical tests of speech recognition in quiet and noise. Twelve NH participants, 24 hearing-aid users, and 24 CI users participated in the study.

Chapter 3 examines the relations of visual lexical-access ability and vocabulary size with different measures of speech-in-noise recognition in young adult NH listeners with various levels of linguistic abilities. To obtain a wide range of linguistic skills, three groups, of 24 participants each, were included: higher-educated native listeners, lower-educated native listeners, and higher-educated non-native listeners. The aim of the study is to evaluate suitable linguistic measures and examine their relation with speech recognition in noise. Additionally, the effect of keyword scoring in the SIN test is examined.

Chapter 4 presents a study that explored the influence of visually measured lexical-access ability compared to other cognitive factors on speech-recognition ability of 24 postlingually deafened CI users. The lexical-access measures and vocabulary-size measure of Chapter 3 are used in addition to two cognitive measures: working-memory capacity (Reading-Span test) and a visual analogue of the sentence-in-noise test (Text reception threshold test). The relations of these abilities with the clinical speech recognition measures that are also used in Chapters 2 and 3 are examined.

Chapter 5 discusses a study on the effect of degraded auditory input on auditory lexical-decision reaction times and accuracy, in Experiment 1 for 11 young NH listeners listening to noise-band vocoded stimuli and in Experiment 2 for 22 CI users for whom the CI itself produces a degraded input signal. Analyses show how CI users differ in this ability from NH listeners listening to degraded input. In addition, the relation of these auditory lexical-decision measures with speech recognition in quiet and in noise are examined.

Chapter 6 comprises a general discussion of the results of all studies reported in this thesis and provides suggestions for clinical implications.

2

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a co A much better
that expl this w nge of perf
needed to improve erative co
rehabilitation progr his thesis aim
between clinica plicable linguis
performance cochlear implant. T
skills tha people w
of curre of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

Assessing speech-recognition abilities with digits in noise in cochlear implant and hearing aid users

ABSTRACT

Objective: The primary objective of the study was to investigate the feasibility, reliability, and validity of the Dutch Digits In Noise (DIN) test for measuring speech recognition in hearing-aid and cochlear-implant users and compare results to the standard sentences-in-noise (SIN) test.

Design: The relation between speech reception thresholds for DIN test and SIN test was analysed to determine the validity of the DIN test. As linguistic skills were expected to make different contributions in these tests, their influence was analyzed.

Study sample: Participants were 12 normal-hearing listeners, 24 hearing aid users and 24 cochlear-implant users.

Results: The DIN test was feasible for more participants than the SIN test. Intraclass correlation coefficients showed high reliability. The standard error of measurement was smaller for the DIN test than for the SIN test. DIN test and SIN test were highly correlated ($r = 0.95$ and $r = 0.56$ for NH+HA and CI users respectively). In the regression analysis no significant contribution of basic linguistic skills or personal factors was found.

Conclusion: In the assessment of speech recognition in noise of aided hearing-impaired listeners with hearing aids or cochlear implants, the DIN test is a feasible, reliable and valid test.

Marre W. Kaandorp
Cas Smits
Paul Merkus
S. Theo Goverts
Joost M. Festen

International Journal of Audiology 2015; 54: 48-57

INTRODUCTION

Assessment of speech-recognition abilities in noise is important to determine the impact of hearing loss in daily life situations. Ideally, a speech-in-noise test should be applicable over a wide range of hearing losses to enable comparison of results between listeners with different capabilities. Within listeners, such a test is useful in diagnostics and in rehabilitation, e.g. over the course of rehabilitative training, for the evaluation of rehabilitation, and for optimizing fitting parameters of hearing aids or cochlear implants.

Smits, Goverts, and Festen (2013) recently developed a digit-triplet speech-in-noise test, the Digits In Noise (DIN) test. The DIN test was developed to serve as an alternative to the standard Dutch speech-in-noise test that uses short meaningful sentences (sentence-in-noise test, SIN test, Plomp and Mimpen 1979). Similar sentence-in-noise tests exist in other languages (e.g., Harvard IEEE sentences, the HINT test, Nilsson, Soli, and Sullivan 1994; Vaillancourt et al. 2005; Kollmeier and Wesselkamp 1997; for a review see Theunissen, Swanepoel de, and Hanekom 2009). Although the SIN test has been widely used in research and clinic, severe hearing impairment or limited linguistic skills (as in children or second language users) limit the applicability of this test. The DIN test has similarities with the digit-triplet screening test by telephone (Smits, Kapteyn, and Houtgast 2004) but was primarily developed as a clinical tool. In contrast to the digit-triplet screening test, the DIN test uses broadband signals and all the digits are included (the bi-syllabic digits 7 and 9 were omitted in the digit-triplet screening test). The digits were uttered separately by a male speaker and concatenated to form the triplets. In the digit-triplet screening test the triplets were uttered as a whole by a female speaker. Recently, in a group of normal-hearing (NH) subjects, it was pointed out that the DIN test measures primarily the auditory speech-recognition abilities, because the use of a closed-set paradigm with easy familiar words limits the effect of top-down processes like linguistic skills (Smits, Goverts, and Festen 2013). The DIN test should be feasible for a wide range of listeners with mild to severe hearing loss, cochlear implant (CI) users, and children. Smits, Goverts, and Festen (2013) evaluated the test for NH listeners. Repetitive testing with the DIN test showed a learning effect only between the very first test run and the second run, even with weeks between separate sessions of repetitive runs. Additionally the test was validated for university students with normal hearing and a wide range of simulated hearing losses to evaluate its outcome, while minimizing the effect of differences in top-down processing. The validity of the DIN test was determined by comparing the DIN test SRTs with those for the SIN test, yielding a high correlation coefficient of 0.90. Digits do not cover the complete phoneme distribution of daily-life speech, but Smits, Goverts, and Festen (2013) showed that, for the overall recognition of digits vowel, consonant, and word length recognition is needed. All in all there is substantial evidence that the DIN test is a valid test for measuring speech-recognition abilities in noise. Currently similar tests are being developed in other countries (e.g., Feeney et al. 2013). The study of Smits, Goverts, and Festen (2013) however does not show whether these promising results also apply to the intended listeners: patients with a wide range of hearing losses and varying cognitive and linguistic abilities.

The objectives of the current study were (1) to investigate the feasibility, reliability, and validity of the DIN test for measuring speech recognition in CI users and hearing aid (HA) users, by comparing results on the DIN test with results on the SIN test for these listeners, and (2) to explore whether differences between listeners in non-auditory aspects like basic linguistic skills explain differences in results between these two speech-in-noise tests.

Where the subjects in the study of Smits, Goverts, and Festen (2013) were supposed to be homogeneous with respect to age and top-down processing abilities, the participants in the current study were naturally more heterogeneous in that respect. It is known that sentences in noise are too difficult to recognize for many CI users and other patients with severe hearing loss. Top-down capacities (especially linguistic abilities) will vary in these patients and may even be affected by the severity and duration of the hearing loss. Linguistic abilities are expected to influence the recognition of highly redundant sentences more than the recognition of a closed set of digits. Knowing how DIN and SIN scores relate to linguistic abilities will be helpful for the selection of speech recognition tests in the evaluation of signal processing settings (mainly auditory) or of overall auditory functioning (most representative for auditory functioning in daily life). To gain a general idea of the influence of linguistic abilities on both speech-in-noise tests in listeners with mild to severe hearing loss, two linguistic tests based on vocabulary size were included in our test battery. We choose for receptive vocabulary size, often referred to as crystallized knowledge, because it is assumed to be a good predictor of overall linguistic skills.

In the present study we evaluated the use of the DIN test in the audiological assessment of aided listeners with a mild to severe hearing loss. Feasibility and reliability of the DIN and SIN test were compared by evaluating the ability to perform the test, the learning effect, standard error of measurement (SEM), and intraclass correlation coefficient (ICC) for all participant groups. The validity of the DIN test was assessed by analyzing the relation with a standard test, which is expected to measure essentially the same *auditory* speech recognition in noise ability (for Dutch the SIN test: Plomp and Mimpen 1979; Versfeld et al. 2000). Additionally, we investigated how linguistic abilities and personal characteristics (for instance age and educational level) influence the results of SIN and DIN test in these listeners.

MATERIALS AND METHODS

PARTICIPANTS

Three groups of listeners participated: young NH listeners (n=12), HA users (n=24) and CI users (n=24). All participants were over 18 years of age and had no relevant additional medical problems. Table 1 lists the characteristics of the three participant groups.

TABLE 1. Characteristics of the three participant groups. Group means, SDs, and range are shown.

	NH listeners		**HA users**		**CI users**	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age [years]	24 (2)	21 - 28	58 (16)	21 - 80	59 (14)	30 - 82
Age at onset of hearing loss [years]	-	-	35 (29)	0 - 80	25 (21)	0 - 63
Age at onset of severe hearing loss [years]	-	-	-	-	45 (22)	6 - 77
Duration of hearing loss* [years]	-	-	23 (18)	0 - 56	31 (16)	10 - 66
Duration of severe hearing loss* [years]	-	-	-	-	12 (14)	0 - 54
Pre-operative aided word score at 65 dB SPL [%]	-	-	-	-	32 (13)	0 - 50
Average pure-tone threshold (1, 2, 4 kHz) [dB HL]	-	-	59 (21)	36 - 107	-	-
CI or HA experience** [years]	-	-	12.8 (13.0)	0.3 - 36.7	3.1 (1.8)	1.1 - 6.6
Education***	7 (0)	7 - 7	5.1 (1.9)	1 - 7	3.6 (1.5)	2 - 7

*Years between self-reported onset of hearing loss and measurement date for HA users; years between self-reported onset of (severe) hearing loss and implantation date for CI users. Severe hearing loss was defined as aided CVC scores at 65 dB SPL of less than 50%.

**Years of HA experience for HA users; years of CI experience for CI users (preceding HA experience was not included for CI users).

***Education rated on a 7-point scale (1-primary school, 2-lower vocational education, 3-lower secondary general education, 4-intermediate vocational education, 5-higher secondary general education, 6-higher vocational education, 7-higher education/university).

NH participants, all university students, were primarily included to serve as control group of the setup of the speech-in-noise tests; sufficient data on NH subjects are available for all other tests. Normal-hearing was defined as pure-tone thresholds equal to or better than 20 dB HL at octave frequencies .25 - 8 kHz in both ears. All NH participants were native Dutch speakers.

HA users were patients of the ENT & Audiology department at VU University Medical Center (VUMC) in Amsterdam. They responded to an invitation to participate in the study during a visit to the audiology clinic, after having been successfully fitted with their first hearing aids or new hearing aids (various brands being used for this purpose). HA users with additional medical problems or prelingually deaf patients were not invited. All had at least two months experience with their hearing aids, but the majority had more than five years experience. During the tests they used their hearing aid with the program and volume setting they preferred for everyday use. All HA users were native Dutch speakers.

CI users were postlingually deaf (with onset of severe hearing impairment after the age of six years). They were patients of the ENT & Audiology department at VUMC, who responded to an e-mail invitation to participate in the study. CI users with additional medical problems or prelingually deaf patients were not invited. All had at least one year experience with their CI, either Cochlear (Sydney, Australia) (n=19), or Advanced Bionics (Valencia, USA) (n=5). During the tests they used their device with the program and volume setting they preferred for everyday use. All CI users were native Dutch speakers, with the exception of one bilingual CI user who was born in the Netherlands, but had English-speaking parents.

Participants provided informed consent before participating, they were reimbursed for travel costs, and received a fee of 7.50 euros per hour for their contribution to the study. The study was approved by the Medical Ethics Committee of VUMC.

SPEECH RECOGNITION MEASURES

Speech recognition was measured using monosyllables, sentences, and digit-triplets. Word lists (NVA; Bosman and Smoorenburg 1995) consisting of 12 meaningful monosyllables of the consonant-vowel-consonant (CVC) type were used to test the recognition of monosyllables. The CVCs were pronounced by a female speaker and presented in quiet at a fixed level. The first word was intended to enhance attention and auditory accommodation and is not included in the results. The test score was defined as the percentage of 33 phonemes that are correctly reproduced. This test is commonly used in Dutch clinical practice.

Sentence lists (VU98; Versfeld et al. 2000) consisting of 13 short, meaningful sentences 8 or 9 syllables in length, pronounced by a female speaker, were used to test sentence recognition. The lists were presented in quiet and in steady-state long term average speech spectrum (LTASS) masking noise (Versfeld et al. 2000). The score in quiet was the percentage of sentences that was correctly reproduced. The SRT in noise (defined as the SNR where on average 50% of the sentences was repeated correctly) was measured by the adaptive procedure described by Plomp and Mimpen (1979). The overall intensity level, speech and noise combined, was kept constant at 65 dBA and the SNR was varied in an adaptive procedure during the test. Each sentence started 1000 ms after the noise was started. Each noise track was 4000 ms. The following presentation started after the response was entered. The first sentence of each trial was presented at -4 dB SNR and was repeatedly presented with a 4 dB increase of SNR until the participant responded correctly or to a maximum intensity of +12 dB SNR to avoid very high SNRs at the start of the adaptive procedure. All subsequent sentences were presented only once with SNRs depending on the response to the previous sentence. After a correct response the SNR was lowered by 2 dB and after an incorrect response the SNR was raised by 2 dB. A response was considered correct only if the entire sentence was repeated without error. 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). The sentence-in-noise test with adaptive procedure is considered as the gold standard for speech-in-noise testing in Dutch clinical practice.

Digit-triplet lists (Smits, Goverts, and Festen 2013) containing 24 broadband, homogeneous digit-triplets were used to test digit recognition. The digits are pronounced by a male speaker and were presented both in quiet and in LTASS masking noise (DIN test; Smits, Goverts, and Festen 2013). The test score in quiet was the percentage of digit-triplets reproduced without error as a function of intensity level. In noise the same adaptive procedure was used for the DIN test as for the SIN test. Again, the overall intensity level was kept constant at 65 dBA and the first digit-triplet was presented at -4 dB SNR. The triplet started 500 ms after the noise was started. Noise offset was 500 ms after the end of the triplet. Both noise intervals were enlarged or reduced by a randomly chosen interval between +50 or -50 ms. The SRT was calculated by taking the average SNRs of triplets 5 to 25. All three digits had to be repeated correctly for the triplet to be qualified as correct.

VOCABULARY SIZE

Two different tests to measure receptive vocabulary size were used. In the Peabody-Picture-Vocabulary-Test-III-NL (PPVT; Schlichting 2005) each test item was presented both written and spoken by the experimenter. Participants were instructed to choose one out of four pictures that best matches the meaning of the word presented. We used the last 4 lists of the PPVT with 60 items in total. The second test was the vocabulary size subtest of the Groningen Intelligence Test-II (GITvs; Luteijn and Barelds 2004), which uses a list of 20 items. The participant was asked to choose the correct synonym out of 5 words for each visually presented test word. In both tests we used raw scores to permit direct comparison of participants.

GENERAL PROCEDURES

Tests were performed in a sound-treated booth by a trained experimenter. Speech tests were administered with a clinical audiometer (Decos Audiology Workstation, Decos Systems, Noordwijk, the Netherlands). Speech and noise were presented through a single loudspeaker (Yamaha MSP5 Studio). Participants were seated facing the loudspeaker at a distance of approximately 70 cm. The measurement protocol is shown in Table 2. Tests were presented in three blocks, in the same order for every participant, in one two-hour session. NH participants only performed the speech-in-noise tests (Block 3), because speech-recognition scores in quiet cited in the literature for young listeners with normal hearing are 100% or near 100% at all intensities used (Bosman and Smoorenburg 1995; Versfeld et al. 2000; Smits, Goverts, and Festen 2013).

TABLE 2. Measurement protocol. Each participant received the three test blocks in the same order, with material being presented in a balanced order within the blocks.

Block	Test order	Test	Practice list	Test list intensities
1 Speech in quiet	Test: counter-balanced Intensities: digram-balanced Latin square	A. CVC	65 dB SPL	35, 45, 55, 65, 65,75 dB SPL
		B. SIN	65 dBA	35, 45, 55, 65, 65,75 dBA
		C. DIN	65 dBA	35, 45, 55, 65, 65,75 dBA
Break				
2 Linguistic tests	Test: fixed order	D. PPVT		
		E. GITvs		
3 Speech in noise	Test: counter balanced	F. SIN	65 dBA	65, 65 dBA
		G. DIN	65 dBA	65, 65 dBA

The first block consisted of the speech tests in quiet. Speech test materials were presented in a counterbalanced order, using a 3x3 Latin square. Within each test six lists were presented at different levels preceded by a practice list. CVC lists were presented at 35 to 75 dB SPL[1] in 10 dB steps, with 65 dB SPL occurring twice. Sentences and digits were presented at 35 to 75 dBA in 10 dB steps, with 65 dBA occurring twice. Lists were presented in a fixed order and presentation levels were counterbalanced in accordance with a 6x6 digram-balanced Latin square (Wagenaar 1969), to avoid list effects and order effects in the average data. Per participant the same order of presentation levels was used for each speech test in this block. The second block consisted of the tests for vocabulary size, PPVT and GITvs, in a fixed order. In the third block the sentences and digits were presented in noise. Half of the participants were given the DIN test first, whereas the other half were given the SIN test first. Both speech-in-noise tests were conducted twice, preceded by a practice list.

In blocks 1 and 3, only one ear was tested. Four CI users who normally used a HA contralaterally were asked not to use the HA on the test day. The contralateral ear was not occluded in CI users. HA users and NH participants were tested on their preferred ear, the contralateral ear being occluded with a hearing protector foam earplug and a circumaural earmuff (Peltor Optime III).

RESULTS

SPEECH RECOGNITION IN QUIET

Figure 1 shows box plots of speech-recognition scores in quiet for CVCs, sentences, and digit-triplets for the hearing-impaired participants.

Each box represents a single measurement at a given level, except at 65 dBA or 65 dB SPL where the test-retest average was used. Results show a large variation in performance and better performance for the HA group than for the CI users for all speech materials. However, for digit-triplet recognition at 65 dBA, all hearing-impaired participants (HA and CI) had scores near 100% (range 92% to 100%). The range was much wider for recognition of CVCs and sentences at this intensity. Most CI users had a relatively poor performance on sentence recognition compared to recognition of CVCs in contrast to the

1 Standard clinical calibration protocols were used for all speech material. In the Netherlands, CVC speech material is calibrated in dB SPL, whereas sentences and digit triplets are calibrated in dBA.

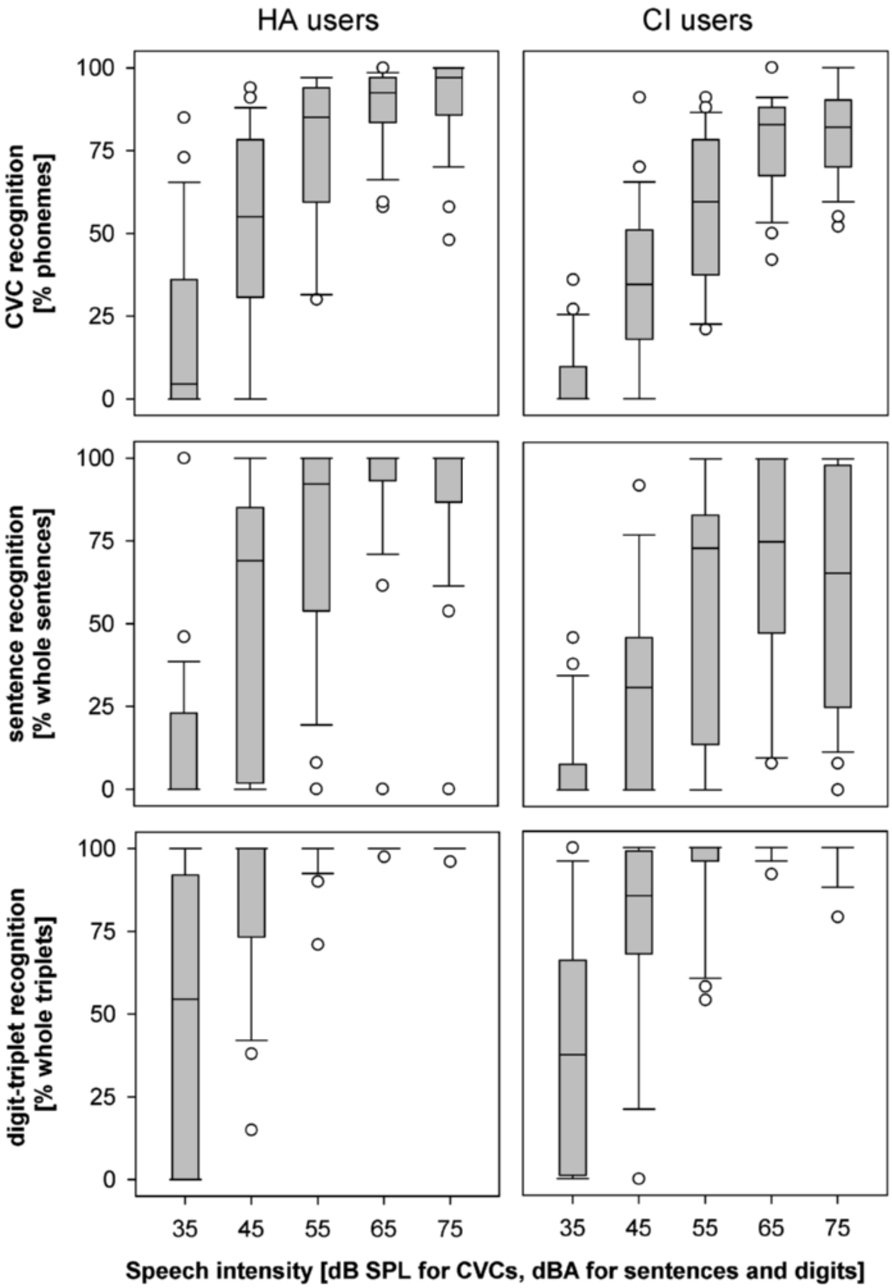


FIGURE 1. Box plots representing the distribution of speech-recognition scores in quiet as a function of intensity level for CVCs, sentences, and digit-triplets for 24 CI users (right panels) and 24 HA users (left panels). Subjects with normal hearing would be expected to achieve 100% scores for sentences and digits at all these intensities. Data at 65 dB SPL represent the average of two measurements per listener. At each intensity level the plot shows upper and lower quartiles (grey box), the median (crossbar), and the extremes of more than 1.5 times the inner quartile range (circles).

HA users. Inspection of individual psychometric curves (not shown) demonstrated that most participants (both CI and HA) reached their maximal speech discrimination scores at 65 dBA. For the sentence test the curves showed a drop in performance at 75 dBA for the CI users, presumably caused by the limited input dynamic range of the CI processor.

SPEECH RECOGNITION IN NOISE

SRTs and feasibility. Figure 2 shows SRTs (test versus retest) for the DIN and SIN test. It shows a large variance in SRTs for HA and CI users, which is also illustrated by the means and SDs in table 3. It also shows that all hearing-impaired participants had worse SRTs than the normal-hearing listeners. Six CI users and one HA user did not complete a the test run of the SIN test because they were unable to repeat any sentence correctly.

Inspection of the raw data showed that these seven participants had the lowest scores on sentences in quiet at 65 dBA. Thus, a group of 53 participants had results on all speech-in-noise tests (12 NH listeners, 23 HA users and 18 CI users). Test-retest differences were determined and were particularly marked for the SIN test at higher SRT values. For many hearing-impaired listeners the lower maximum performance in quiet and shallower psychometric curve result in a less steep slope at the 50% point. Thus, poor performance in quiet (see Figure 1) will lead to increasing uncertainty in SRT outcome, which can result in high test-retest differences. For these listeners the standard adaptive procedure (Plomp

TABLE 3. Mean SRTs in dB SNR, between subject standard deviations (SDs), intraclass correlation coefficient (ICC), and standard error of measurement (SEM) for sentences (SIN test) and digit-triplets (DIN test) in noise for participants with normal hearing (NH), hearing aid (HA), and cochlear implant (CI) users. All tests were performed at a fixed overall sound level of 65 dBA.

Test	Participants	Mean (SD) [dB SNR]		ICC		SEM [dB]	
		All data (n=53)	Data SRT<15 (n=48)	All data (n=53)	Data SRT<15 (n=48)	All data (n=53)	Data SRT <15 (n=48)
DIN test	NH	-9.3 (0.7)	-9.3 (0.7)	0.71	0.71	0.5	0.5
	HA	-4.4 (3.5)	-4.7 (3.3)	0.98	0.98	0.7	0.7
	CI	-1.8 (2.7)	-2.6 (2.3)	0.85	0.83	1.5	1.1
	All	-4.6 (4.0)	-5.3 (3.6)	0.97	0.97	1.1	0.9
SIN test	NH	-4.2 (0.8)	-4.2 (0.8)	0.68	0.68	0.6	0.6
	HA	2.1 (4.8)	1.4 (3.7)	0.85	0.93	2.5	1.3
	CI	8.0 (6.1)	5.1 (2.8)	0.90	0.91	2.7	1.2
	All	2.7 (6.6)	1.1 (4.5)	0.94	0.97	2.3	1.2

The ICC for average measures was calculated with a two way random model with absolute agreement.

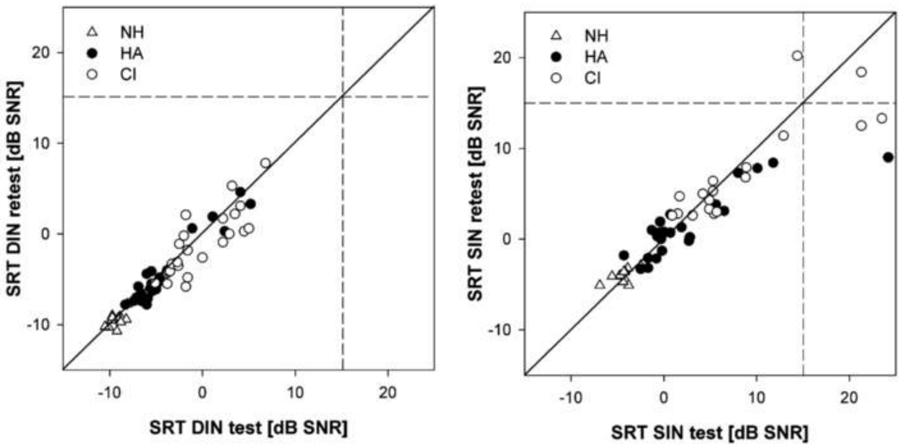


FIGURE 2. Test versus retest scores of digits in noise (left panel) and sentences in noise (right panel) for NH participants (triangles), HA users (filled circles) and CI users (open circles). The solid line represents equal values. The dashed lines represent the SNR +15 boundary.

and Mimpen 1979) will more often result in a higher SRT than the actual SRT (Smits and Festen 2011). There is no commonly used criterion to detect unreliable SRTs. In the speech intelligibility index (SII, ANSI 1997), it is assumed that all speech information is available when the SNR in steady-state LTASS noise is +15 or more. Thus, in theory SRTs of 15 dB SNR or higher do not reflect the ability to recognize speech in noise, and these SRTs should be classified as unreliable because the adaptive procedure does not work properly. The SII is useful for NH and hearing-impaired listeners, but the upper limit of +15 dB SNR may not be valid as the upper limit for signals processed by hearing aids or cochlear implants. For HA and CI users the upper limit could be different. Nevertheless, we considered SRTs of 15 dB SNR or higher as unreliable in this study. Excluding participants with SRTs of 15 dB SNR or higher yielded a group of 48 participants of whom complete and reliable data were available (12 NH listeners, 22 HA users and 14 CI users).

Reliability. Reliability was assessed by the ICC with a two-way random model with absolute agreement (McGraw and Wong 1996). To determine the effect of including unreliable SRTs, calculations were performed for (1) all participants with results on both DIN and SIN test, including unreliable SRTs (n=53); and (2) the subset of 48 participants with reliable SRTs. The ICC's (Table 3) were high (0.94-0.97) for the entire group (1) for both DIN test and SIN test. ICC's for NH participants were lower (approximately 0.7), which is a result of the homogeneity of this group and consequently minimal variance in SRTs. Hence, reliability of both DIN test and SIN test was high for all participants (a value of 0.7 or more for the ICC is considered to reflect good reliability, Nunnally and Bernstein 1994). This indicates that DIN and SIN are equally able to distinguish patients from each other (de Vet et al. 2006).

The SEM (Table 3) reflects the agreement between two measures and was calculated as the standard deviation of the mean differences between two measurements divided by √2 (de Vet et al. 2006; Plomp and Mimpen 1979). The impact of the 5 unreliable SRTs on the SIN test results is reflected by the difference of more than 1 dB between the SEM values of the entire group of participants (1) and the subset of participants (2) (see Table 3). Therefore, in the following analyses the group of 48 participants was used. For NH the SEM for the DIN test is comparable to the value of 0.7 dB Smits et al. (2013) found. Even for the subset of participants without unreliable SRTs (2), the agreement between two measures (SEM) was better for the DIN test than for the SIN test. Together the ICC and SEM values reflect a good reproducibility for both tests for all participant groups.

Learning effect. To investigate a possible learning effect, data (n=48) were at first analysed with a mixed repeated-measures analysis of variance, with Test Number (test, retest) and Speech Material (DIN, SIN) as within subject variables and Participant Group (NH, HA, CI) as between subjects factor. Since the first list was a practice list, it was not included in the analysis. As expected, a major effect was found for Speech Material, $F(1, 45)=638.06$, $p<0.001$ as well as an interaction between Speech Material and Participant Group $F(2, 45)=8.36$, $p<0.001$. There was also a major effect of Test Number, $F(1, 45) = 4.62$, $p = 0.037$, but no interaction between Test Number and Speech Material $F(1, 45) = 1.39$, $p = 0.244$ or between Test Number and Participant Group, $F(2, 45) = 2.35$, $p = 0.107$. Hence, we conclude that, despite the practice list, there was for both DIN and SIN test still a small learning effect between test and retest. This effect of 0.3 dB was not statistically different between DIN and SIN test or between participant groups. The size of the learning effect was small compared to the SRT increment and may be considered irrelevant in many situations. Without prior practice with the speech material in quiet the learning effect and measurement error could be somewhat higher.

Validity: relation between DIN test and SIN test. The validity of the DIN test was assessed by analyzing the relation between DIN and SIN tests (Figure 3). The right-hand panel shows DIN test results (mean: 2.0, SD: 2.4) for participants who were unable to do the SIN test or had unreliable SIN results. Stepwise linear regression was used to investigate whether the relation between test results is different for the participant groups. Because not all

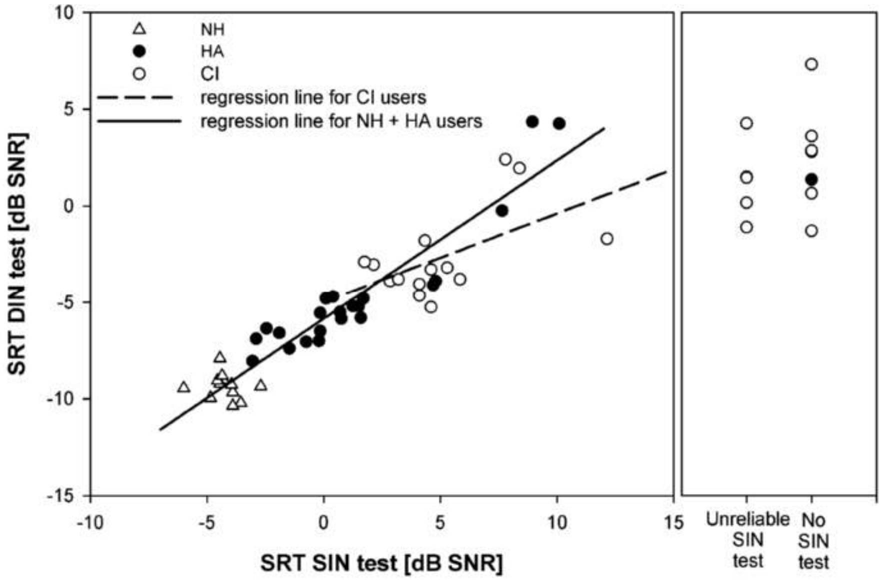


FIGURE 3. Relation between digit-triplet SRTs in noise (DIN) and sentence SRTs in noise (SIN). Shown SRTs are the average of test and retest scores. The solid line represents the linear regression line for NH and HA users. The dashed line represents the regression line for CI users. The right panel shows DIN results for participants who were unable to perform the SIN test or had unreliable SIN results.

participants were able to perform the SIN test and some SIN results were considered unreliable, the DIN test was chosen as dependent variable and SIN test as independent variable to avoid a sample selection bias on the dependent variable. Dichotomous dummy variables for participant group (CI-use and HA-use), with NH as reference group, were added to the model to account for interaction effects. Regression analysis showed significant interaction for CI-use with the SIN test (t = -2.363, p<0.023). There was no interaction for HA-use with the SIN test (t = 1.676, p=0.101). DIN scores are predicted reasonably well by the following model: DIN = -5.85 + 0.82*SIN for NH and HA users, and DIN = -4.98 + 0.46*SIN for CI users (R^2 = 0.86, p<0.001). The regression line for CI users was less steep than that for NH and HA users. Pearson's correlations were computed for the NH+HA users and CI users separately (r = 0.95 and r = 0.56 respectively).

EFFECTS OF PERSONAL CHARACTERISTICS ON SPEECH-IN-NOISE PERFORMANCE

For ten CI users and two HA users the SIN test was too difficult or unreliable. Therefore we examined possible predictor variables for the ability to perform the SIN test in all hearing-impaired participants (24 CI users and 24 HA users).

Mann-Whitney U tests were carried out on independent samples, shown in Table 4, to evaluate possible differences in personal characteristics that can predict the ability to perform the SIN test. Next to age, these were self-reported variables: duration of hearing impairment, age at onset of hearing impairment, years of CI or HA experience, and education level. All were not significantly different between the group where the SIN test was feasible and that where it was not.

Next, we looked at differences in performance on the linguistic tests and the CVC test between the groups, since these are aspects that should be taken into account in clinical practice. PPVT raw scores showed a normal distribution and ranged from 4 to 36 errors (mean = 14.9 errors, SD = 7.7 errors) for the HA users, and 7 to 35 errors (mean = 19.0 errors, SD = 8.3 errors) for the CI users. GITvs raw scores showed a normal distribution and ranged from 0 to 12 errors (mean = 5.5 errors, SD = 3.6 errors) for the HA users and 2 to 13 errors (mean = 6.7 errors, SD 2.9 errors) for the CI users. A composite measure "Vocabulary Size" was created for the purposes of further analysis by converting both test results to *z*-scores and averaging them. The composite score showed good reliability (Cronbach's alpha = 0.887). Vocabulary Size was not statistically different (p = 0.07) between the group that was able to perform the SIN test and the group that was not. Aided recognition of CVCs in quiet at 65 dB SPL was significantly lower (p < 0.001) for the group unable to perform the SIN test.

2

TABLE 4. Comparison of hearing-impaired participants who were able to perform the SIN test and those who were unable to perform this test. Group means, SDs, and Mann-Whitney U test of the personal characteristics and test outcomes are shown. Group differences are considered significant if p<0.05.

	Personal characteristics and test scores	**SIN feasible (n = 36)**	**SIN not feasible (n = 12)**		
		Mean (SD)	Mean (SD)	z	p
Personal characteristics	Age [years]	59.8 (15.7)	54.5 (13.0)	-1.34	0.18
	CI or HA experience [years]	8.7 (11.1)	5.6 (8.0)	-0.07	0.94
	Duration of hearing-impairment [years]	26.0 (17.1)	31.0 (18.4)	-0.89	0.37
	Age at onset hearing-impairment [years]	33.3 (26.4)	20.8 (20.2)	-1.30	0.19
	Education	4.6 (1.9)	3.7 (1.8)	-1.45	0.15
Test scores	CVC (at 65 dB SPL) [% phonemes]	89.5 (7.1)	63.1 (11.5)	-4.98	**<0.001**
	Vocabulary size [z-score]	-0.19 (0.8)	0.52 (1.1)	-1.79	**0.07**

Forward stepwise logistic regression analysis was used to evaluate the predictive value of aided CVC score at 65 dB SPL on the ability to perform the SIN test. We decided to also include Vocabulary Size in the analysis, because the Mann-Whitney U test result showed a *p*-value of 0.07 in this relatively small study sample for this variable which suggests a possible influence. The regression result showed CVC score to be the sole predictor of the ability to perform the SIN test (model $\chi^2 = 39.28$, $p < 0.001$, Nagelkerke $R^2 = 0.83$) in this study group. Addition of Vocabulary Size did not significantly improve the model. The odds ratio for CVC score is 0.73 with a 95% interval of [0.60 – 0.90]. This suggests that a decrease in the aided CVC score at 65 dB SPL reduces the likelihood that the participant will be able to perform the SIN test.

We were also interested in the contribution of the two above-mentioned variables to the relation between the DIN test and the SIN test. As mentioned in the introduction, we expected linguistic skills to have a larger influence on SIN scores than on DIN scores. We therefore examined whether these variables would improve our linear regression model for prediction of DIN test scores. For this purpose we entered Vocabulary Size and CVC scores in a second block, next to SIN scores and 'CI-use' in the linear regression analysis. The regression run with these variables showed that none of the variables significantly improved the model ($p > 0.1$). This indicates that the residual variance in this study group cannot be explained by differences in vocabulary size or CVC score at 65 dB SPL.

DISCUSSION

The primary objective of this study was to investigate the feasibility, reliability, and validity of the DIN test for hearing-impaired participants with either cochlear implants or hearing aids and compare the results with results of the standard SIN test.

Most participants, even those with a severe hearing loss, achieved nearly perfect digit-triplet scores in quiet at 65 dBA. The range in scores for sentence recognition in quiet was much wider. High speech discrimination performance in quiet is needed for reliable speech-in-noise testing in tests with an adaptive procedure. Seven participants with low sentence recognition in quiet (<50%) were unable to do the SIN test. Another five participants had one or two SRTs of more than 15 dB SNR, that were considered unreliable. Hence, the DIN test is applicable over a wider range of hearing acuities than the SIN test, where good sentence recognition in quiet is not always evident - especially for CI users. The DIN test appears to be feasible for almost all CI and HA users.

The low SEM and the high ICC for the DIN test indicate a high level of reliability for this test. Jansen et al. (2012) showed a comparable pattern of SEM scores for normal-hearing and hearing-impaired listeners. They compared the French Intelligibility Sentence test (FIST) with the French digit-triplet test (FrDigit3) and report for normal-hearing listeners a SEM of 0.4 and 1.2 dB for FrDigit3 and FIST respectively, and 1.1 and 1.3 dB for hearing-impaired listeners with mild to moderate hearing loss. The DIN test seems to show less variability between different listeners. However, because SEM was higher for the SIN test for HA and CI users than for the DIN test, the ICC was equal for both tests, indicating equal ability to distinguish between individual participants (de Vet et al. 2006). The lower SEM, reflecting the agreement between two measures (de Vet et al. 2006) makes the DIN test more suitable for comparison of different conditions like settings of the CI or HA, bimodal versus CI only situations, or unilateral versus bilateral conditions.

To validate the DIN test we compared it to the standard SIN test, which is widely used but mainly for listeners with normal hearing or mild to moderate hearing loss. The present study showed that the SIN test was too difficult or SIN results were unreliable for some of the CI users and HA users participating. A comparable test that is applicable to all participants and could be used as standard for validation is not available. We found two linear relations between DIN test and SIN test scores, i.e. for the CI users and for the NH + HA users. The regression line was shallower for CI users than for NH + HA users: 1-dB change in SIN threshold resulted in a smaller change in DIN threshold for CI users than for the other two groups. This effect of participant group on the relation between DIN test and SIN test could be caused by the difference in signal processing in hearing aid and cochlear implants, but also by the difference in hearing loss or other personal characteristics in the HA and CI group. Additional research might shed more light on this.

We found good correlations between the results obtained with the DIN test and those obtained with the well-established SIN test, suggesting that the DIN test may be able to play an important role in the measurement of the ability of aided hearing-impaired listeners to recognize speech in noise. This is supported by the correlations found in literature (Smits, Goverts, and Festen 2013; Jansen et al. 2012). Hence, the relation between DIN test and SIN test showed that the DIN test measures approximately the same ability as the SIN test and therefore is a valid test to measure speech recognition in noise. For participants that were unable to do the SIN test, the DIN test is of special additional value, because using the DIN test makes it possible to assess speech recognition in noise, also for follow-up in rehabilitation or comparison of HA or CI settings.

EVALUATION OF EXTREME SRT VALUES

Many studies have used the adaptive SRT procedure to measure the ability to recognize speech in noise, but it is not always clear when SRT results are considered unreliable. We found large test-retest differences for some participants indicating that these test results might be unreliable. SRTs of +15 dB SNR or higher were considered unreliable because, for steady-state LTASS noise for those SNR, all speech information should in theory be available (SII model) and these SRTs do not represent speech-in-noise recognition. Although we are aware of the fact that for HA or CI users the criterion of 15 dB SNR may be too strict. In other studies for instance a minimal speech-recognition performance (e.g. 70%) in quiet is required to avoid unreliable values. In the present study standard test procedures were used. The influence of the SRTs that we considered unreliable was especially visible in the difference (1.2 dB or more) in SEM values for hearing-impaired listeners (Table 3). Hence, there is lack of agreement on the criterion to judge whether an individual result for a speech-in-noise test with an adaptive procedure is reliable. Such a well-defined criterion is needed to avoid a selection bias in the results. This problem is much less present in the DIN test.

INFLUENCE OF LINGUISTIC AND PERSONAL FACTORS

Although the SIN test is assumed to call more heavily upon linguistic and cognitive skills than the DIN test, we did not find evidence in this exploratory study that vocabulary size and education level play a major role in the difference between sentence recognition and digit-triplet recognition in noise in this population. Also other personal characteristics were not significantly different between groups of participants that were able to do the SIN test and that were not in this study. However, the group that is not able to perform the SIN test was small (n=12). In a larger study sample significant differences might be found.

Nevertheless, participants who were unable to perform the SIN test did have smaller vocabulary (near significant) than other participants. Logistic regression did not reveal, however, that this variable was valuable as a predictor of the ability to perform the SIN test in this small study sample. Further research is needed to shed more light on the role of linguistic abilities in speech recognition of listeners with hearing impairment, especially CI users. A recently published study by Benichov et al. (2012) showed that verbal ability (crystallized knowledge) was not a predictor of word recognition in noise under conditions involving different levels of contextual probability, but hearing acuity and cognitive function (fluid abilities) were significant contributors to the variance obtained. In future studies we will also explore other, more process-related, measures of linguistic abilities and their role in speech recognition. We conclude that the DIN test measures approximately a similar ability as the SIN test.

The logistic regression did show that CVC score at 65 dB SPL (which is, in the clinic, usually known for hearing-impaired listeners) was a good predictor of the ability to perform the SIN test. Smits, Goverts, and Festen (2013) found near perfect recognition of spectrally smeared and low-pass filtered digit-triplets in quiet at 65 dBA, whereas recognition scores for CVCs at 65 dBA in this condition were only approximately 40%. For the more difficult filtered signal conditions the performance of digit-triplets in quiet also dropped. Hearing-impaired listeners in the current study all had CVC scores of 42% or higher in quiet at 65 dB SPL. These results confirm that the DIN test is at least applicable for persons with CVC scores in quiet of 40% or more at 65 dB SPL.

COMPARISON TO OTHER TESTS

It is already well known that severe hearing impairment or limited linguistic skills limit the clinical applicability of testing speech recognition using sentences. Other studies have focused on tests and materials for quantifying speech recognition in severely hearing-impaired listeners. Van Wieringen and Wouters (2008) developed the Leuven Intelligibility Number test (LINT) and Sentence test (LIST) that use sentences with a lower speech rate. Some studies use closed set materials (e.g., Matrix test, Hagerman 1982; Wagener, Brand, and Kollmeier 1999; Zokoll et al. 2013), keyword scoring or an adaptation of scoring methods (Wong and Keung 2013). The DIN test is intended to be applicable to a very wide range of listeners and to serve as an alternative to the sentence-in-noise test. The DIN test might be more suitable for follow-up of patients or comparison of different HA or CI settings, because it can be repeated many times (content learning is less likely to occur), and the results are less influenced by top-down processing. The Matrix test has similarities to the DIN test (closed set of stimuli, less possibility of content learning, and less influence of linguistic factors). In contrast to the SIN test, the Matrix test uses sentences of five words that are semantically unpredictable (Zokoll et al. 2013). However, linguistic knowledge is required on the level of word form. Where the DIN test was developed to measure primarily bottom-up (auditory) speech recognition in noise, the Matrix test may require some down processing. The specific properties of the DIN test make it applicable to even a broader group of listeners (worse hearing losses and young children). Although digits do not cover the complete range of phonemes in daily life speech, Smits, Goverts, and Festen (2013) showed that not only vowel but also consonant recognition is important in this test. The use of digits makes it easy to translate the test to other languages, and therefore might enable comparison of large groups of listeners.

CONCLUSIONS

The results of this study demonstrate that the DIN test is a feasible, reliable and valid test for measuring speech recognition in noise in HA and CI users with a wide range of hearing losses. The validity of the DIN test is demonstrated by the high correlation with the SIN test, though the relation between DIN and SIN scores is slightly different for CI users than for HA users and NH listeners. Vocabulary size was not significantly related to the difference between SIN or DIN results. The DIN test can be repeated many times and is not time consuming, which makes it suitable for the follow-up of HA and CI users.

Declaration of Interest: The authors would like to thank Cochlear Europe LTD for their financial support for this study.

3

Many postlingually deafened individuals

reasonably good speech-recognition

adverse conditions there still is a large

with a co much better

that expl this ge of per

needed to improve erative co

rehabilitation his thesis aim

between clinically ble linguis

performance with a co implant. T

skills th n people w

of current easures of speec

speech-recognition in CI users was inves

speech recognition performance of

examined in addition to other auditory

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

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: high-educated native, lower-educated native, and high-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.

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

International Journal of Audiology 2016; 55: 157-167.

INTRODUCTION

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 and 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., Houtgast and Festen 2008; Akeroyd 2008; Pichora-Fuller 2003; Ronnberg 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., Bradlow and Alexander 2007; Goverts et al. 2011; Van Rooij and Plomp 1990; Van Wijngaarden, Steeneken, and Houtgast 2002; Weiss and Dempsey 2008). For instance, Van Wijngaarden, Steeneken, and Houtgast (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.

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 and 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 and 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 and Cattell 1967). Benichov et al. (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 et al. (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 and 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., wordnaming, Balota and Chumbley 1985; perceptual identification, Grainger and Segui 1990; lexical decision, Rubenstein, Lewis, and Rubenstein 1971). 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. Lyxell et al. 1998; Ronnberg, Samuelsson, and Borg 2000; Van Rooij and Plomp 1990). 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 and 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, Jared, and Seidenberg (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 and 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 high-educated non-native listeners have non-verbal fluid abilities (e.g. processing speed, working memory) comparable to those of the high-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, Goverts, and Festen 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 and Plomp 1990). In the Netherlands performance of sentence recognition in noise is commonly evaluated with the SRT test (Plomp and 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, Soli, and Sullivan 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 high education (HE-N), 24 native Dutch listeners with lower education (LE-N), and 24 non-native listeners with high 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 and 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 8 or 9 syllables in length. . They were presented in LTASS masking noise or in fluctuating noise (Festen and 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 (SRTstat)1 and -15 dB SNR for sentences in fluctuating noise (SRTfluc) 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 and 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, Goverts, and Festen 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 and Barelds 2004), which uses a list of 20 items. Participants had to choose the correct synonym out of 5 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.

---

1 In this study the abbreviation SRT_{stat} was used for the sentence-in-noise test in stationary noise to discriminate from the test in fluctuating noise (SRT_{fluc}). The SRT_{stat} is the same test as the SIN test that was used in the other studies described in this thesis.

For the LDT test (Rubenstein, Lewis, and Rubenstein 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, Piepenbrock, and Van Rijn 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, Perea, and Grainger 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 and 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.

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

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

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

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, Goverts, and Festen 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 χ^2 = 51.8, df = 2, p < 0.001) and for SRT-type (Wald χ^2 = 1042.0, df = 2, p < 0.001) and an interaction for Group × SRT-type (Wald χ^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 dB 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.

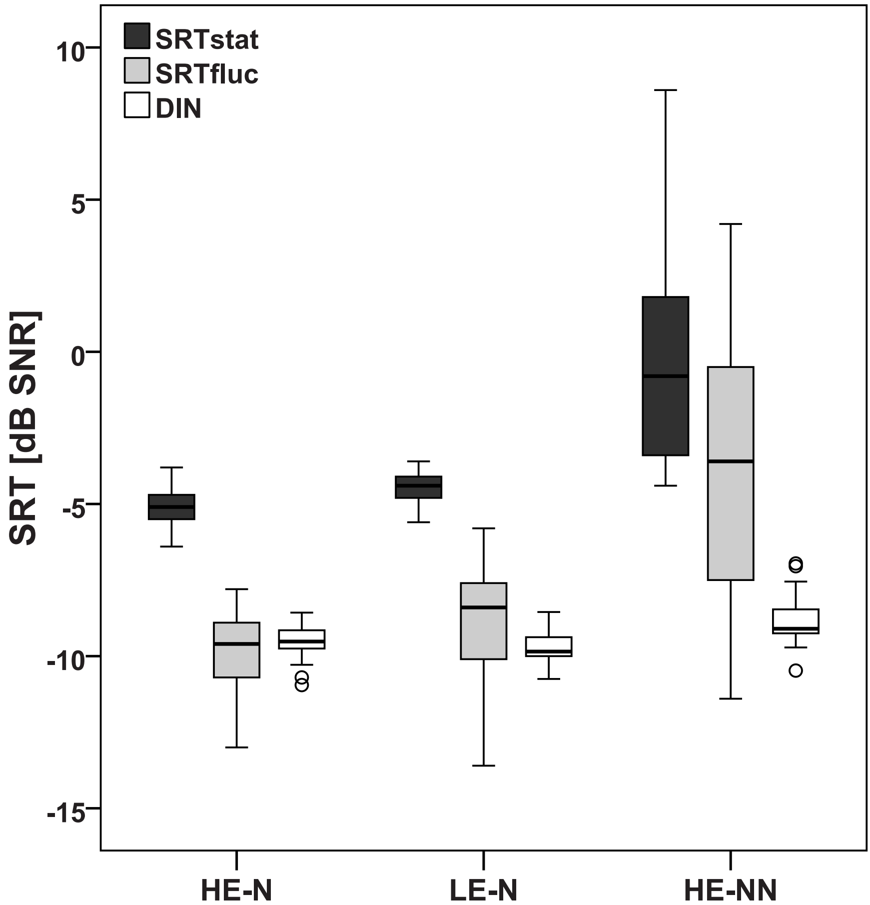


FIGURE 1. Boxplots of speech-reception thresholds for sentences in stationary noise (SRT_{stat}), sentences in fluctuating noise (SRT_{fluc}), and digits in noise (DIN) for three participant groups: High Educated Native (HE-N), Lower Educated Native (LE-N), and High Educated Non-native (HE-NN).

LINGUISTIC MEASURES

Lexical-decision test setup. First we evaluated whether the current LDT test 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

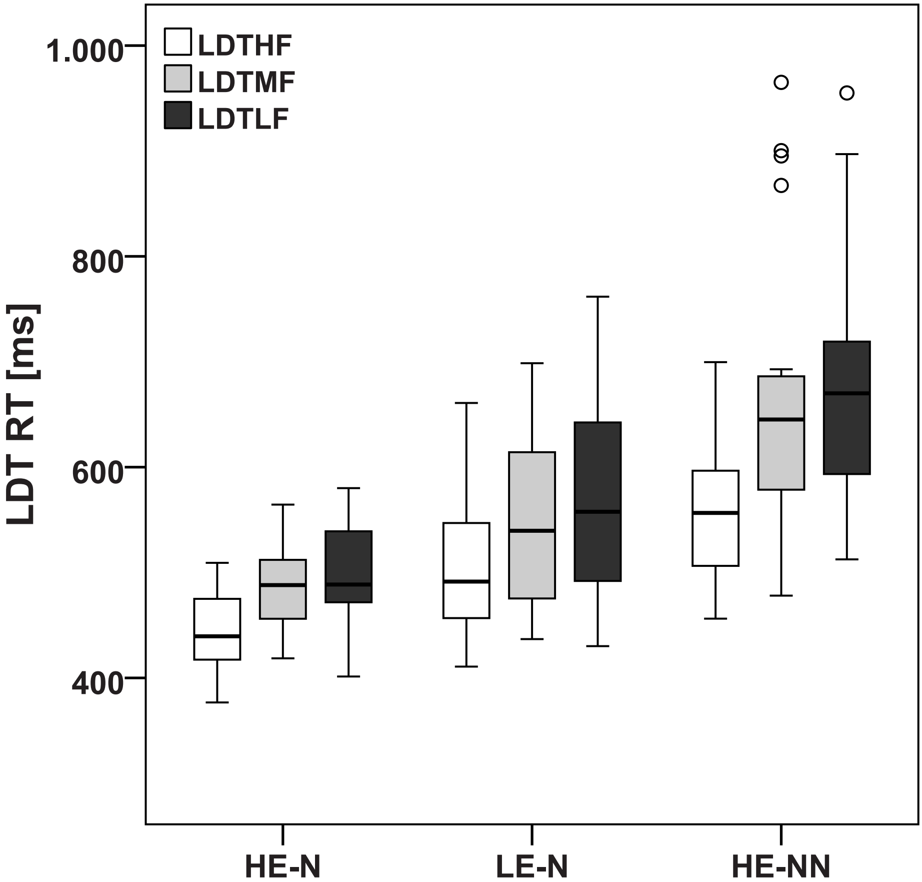


FIGURE 2. Boxplots for Lexical-Decision Test (LDT) results for three lists with different average word frequencies: high frequent (LDT_{HF}), middle frequent (LDT_{MF}), and low frequent (LDT_{LF}). For each frequency the mean of two lists was used. Results are shown for three groups of participants: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

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, $p < 0.001$).

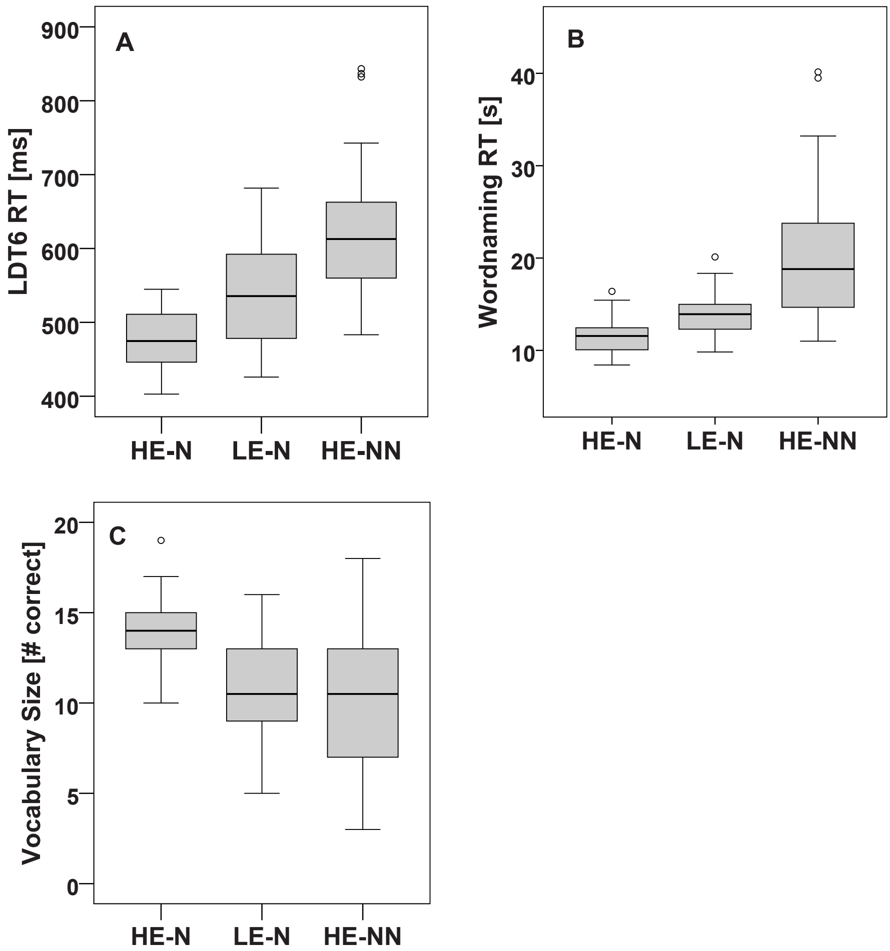


FIGURE 3. Boxplots for (A) Lexical-decision RTs (LDT_6), (B) Wordnaming RTs and (C) Vocabulary Size scores for the three groups of participants: High Educated Native (HE-N), Lower Educated Native (LE-N), and High Educated Non-native (HE-NN).

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 Behavior 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 6 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 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.

3

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 3A. 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 3B. 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

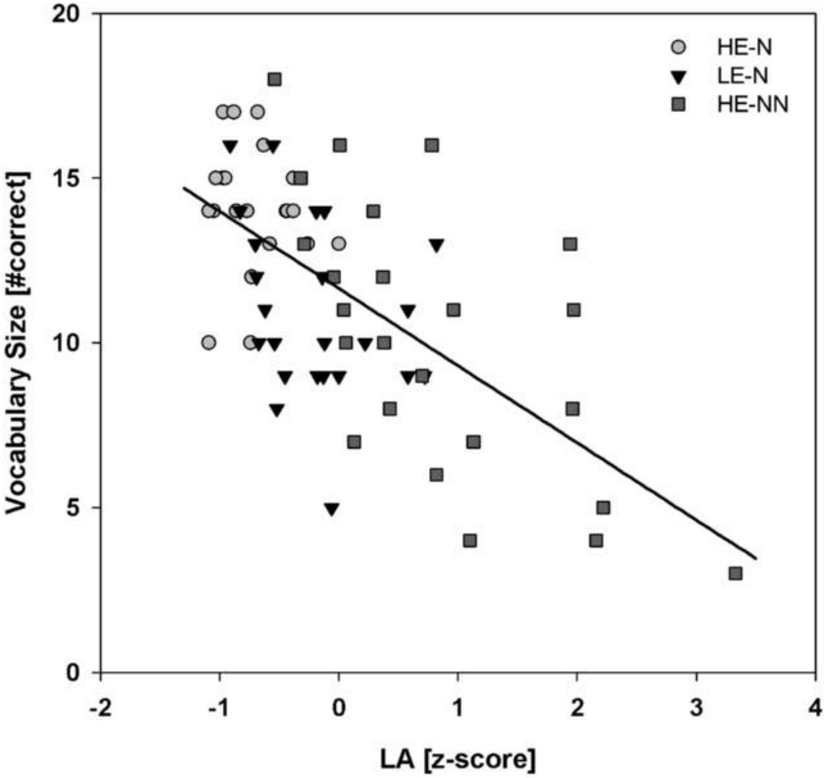


FIGURE 4. Relation between Lexical Access (LA) and Vocabulary Size for three groups of participants: High Educated Native (HE-N), Lower Educated Native (LE-N), and High Educated Non-native (HE-NN).

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 high and lower education ($p = 0.43$).

Vocabulary Size scores, shown in Figure 3C, 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. Grainger and Jacobs 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.

TABLE 2 Prediction of SRTs by mean lexical-decision score (LDT_6), Wordnaming, Vocabulary Size, and the combined variable Lexical Access.

	SRT_{stat}		SRT_{fluc}		DIN	
Predictor	Adj. R^2	*p*	Adj. R^2	*p*	Adj. R^2	*p*
Linear regression						
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.

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 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.

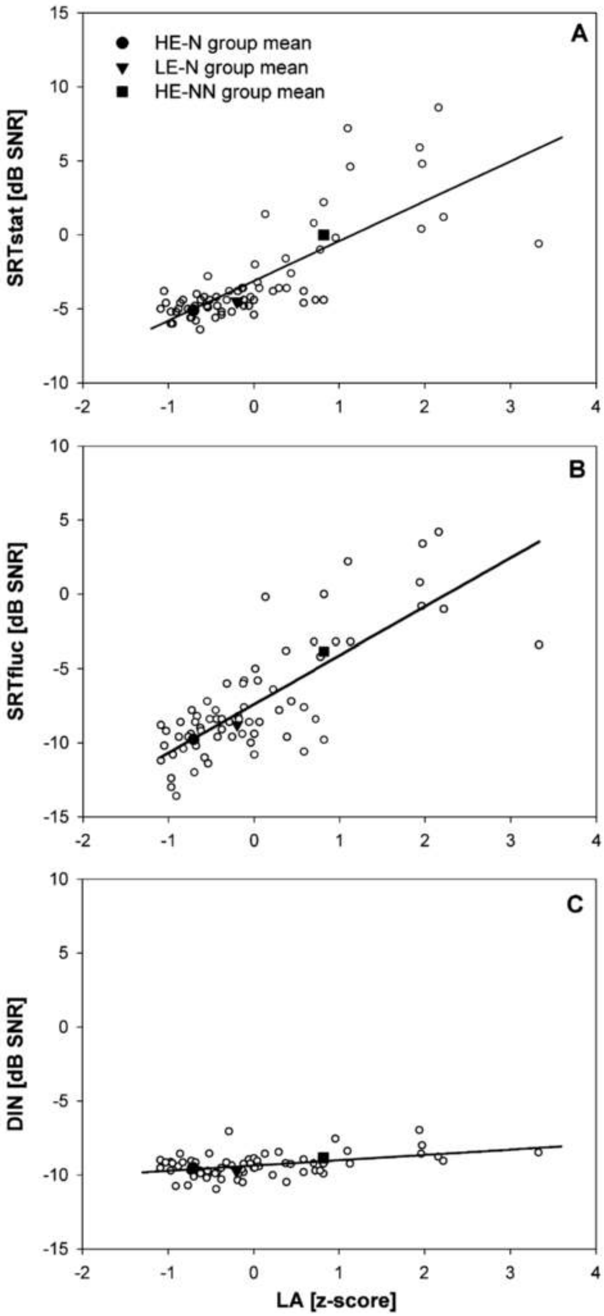


FIGURE 5. Relation between lexical-access ability (LA) and speech recognition of sentences in (A) stationary noise (SRT_{stat}), (B) fluctuating noise (SRT_{fluc}), and (C) digits in noise (DIN). Mean scores for three groups of normal-hearing participants are shown: High Educated Native (HE-N), Lower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

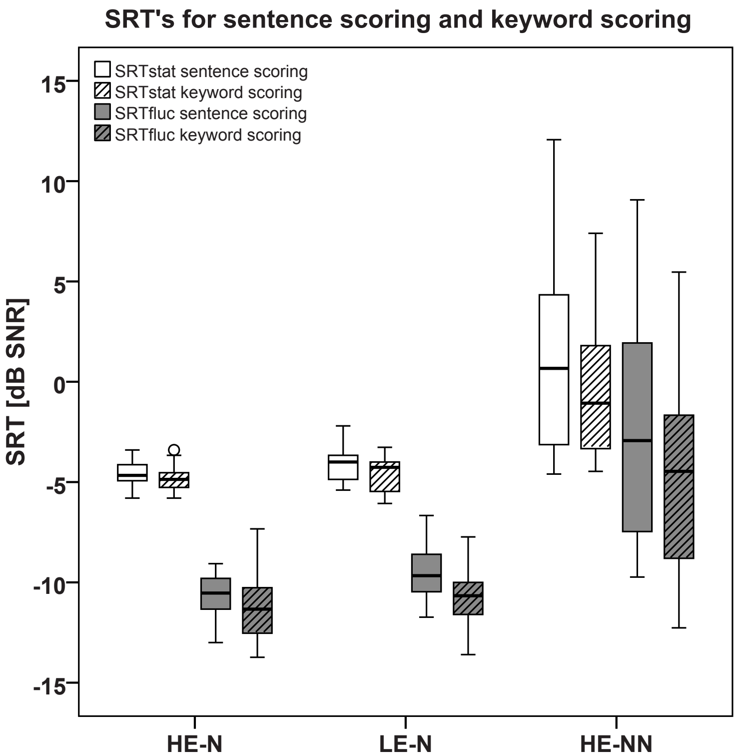


FIGURE 6. Comparison of SRTs obtained with different scoring methods. Boxplots for SRTs with sentence scoring (plain boxes) and for SRTs with keyword scoring (striped boxes) are given for three groups of participants: High Educated Native (HE-N), Lower Educated Native (LE-N), and High Educated Non-native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

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 high and low-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 high-educated native listeners that are often studied, two groups with lower levels of proficiency in Dutch: lower-educated native listeners and high-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, Steeneken, and Houtgast 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 and Festen 2011). 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, Goverts, and Festen 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 et al. (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, 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 (8-9 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 and 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 are 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 and 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, Goverts, and Festen 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 Wordnaming did not discriminate between the two groups of native listeners (high 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; Ronnberg, Samuelsson, and Borg 2000; Van Rooij and 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, Soli, and Sullivan 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.

CONCLUSIONS

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.

4

Lexical-access ability and cognitive predictors of speech recognition in noise in adult cochlear implant users

ABSTRACT

Not all of the variance in speech-recognition performance of cochlear-implant (CI) users can be explained by biographic and auditory factors. In normal-hearing listeners linguistic and cognitive factors determine most of speech-in-noise performance. The current study explored specifically the influence of visually measured lexical-access ability compared with other cognitive factors on speech recognition of 24 postlingually deafened CI users. Speech-recognition performance was measured with monosyllables in quiet (consonant-vowel-consonant [CVC]), sentences in noise (SIN) and digit-triplets in noise (DIN). In addition to a composite variable of lexical-access ability (LA), measured with a lexical-decision test (LDT) and word-naming task, vocabulary size, working-memory capacity (Reading Span test [RSpan]), and a visual analogue of the SIN test (text reception threshold test) were measured. The DIN test was used to correct for auditory factors in SIN thresholds by taking the difference between SIN and DIN: SRTdiff. Correlation analyses revealed that duration of hearing loss (dHL) was related to SIN thresholds. Better working-memory capacity was related to SIN and SRTdiff scores. LDT RT was positively correlated with SRTdiff scores. No significant relationships were found for CVC or DIN scores with the predictor variables. Regression analyses showed that together with dHL, RSpan explained 55% of the variance in SIN thresholds. When controlling for auditory performance, LA, LDT, and RSpan separately explained, together with dHL, respectively 37%, 36% and 46% of the variance in SRTdiff outcome. The results suggest that poor verbal working-memory capacity and to a lesser extent poor lexical-access ability limit speech-recognition ability in listeners with a CI.

Marre W. Kaandorp
Cas Smits
Paul Merkus
Joost M. Festen
S. Theo Goverts

Trends in Hearing, 2017; 21: 1-15.
DOI: 10.1177/2331216517743887

INTRODUCTION

A better understanding of the underlying factors that explain the wide range of speech-recognition performance in postlingually deafened cochlear implant (CI) users is increasingly important to improve pre-operative CI counselling and optimize rehabilitation programs. Several factors have been the subject of investigation, but a major part of the variance in speech-recognition outcomes in CI users can still not be explained (e.g., Blamey et al. 2013; Lazard et al. 2012; Roditi et al. 2009). The current study aimed to determine the influence of lexical-access (LA) ability and cognitive factors in speech-recognition performance in CI users.

It is well recognized that various factors play a role in speech-recognition performance of CI users: for example, duration of severe to profound hearing loss before implantation (Blamey et al. 1996; Blamey et al. 2013; Budenz et al. 2011; Holden et al. 2013; Mosnier et al. 2014; Roditi et al. 2009), the position of the electrodes, (e.g., Finley et al. 2008; Lazard et al. 2012), residual hearing, and preoperative speech recognition (Lazard et al. 2012; Leung et al. 2005). Besides these specific CI-related factors, some studies have shown the importance of cognitive skills in speech-recognition tasks in CI users (Heydebrand et al. 2007; Lyxell et al. 2003; Pisoni 2000) as well as for people with normal hearing (NH) or a mild hearing loss (Akeroyd 2008; Van Rooij and Plomp 1990; Zekveld et al. 2007).

Moreover, there is evidence that linguistic factors play a role in speech recognition as well, especially in demanding listening situations. This has been demonstrated for CI users (Heydebrand et al. 2007; Kaandorp et al. 2015; Lyxell et al. 1996), and also for children (Munson 2001), non-native listeners (Kaandorp et al. 2016; Van Wijngaarden, Steeneken, and Houtgast 2002), and congenitally hearing impaired listeners (Huysmans et al. 2014). Lyxell et al. (1996) found in a group of 11 CI candidates that preoperatively measured verbal information processing speed and working-memory capacity were predictors of CI users' levels of speech understanding. Heydebrand et al. (2007) found in a study sample of 33 postlingually deafened CI recipients that improvement in spoken word recognition in the first 6 months after cochlear implantation was not correlated with processing speed or general cognitive ability, but it was associated with higher verbal learning scores and better verbal working-memory capacity measured before implantation. In a larger study with 114 postlingually deafened adult CI users, Holden et al. (2013) also found a positive correlation between a composite measure of cognition (including verbal learning and working memory) with monosyllable word recognition scores with CI, although after controlling for age, cognition no longer significantly affected CI outcome. Thus, possibly, age-related declines in cognition were responsible for the lower speech-recognition performance in that study.

Although several studies support the assumption that cognitive and linguistic factors are relevant in CI speech-recognition performance, results vary among studies and do not point to a test or combination of tests that can be easily used in the clinic for evaluation of CI candidacy or following rehabilitation progress. Some of these studies examine only the influence of cognitive and linguistic factors on recognition of words and sentences in quiet, which may be less cognitively and linguistically demanding than daily life speech understanding. Kaandorp et al. (2016) recently showed a relationship between linguistic skills and speech recognition of sentences-in-*noise* (SIN) in three groups of 24 young adult normal-hearing (NH) listeners with varying linguistic skills: native Dutch listeners with high education, native listeners with lower education, and non-native listeners with high education. Visually measured lexical-access ability (which we consider a mainly fluid cognitive ability, i.e., the capacity for processing information or reasoning) was a better predictor of speech-in-noise recognition than vocabulary size (VS; a type of cognitive ability based on accumulated knowledge, referred to as a crystallized ability, Horn and Cattell 1967). In the Kaandorp et al. (2016) study, speech-recognition abilities in noise were assessed by measuring the speech reception threshold (SRT; Plomp and Mimpen 1979) with a SIN test (Versfeld et al. 2000) and a digits-in-noise (DIN) test (Smits, Goverts, and Festen 2013). It was found that the DIN test measures mainly auditory performance and minimal linguistic aspects, because it was less related to the linguistic variables.

Based on the findings of Kaandorp et al. (2016), the current study explored the predictive value of visual lexical-access ability and cognitive measures on speech-recognition performance both in quiet and in noise of postlingually deafened CI users. VS was included as a second measure of linguistic skills because crystallized knowledge is known to be preserved or even improve with age, whereas fluid abilities (lexical access) tend to decline with increasing age (Horn and Cattell 1967). In the current study with a larger range of ages, vocabulary knowledge might play a bigger role than in the Kaandorp et al. (2016) study with young adult listeners. Two cognitive measures that are known to relate to speech recognition (in noise) were included to evaluate their predictive value in CI performance compared to that of lexical access: the Reading Span test (RSpan, Besser et al. 2012) as a test of verbal working-memory capacity and the Text Reception Threshold test (TRT, Zekveld et al. 2007) as a visual analogue of the SRT for SIN. Zekveld et al. found that, in a group of 34 NH listeners aged 19 to 78 years, 30% of the variance in TRT was shared with variance in SRT, indicating that the same modality-aspecific cognitive skills were needed in both tests to recover written text or speech from noise. Haumann et al. (2012) found a relation between presurgical TRT measures and postsurgical SRT results in postlingually deafenend adults using a CI.

We were essentially interested in predicting SIN recognition with predictors that can be measured *preoperatively*, like cognitive and linguistic factors. We assumed that these visually measured skills would not change significantly after surgery. However, when measuring these factors in speech-recognition performance in CI users, we expected a large additional influence of auditory factors (e.g. temporal and spectral resolution), compared with the study in listeners with normal hearing (Kaandorp et al. 2016). To investigate the explanatory value for performance with CI *postoperatively*, we chose to use the DIN scores to correct for auditory factors, as DIN performance is expected to represent mostly auditory performance. We hypothesized that controlling for auditory performance by looking at the difference between SIN and DIN thresholds after implantation might give a clearer view of the influence of cognitive and linguistic ability in SIN test results.

METHOD

PARTICIPANTS

Twenty-four postlingually deafened CI users (5 men, 19 women) participated in the study. Table 1 lists characteristics of the participants. No dyslexia, reading problems, or relevant medical problems were reported in an interview prior to participation. Etiology of hearing loss was hereditary (*n* = 13), meningitis (*n* = 1), measles (*n* = 1), hearing loss after brain tumor surgery (*n* = 1), otoscleroses (*n* = 1), sudden deafness while giving birth (*n* = 1), or unknown (*n* = 6). Participants were all patients of the Otolaryngology department, Section Ear & Hearing of the VU University Medical Center, Amsterdam, The Netherlands, who responded to an e-mail invitation to participate in the study. CI users with known relevant medical problems (i.e., that would influence their participation) or prelingually deaf patients (i.e., profound hearing loss before the age of four) were not invited. All patients had at least 1 year of experience with their CIs. The majority used Cochlear™ devices (Sydney, Australia; *n* = 23) and there was one Advanced Bionics user (CA, Valencia, CA). During the tests they used their device with the program and setting they preferred for everyday use. If participants wore a contralateral hearing aid, they were allowed to wear it during testing.

All were native Dutch speakers. Participants' vision, with corrective eyewear if needed, was checked with a near-vision screening test (Bailey and Lovie 1980). All participants were able to read the words of a chart with different print sizes down to a size of at least 16 points at approximately 50 cm from the screen and were assumed to have good visual ability.

Participants provided informed consent before participating according to the declaration of Helsinki, and they received reimbursement for travel costs and additionally a fee of 7.50 euros per hour for their contribution to the study. The study was approved by the Medical Ethics Committee of VU University Medical Center Amsterdam (reference number: 2010/246).

TESTS

Speech-recognition measures. Speech recognition was measured using monosyllables, sentences, and digit-triplets.

Monosyllabic word lists (NVA; Bosman and Smoorenburg 1995) consisting of 12 meaningful consonant-vowel-consonant (CVC) monosyllables were used. The first word was used to focus the listener's attention. CVCs were produced by a female speaker and presented in quiet at 65 dB SPL. The test score was defined as the percentage of phonemes correctly reproduced from the last 11 words.

SIN recognition was measured with sentence lists (VU98; Versfeld et al. 2000) consisting of 13 short meaningful sentences, produced by a female speaker, that were eight or nine syllables in length. They were presented in steady-state long-term average speech spectrum (LTASS) masking noise (Versfeld et al. 2000). The SRT in noise, defined as the signal-to-noise ratio (SNR) at which on average 50% of the sentences were repeated correctly, was measured by the adaptive procedure described by Plomp and Mimpen (1979). The speech level was kept constant at 65 dB A, and the noise level was varied. The first sentence of each trial was presented at 0 dB SNR 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. A response was considered correct if all two to seven predefined keywords in the sentence were repeated correctly in the presented order (Kaandorp et al. 2016). After a correct response, the SNR was lowered by 2 dB, and after an incorrect response, the SNR was raised by 2 dB. The SIN score was calculated by taking the average SNR for Sentences 5 to 14 (where Sentence 14 does not exist, but its SNR was calculated from the

TABLE 1. Characteristics of the participants.

Participant number	Age [years]	Age at implantation [years]	Age at onset of HL [years]	Duration of HL [years]	Duration of SHL [years]	Duration of CI use [years]	Aided preoperative CVC score[% correct phonemes]	Sound field detection thresholds [dB HL]	Etiology	Implant and processor type	Strategy, no. of active electrodes	Contralateral hearing aid
1	84	80	63	17	3	3	36	35	unknown	CI24RE, Freedom	ACE, 22	no
2	74	69	19	50	<1	4	0	22	hereditary	CI24RE, CP810	MP3000, 22	no
3	80	74	9	66	6	5	30	39	unknown	CI24RE, Freedom	ACE(RE)22	no
4	79	76	35	42	7	2	33	35	hereditary	CI24RE, Freedom	ACE, 22	no
5	53	51	12	40	5	1	43	20	measels	CI512, CP810	ACE, 22	no
6	56	50	35	15	1	5	50	33	hereditary	CI24RE, CP810	ACE, 20	no
7	59	54	36	18	2	4	26	24	hereditary	CI24RE, Freedom	ACE(RE), 22	no
8	62	56	50	7	1	4	50	39	hereditary	CI24RE, Freedom	ACE(RE), 22	no
9	61	59	52	7	4	1	43	23	unknown	CI24RE, Freedom	ACE, 22	yes
10	78	74	57	18	2	3	45	31	unknown	HR90K HiFocus1J, Harmony	HiRes-S/ Fidelity120, 16	yes
11	70	65	26	40	40	3	12	16	Sudden HL after giving birth	CI24RE, Freedom	ACE, 22	no
12	58	52	0	53	47	5	42	29	hereditary	CI24RE, CP810	ACE, 22	yes
13	64	60	27	33	1	3	47	24	hereditary	CI24RE, Freedom	ACE, 22	no
14	63	60	14	47	1	1	51	28	otoscleroses	CI24RE, Freedom	ACE, 22	yes
15	63	60	8	53	35	2	43	38	hereditary	CI24RE, Freedom	ACE, 22	yes
16	50	47	23	24	16	2	42	23	after brain-tumor surgery	HR90K HiFocus1J, Harmony	HiRes-S/ Fidelity120, 16	no
17	72	68	40	29	2	2	43	35	hereditary	CI24RE, Freedom	ACE, 22	no
18	73	69	55	14	4	3	45	26	unknown	CI24RE, Freedom	ACE, 22	no
19	42	39	6	33	8	2	45	31	hereditary	CI24RE, Freedom	ACE, 22	no
20	66	62	2	61	3	2	30	26	meningitis	CI24RE, Freedom	ACE, 22	no
21	64	62	52	11	<1	1	52	19	hereditary	CI512, CP810	ACE, 22	yes
22	67	66	41	25	1	1	50	21	hereditary	CI512, CP810	ACE, 22	no
23	57	56	35	21	<1	1	48	21	unknown	CI512, CP810	ACE, 22	yes
24	67	58	26	33	10	8	25	26	hereditary	CI24R(CS), Freedom	ACE, 20	no
M (SD)	64 (10)	61 (10)	30 (19)	32 (17)	8 (13)	3.3 (1.8)	39 (13)	28 (7)				

Note. Duration of hearing loss (HL) and duration of severe to profound hearing impairment (SHL) were obtained from questionnaires and medical records. SHL was defined as either not being able to use the telephone or aided monosyllable (CVC) recognition scores of less than 50% phonemes correct at 65 dB SPL. Participants in Group 2 are marked grey. SD = standard deviation; CI = cochlear implant; CVC = consonant-vowel-consonant.

response to Sentence 13; Plomp and Mimpen 1979). In a previous study, we concluded that not all CI users were able to obtain reliable results on the SIN test (Kaandorp et al. 2015). Nevertheless, in the current study, the SIN test was used again because we aimed to match the test battery that is regularly used in our clinic at present, and there is no common alternative for this relevant test. Participants were not screened before inclusion on SIN performance to obtain a representative group of postlingually deafened CI users. It is known that recognition of sentences in noise can be difficult for CI users. There is no commonly used criterion to identify unreliable SRTs. The speech intelligibility index model (SII, ANSI 1997) assumes that all speech information is available when the SNR in steady-state LTASS noise is higher than +15 dB SNR. Thus, in theory, SRTs higher than +15 dB SNR do not reflect the ability to recognize speech in noise, and these SRTs should be classified as unreliable, because (a) the adaptive procedure does not work properly and (b) the SRT reflects no longer the construct of speech-in-noise ability. The upper limit of +15 dB SNR (based on the SII) could be different for signals processed by hearing aids or cochlear implants, and this complex topic could be the subject of future research. Nevertheless, we considered SRTs higher than 15 dB SNR as unreliable in line with a previous study (Kaandorp et al. 2015).

Recognition of digit-triplets in noise was measured using the DIN test (Smits, Goverts, and Festen 2013) that uses digit-triplet lists containing 24 broadband, homogeneous digit-triplets. The digits were produced by a male speaker and were presented in steady-state LTASS masking noise (Smits, Goverts, and Festen 2013). The same adaptive procedure was used for the DIN test as for the SIN test. 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.

All speech-recognition tests were administered three times, where the first list was used as a practice list. Scores on the second and third runs were averaged for the analyses. Recognition of whole sentences will show larger differences between listeners because of differences in cognitive and linguistic abilities, whereas recognition of closed set words (digits) reflects phoneme recognition and is (mainly) associated with auditory capacity (Kaandorp et al. 2016). Therefore we used the DIN test to eliminate the major auditory effects in speech recognition in noise, and hence isolate the additional role of cognitive and linguistic skills. For this purpose a derived variable SRT_{diff} (SIN-DIN) was calculated.

Cognitive and linguistic measures. Linguistic skills were measured with a VS test and two tests of lexical access: a lexical-decision test (LDT) and a word-naming (WN) test. Two commonly used tests that measure combinations of linguistic and non-verbal aspects of cognition, were also included: the TRT and the RSpan. All tests will be explained later.

VS was measured with a subtest of the Groningen Intelligence Test-II (Luteijn and Barelds 2004), which uses a list of 20 visually presented items. For each test word, the participants had to choose the correct synonym out of five alternatives. In this test, raw scores were used to permit direct comparison of participants.

For the LDT (Rubenstein, Lewis, and Rubenstein 1971), the measurement protocol of Kaandorp et al. (2016) was used. Words were used from a previous study by De Groot et al. (2002). Two lists were used for each of three different average word frequencies: low-frequent (LDT_{LF}), middle-frequent (LDT_{MF}), and high-frequent (LDT_{HF}). The lists were presented in two test blocks of three lists (in the order: LDT_{MF}, LDT_{HF}, and LDT_{LF}), each containing 30 words and 20 pseudowords. The pseudowords were constructed from words of the CELEX database (Baayen, Piepenbrock, and Van Rijn 1993) that were altered by changing at least one letter in such a way they represent orthographically correct, but meaningless letter strings. All words and pseudowords were four to seven letters long and were presented in the middle of the screen. Participants were instructed to press a green button with their right hand for each word and a red button with their left hand for each pseudoword and to respond as quickly and accurately as possible. Reaction times (RTs) for correct responses to words and pseudowords were used as well as the number of errors. RTs under 300 ms or above 1,500 ms as well as RTs that deviated more than 2.5 *SD*s from the participants' resulting list average were omitted. The average RT for words of all six lists was used as the test score.

4

WN was measured with a short test with 30 words simultaneously presented on the screen (Kaandorp et al. 2016). 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. The total time needed to read all the words was used as test score. A combined variable of LDT and WN was calculated by converting the RT's of both measures into z scores and averaging them. For calculation of the *z* scores, the mean and *SD* of the NH data of Kaandorp et al. (2016) were used (LDT: $M = 550$, $SD = 115$; word naming: $M = 15.6$ $SD = 6.4$). This combined variable was used as a more pure measure of lexical-access ability (LA).

The TRT test (Zekveld et al. 2007; Besser et al. 2012) was used as a visual analogue of the SIN test. Three lists of 13 sentences (Versfeld et al. 2000) were used, which did not overlap with the sentences used in the SIN test. Sentences were partly masked by a vertical bar pattern and were presented on a computer screen. Participants were instructed to read the sentence out loud as accurately as possible. The test result indicated the percentage of unmasked text, at which the participant was able to read 50% of the sentences correctly. Masking patterns were, analogous to the SIN test, adaptively changed depending on the response. The first sentence was initially presented at a level of 40% unmasked text. It was then repeated with a decrease in masking of 12% for every repetition until the participant was able to read the complete sentence correctly. All subsequent sentences were presented only once. The change in masking for each following sentence was 6% up after a correct response and down after an incorrect response. The test result was the mean percentage of unmasked text of Sentences 5 to 14. In contrast to the original TRT (Zekveld et al. 2007) the TRT_{Center} was used, because of its higher correlation with the SRT for sentences in stationary noise (Besser et al. 2012). In the TRT_{center} test, sentences are presented word by word in the center of the screen. The presentation time of each word corresponded to the duration of the word in the respective audio recording of the sentence. Each participant did three runs. Scores on the second and third runs were averaged for the analyses.

Verbal working memory was measured with the RSpan test (Besser et al. 2012). In this complex dual task, test sets of sentences were presented on a computer screen. Sentences constructed of five words in past tense were presented in three parts (subject – verb – object) in the center of the screen. Half of the sentences were semantically sensible; the other half were absurd. Twelve sets of sentences in increasing set-size order were used, with each set-size presented three times (3 x 3, 3 x 4, 3 x 5, 3 x 6). After every sentence, the participant had to indicate whether it was semantically sensible or absurd. At the end of every set, participants were asked to recall either all first words (subjects) or all last words (objects) in the set. Which words to recall was unknown in advance. Participants were given a maximum of 80 s to recall the requested subjects or objects. The test result was the total number of correctly recalled target words.

PROCEDURE

Tests were presented in a fixed order with the same lists and test setup for each participant to enable comparisons between listeners. Each participant completed a single 2-hour test session with a 15-min break. The order of tests was as follows: Sound field thresholds, CVCs, RSpan, LDT, DIN, SIN, TRT, VS, and WN. Tests were performed in a sound-treated booth by a trained experimenter. Sound field thresholds were measured with the aid of a clinical audiometer (Decos Audiology Workstation, Decos Systems, Noordwijk, The Netherlands) and a loudspeaker (Yamaha MSP5 Studio). Speech-in-noise tests were measured using a Soundblaster Audigy soundcard and a Soundblaster T20 loudspeaker. The signal was calibrated with a sound level meter at the expected position of the participants heads. Participants were seated facing the loudspeaker at a distance of approximately 70 cm or at a comfortable distance from the display monitor.

The effects of several personal factors and test scores on speech recognition were analyzed first with correlation analyses. Second, variables that showed significant correlations were included in multiple linear regression analyses. Also, the data of CI users were compared with the data of listeners with normal hearing with a wide range of linguistic abilities (Kaandorp et al., 2016) to show the effect of lexical-access ability on speech-in-noise performance for these groups.

RESULTS

OUTCOME MEASURES AND PREDICTORS

CVC, DIN, and SIN scores are shown in Figure 1 for each participant. The variables were approximately normally distributed (checked by visual inspection of histograms and quantile-quantile [Q-Q] plots), except for the CVC phoneme scores. CVC phoneme scores ranged from 47% to 97% with a mean of 86%. DIN thresholds ranged from -5.8 to 7.8 dB SNR with a mean of -0.7 dB SNR. For the SIN test there were useful data for only 20 CI users. For the remaining four CI users, both SIN thresholds were higher than 15 dB SNR. For two CI users, one SIN threshold was higher than 15 dB and so their result was based on one list. For the 20 participants, SIN thresholds ranged from 3.0 to 13.6 dB SNR with a mean of 7.4 dB SNR. To examine whether the poor sentence recognition performance of the four participants with unreliable SIN thresholds was related to cognitive or linguistic abilities, participants were divided into two performance groups (see Figure 1), and the individual data of participants in group 2 were also examined.

Summary data of all cognitive and linguistic measures are shown in Table 2. All variables were approximately normally distributed (checked by visual inspection of histograms and Q-Q plots).

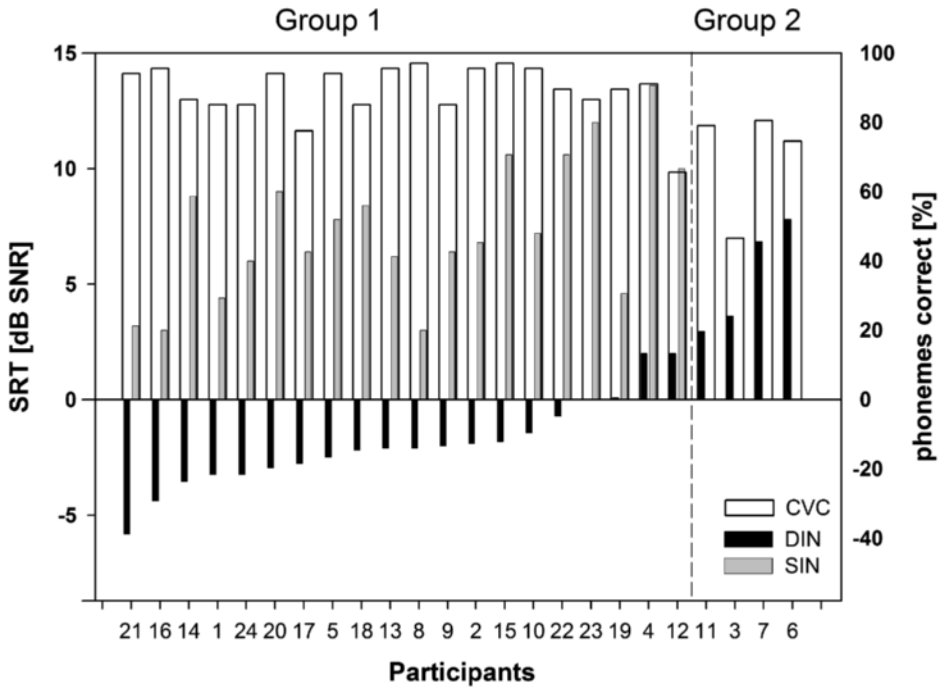


FIGURE 1. Speech reception thresholds (left axis) for sentences (SIN) and digit-triplets (DIN) in stationary noise and % phonemes correct (right axis) for monosyllables in quiet (CVC). Participants were arranged in order of DIN threshold (from best to poorest) and divided into two performance groups.

TABLE 2. Means, standard deviations, and ranges are given for scores on the linguistic and cognitive tests.

	NH data (from literature)		**Total CI group** (n=24)		***t*-tests**	
Predictor variables	*M (range)*	*SD*	*M (range)*	*SD*	*t*	*p*
LDT [ms]	550 (419 – 965)	115	585 (439 – 825)	101	-1.51	.134
WN [s]	15.6 (8.4 – 40.2)	6.4	16.8 (10.5 – 29.2)	3.8	-0.86	.394
VS [# correct]	11.8 (3 – 19)	3.5	13.1 (7 – 18)	3.0	-1.58	.117
RSpan [# correct]	19.7 (4 – 34)	6.1	16.8 (5 – 33)	7.6	1.80	.076
TRT [% unmasked text]	59.7 (49 – 75)	5.4	59.8 (53 – 74)	5.2	0.08	.939

Note. Data of normal-hearing (NH) listeners from previous studies are also given. Independent t-tests were done to compare the groups. LDT = lexical-decision test; WN = word-naming test; VS = vocabulary size test; RSpan = reading span test; TRT = text reception threshold test. NH data was obtained from Kaandorp et al. (2016) for LDT, WN, and VS (n = 72); and from Besser et al. 2012 for TRT and RSpan (n = 55).

Results for LDT, WN, and VS for CI users were comparable to the results of the study with 72 young NH listeners (Kaandorp et al. 2016), with a wide range of linguistic skills and different educational levels. Independent sample *t* tests (Table 2) confirmed that results for the CI users were not statistically different from the total group of NH listeners in that study. Also, results for the RSpan test and TRT of the CI users were not statistically different from the results of the 55 NH listeners with average age of 44 years (range 18 – 78 years) tested by Besser et al. (2012). In the Kaandorp et al. (2016) study, a correlation of 0.69 was found between WN and LDT in NH listeners. In the current study population, we found a moderate correlation but the sample size was too small for such effects to be significant (r = 0.34, p = .16). Nevertheless, the results of LDT and WN were, in line with the NH study, combined into a composite measure LA and included in the analyses, in addition to WN and LDT RTs.

Table 3 shows the individual test results of the participants in Group 2. Inspection of the individual data showed that participants in Group 2 performed poorer than Group 1 with respect to DIN thresholds. Participant 3 performed poor on all tests, auditory and linguistic/cognitive. The other three participants all had varying but not clearly poor results on the cognitive/linguistic tests. For the following analyses, only participants with SIN scores were included (n = 20). Because a high number of correlations were calculated, results have to be interpreted carefully as type I errors might occur.

TABLE 3. Individual data of the participants with unreliable SIN thresholds, Group 2 (n = 4, SIN > 15 dB SNR)

Participant	3	6	7	11
DIN	3.6	7.8	6.9	3.0
CVC	47	75	81	79
LDT	670	494	493	652
WN	20.4	17.0	21.6	17.3
VS	10	11	11	14
RSpan	8	27	21	19
TRT	62	56	59	58
SFT	39	33	24	16

Note. SIN = sentences-in-noise test in dB SNR; SNR = signal-to-noise ratio; DIN = digits-in-noise test in dB SNR; CVC = consonant-vowel-consonant monosyllable test, in % correct phonemes; LDT = lexical-decision test, in ms; WN = word-naming test, in s; VS = vocabulary size test, in number of correct responses; RSpan = Reading Span test, in number of correct responses; TRT = text reception threshold test, in % unmasked text needed to reach 50% correct responses; SFT = sound field threshold, in dB HL.

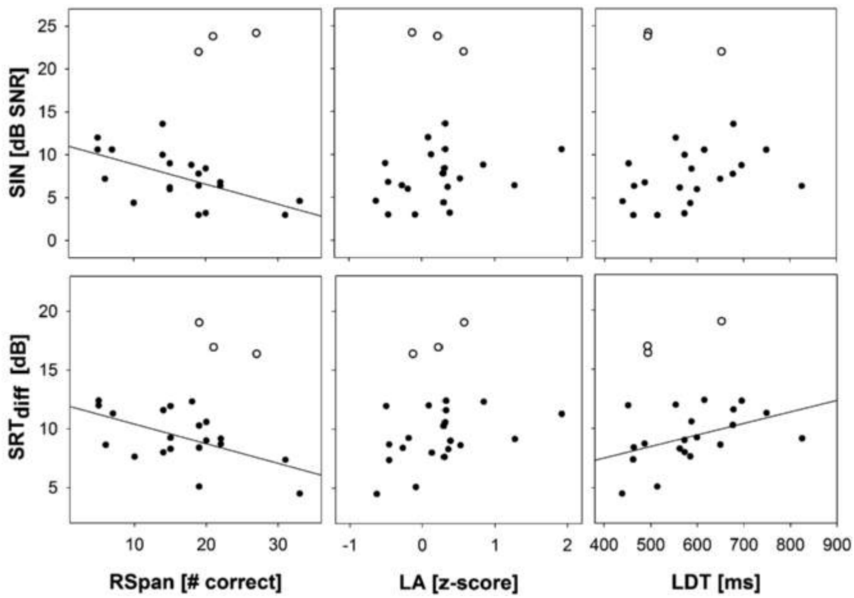


FIGURE 2. Relations between working-memory capacity (RSpan), lexical-access ability with a composite measure (LA), and with a lexical-decision test (LDT) with sentence-in-noise recognition (SIN) and the derived variable SRT_{diff} (difference between sentence-in-noise recognition and digits-in-noise recognition). Open symbols reflect the unreliable SIN results of listeners in Group 2. Lines represent significant correlations.

BIOGRAPHIC AND AUDIOLOGIC FACTORS

The influence of biographic and audiologic factors on speech-recognition outcomes was analyzed first. Pearson's correlations were computed between speech-recognition variables and age, age at onset of hearing loss, age at implantation, duration of hearing loss (dHL), duration of severe to profound hearing loss, years of experience with CI, and aided preoperative CVC phoneme score. Correlations were not significant, except for dHL with SIN threshold (r = 0.49, p = .029), suggesting that poorer SIN results were related to a longer dHL.

COGNITIVE AND LINGUISTIC FACTORS

Next, Pearson's correlations were computed between the speech-recognition outcome variables and the cognitive and linguistic variables as well as the sound field thresholds (Table 4). There were no significant correlations between CVC or DIN scores with any of the cognitive and linguistic tests. For the SIN test and SRT_{diff}, significant strong correlations were found with the RSpan test (r = - 0.59, p = .006 and r = -0.57, p = .009 respectively). Also, a significant correlation was found for SRT_{diff} with LDT score (r = 0.45 , p = .047).

Relations between RSpan, LA, and LDT with SIN and SRT_{diff} are shown in Figure 2 (Group 2 data are shown by open symbols).

Because age is related to some cognitive and linguistic measures, Pearson's correlation analyses for the total group (*N* = 24) were done between age, the cognitive, and the separate linguistic measures. They revealed significant correlations for age with RSpan (r = -0.54, p = .006) and LDT (r = 0.47, p = .019) indicating poorer performance on both tests with higher age.

REGRESSION ANALYSES

Multiple linear regression analyses were done to evaluate predictive value of the cognitive and linguistic variables for sentence recognition performance in addition to other relevant factors. Predictor variables that showed significant strong correlations with SIN threshold or SRT_{diff} score in the correlation analyses were considered. Those were dHL, RSpan, and LDT score. Because we expected a relationship with composite measure LA based on Kaandorp et al. (2016) and there was a trend for LA with SRT_{diff} in the correlation analyses, also LA was examined as predictor variable. First, regression models were examined that could be used *preoperatively* in relation to the application for determining CI candidacy. Separate regression analyses were performed for Group 1 (n = 20) with SIN as dependent variable and LA, LDT or RSpan as independent factors to compare predictive value of these measures, with each time dHL first entered into a separate block. Table 5 shows that the model with dHL and RSpan together explains 55% of the variance in SIN outcome. LA or LDT did not significantly improve the model on top of dHL.

TABLE 4. Pearson's correlation coefficients for speech recognition measures and cognitive and linguistic factors for CI users (*n* = 20) without poorest performers.

Variables (n=20)	**Word naming**	**LDT**	**LA**	**VS**	**RSpan**	**TRT**	**SFT**
CVC	-0.02	-0.25	-0.19	-0.22	0.03	-0.13	-0.08
DIN	-0.03	0.07	0.03	-0.19	-0.26	0.03	0.27
SIN	0.12	0.38†	0.34	-0.33	-0.59**	0.08	0.02
SRT_{diff}	0.18	0.45*	0.42†	-0.27	-0.57**	0.08	-0.20

Note. Significance levels are given. CI = cochlear implant; CVC = consonant-vowel-consonant monosyllables; DIN = digits-in-noise test; SIN = sentences-in-noise test; SRT_{diff} = difference measure (SIN-DIN); LDT = lexical decision test; LA = lexical access; VS = vocabulary size; RSpan = Reading Span test; TRT = text reception threshold; SFT= sound field thresholds.

† $p \leq .10$, * $p \leq .05$, ** $p \leq .01$

TABLE 5. Multiple regression analyses for preoperative prediction of SIN performance of CI users (*n* = 20) at 1 year or more post CI activation.

Predictor	*B*	*SE*	*p*	R^2 *change*	R^2
Model with LA					
dHL	0.09	0.04	.022	0.215	
LA	1.87	0.94	.065	0.155	0.37
Model with LDT					
dHL	0.09	0.04	.026	0.238	
LDT	0.01	0.01	.082	0.128	0.37
Model with RSpan					
dHL	0.08	0.03	.013	0.238	
RSpan	-0.22	0.06	.003	0.314	0.55

Note. Separate regression analyses were performed for LA, LDT and RSpan after controlling for duration of hearing loss. SIN = sentences-in-noise test; CI = cochlear implant; dHL = duration of hearing loss; LA = lexical access; LDT = lexical-decision test; RSpan = Reading Span test.; *B* = Unstandardized regression coefficient; *SE* = standard error; *p* = level of significance; R^2 = proportion of variance.

TABLE 6. Multiple regression analyses for postoperative prediction of SRT_{diff} performance of CI users (*n* = 20) at 1 year or more post CI activation.

Variables	*B*	*SE*	*p*	R^2 *change*	R^2
Model with LA					
dHL	0.06	0.03	.04	0.156	
LA	1.68	0.72	.032	0.216	0.37
Model with LDT					
dHL	0.06	0.03	.058	0.173	
LDT	0.01	<0.01	.041	0.184	0.36
Model with RSpan					
dHL	0.05	0.03	.048	0.173	
RSpan	-0.16	0.05	.007	0.291	0.46

Note. Separate analyses were performed for LA, LDT and RSpan after controlling duration of hearing loss. CI = cochlear implant; SRT_{diff} = difference between sentences-in-noise (SIN) and digits-in-noise (DIN) thresholds; dHL = duration of hearing loss; LA = lexical access; LDT = lexical-decision test; RSpan = Reading Span test.; *B* = Unstandardized regression coefficient; *SE* = standard error; *p* = level of significance; R^2 = proportion of variance.

TABLE 7. Multiple regression analyses for prediction of SRT_{diff} scores of CI users (*n* = 20) and normal-hearing listeners (*n* = 72).

Variables	*B*	*SE*	*P*	R^2 *change*	R^2
LA	2.23	0.25	<.001	0.44	
CI	2.66	0.52	<.001	0.13	0.57

Note. SRT_{diff} = difference between sentences-in-noise (SIN) and digits-in-noise (DIN) thresholds; LA = lexical access; CI = dummy variable for CI use; *B* = Unstandardized regression coefficient; *SE* = standard error; *p* = level of significance; R^2 = proportion of variance.

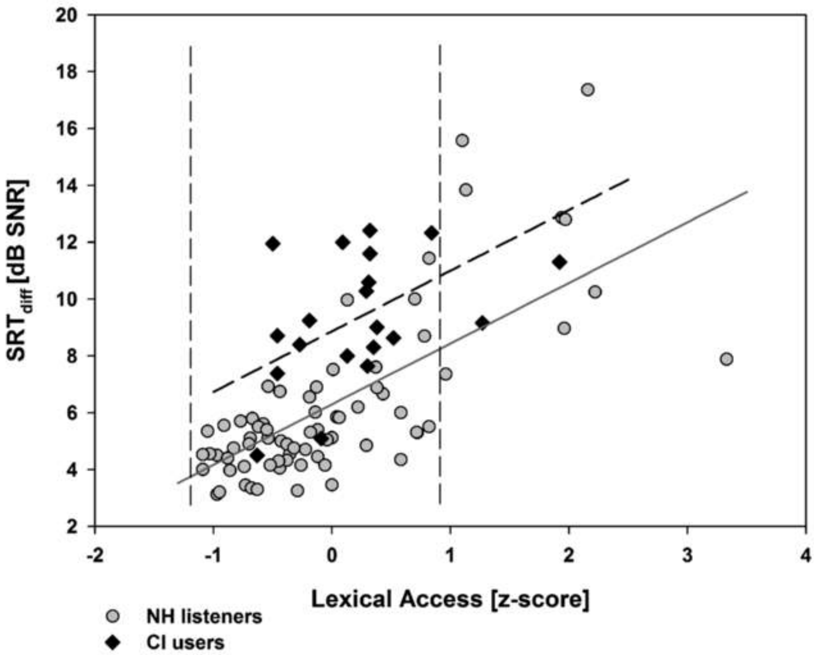


FIGURE 3. Difference in speech reception threshold for sentences in noise and digits in noise (SRT$_{diff}$) for cochlear implant users (diamonds) together with the normal-hearing (NH) data of listeners with various levels of linguistic skills and nonnative listeners of Kaandorp et al. (2016)(circles). Lines represent regression lines for both groups (black dashed for CI users, grey solid for NH listeners). The area between the vertical dashed lines represents the range of native NH listeners.

Next, multiple linear regression models were examined that could be used to explain results *postoperatively*. The regression analyses were repeated using the DIN thresholds to correct for auditory factors by using SRT$_{diff}$ as dependent variable. Table 6 shows that LA, LDT and RSpan accounted for an additional 22%, 18% and 29% of variance, respectively, in SRT$_{diff}$ scores after forcing dHL first into the regression. Together these variables explained 37%, 36% and 46% of SRT$_{diff}$ scores, respectively with LA, LDT or RSpan in the model.

COMPARISON TO NORMAL-HEARING LISTENERS

Figure 3 shows that CI users had similar LA scores compared with the NH group from the Kaandorp et al. (2016) study. To examine the relation between LA and speech recognition in noise for CI users compared with NH listeners, a multiple linear regression analysis was done on the total of NH and CI data (*n* = 92). For this analysis, we used SRT$_{diff}$ values instead of the SIN thresholds to eliminate auditory factors from SIN results. In this combined group, SRT$_{diff}$ was positively skewed. Log transformation, however, did not result in a normal distribution. Dummy variables were used to evaluate group differences ('CI' and 'No CI'). The results showed that the slope of the regression line for CI users was not statistically different from that of the group of NH listeners, but on average CI users had a 2.7 dB poorer SRT_{diff} than NH listeners with the same LA performance (Figure 3, Table 7). Thus LA seems to predict differences between SRT_{diff} scores in CI users comparable to the effect that LA has in NH listeners.

DISCUSSION

The results of this study, with 20 postlingually deafened CI users, support the hypothesis that there is an influence of linguistic and cognitive abilities on speech recognition in noise performance in CI users. Verbal working-memory capacity was a stronger predictor of CI SIN recognition performance than lexical-access ability. Lexical-access ability, measured with the LDT test, was only correlated with the derived speech-in-noise measure SRT_{diff}. The role of lexical access comes on top of auditory factors that account for most of the variance in performance as was expected. Analyses of speech-recognition outcome measures included in this study showed that word recognition in quiet and DIN recognition were not influenced by cognitive and linguistic measures. However, for sentences recognition with the SIN test 55% of the variance in results was predicted by verbal working memory together with dHL (both variables that can be determined preoperatively). When controlling for both auditory performance (with the DIN score that can be measured postoperatively) and dHL, lexical-access ability accounted for an additional 22% using the combined variable LA or 18% using LDT score and verbal working memory for an additional 29% of the variance in SRT_{diff}.

OUTCOME MEASURES

The word recognition results (CVC phoneme scores, a commonly used CI outcome measure) showed relatively good performance in quiet for the current study group with an average score of 85% of phonemes correct. Only two participants had CVC scores lower than 77%. For speech-recognition performance in noise there was a much larger range of outcomes. For four participants the SIN test produced results in an unreliable range, reflecting very poor performance. This poor speech-recognition performance might be caused by poor cognitive and linguistic processing but could also originate from poor auditory processing. Only one of the four participants had scores on the linguistic or cognitive tests that were among the poorest scores (Table 3), but this participant also had the lowest CVC scores and highest SFT. It is, thus, difficult to say what might have caused the poor results. Table 1 shows no clearly deviating or extreme variables for these poor performers, except that one CI user had a long duration of severe hearing loss, of 40 years. For Participant 7, an implant failure was identified as the cause of a gradual decreasing speech-in-noise recognition performance about two years after data collection for this study. Six months after reimplantation, her speech-in-noise recognition thresholds had drastically improved, which agrees with our assumption that her poor performance in noise during the study was primarily caused by a very poor auditory input.

We expect that the other 3 poor performers in noise also received a more degraded input. Although the speech recognition in quiet was for three listeners near that of the other participants, their received spectral detail could be much less due to differences in for instance electrode placement, neural survival, and the amount of channel interactions by spread of current in the cochlea. Several studies have shown that poor speech recognition in noise is not necessarily related to poor speech recognition in quiet. For instance, Shannon, Fu, and Galvin (2004) concluded, in a CI simulation study, that speech recognition in quiet requires only four spectral channels, whereas more complex materials can require 30 or more channels for the same level of performance. Results of Friesen et al. (2001) suggested that not all CI users were able to use all channels of spectral information provided by the implant, resulting in poorer speech recognition especially at lower SNR. They hypothesized that the use of multiple electrodes was limited by electrode interactions in these listeners. Thus, for some listeners with reasonable speech recognition in quiet, the available spectral detail can be limiting speech-recognition performance in noise. A very poor quality of auditory input in some CI recipients may result in a different effect of cognitive and linguistic factors. Very degraded signals might not contain enough usable information to put good cognitive or linguistic skills into action. The results of a study on phonemic restoration (Baskent, Eiler, and Edwards 2010) indicated that listeners with mild hearing loss were able to benefit from top-down processing, while listeners with moderate hearing loss were not. Collison, Munson, and Carney (2004) also studied a group of CI users and hypothesized that differences in signal perception elicited by the heterogeneity of the study group might explain the lack of a relation of linguistic and cognitive skills with speech recognition in their study. They concluded that predictive relations of cognitive and linguistic variables with spoken word recognition might exist only in groups of listeners that are homogeneous with respect to other variables that affect implant use. In addition, Baskent et al. (2016) concluded that the interaction between bottom-up information in case of degraded speech and how this degradation can be compensated for using cognitive mechanisms is complex. In the current study, more clear relations were indeed found for the subset of CI users without the poorest performers on speech-in-noise recognition. Thus, the degree of degradation of the auditory signal seems to interact with the influence of top-down processes on speech recognition in noise. This limits some of our conclusions on the relations discussed later only to relatively good performing CI users.

PREDICTOR VARIABLES FOR SPEECH RECOGNITION

In many studies, duration of (severe) hearing loss has been found to predict CI outcome (e.g., Blamey et al. 2013; Holden et al. 2013; Mosnier et al. 2014). In the current study, this variable was related to SIN recognition, but not to phoneme recognition in quiet or DIN recognition. Compared with those studies, the range of scores and sample size of the current study was smaller, which was sufficient to detect strong correlations but could be too small for milder effects to be significant. Another explanation could be that possibly a longer duration of deafness has a stronger impact on the linguistic and more central auditory system, which is mostly reflected in SIN thresholds, than on peripheral auditory pathways. The other biographic or audiologic variables that were found to influence CI outcome in other studies (with very large study samples) were not significantly correlated with any of the speech-recognition scores in the current study sample.

Cognitive and linguistic abilities of the studied CI users did not clearly deviate from NH listeners with a broad range of cognitive and linguistic abilities from previous studies. This could be expected because these CI users acquired their linguistic skills with NH and the tests were presented visually. For instance, only two CI users had LA scores worse than the native young listeners of Kaandorp et al. (2016), in the range of the non-native highly educated young listeners performance (see Figure 3). Other studies concerning postlingually deafened adult CI users found similar results (e.g., Collison, Munson, and Carney 2004). Correlation analyses showed that word recognition in quiet and DIN recognition were not correlated with any of the cognitive and linguistic measures. For word recognition in quiet, this could be expected because most CVC phoneme scores were relatively high, possibly causing a ceiling effect at the chosen presentation level. Heydebrand et al. (2007) did find correlations in 33 participants between improvement of CVC word scores in quiet 6 months postoperatively and some cognitive measures, possibly because of the larger range in scores. For DIN recognition, the absence of a correlation confirms the earlier findings that the DIN test is mainly associated with auditory capacity with only a small cognitive component (Kaandorp et al. 2016). This was also shown by

Moore et al. (2014), who found only a 0.7 dB better DIN score for listeners with higher cognitive function in a Biobank study with a very large group (n > 500,000) of listeners.

SIN recognition and the derived variable SRT_{diff} were, however, influenced by verbal working memory, which is in line with the findings of, for instance, Akeroyd (2008) for hearing impaired listeners. In the current study sample, the correlation between SIN thresholds and the composite measure LA was not significant, possibly because auditory factors caused more variation in results and the much larger age range of the CI users than the NH listeners in our previous study (Kaandorp et al. 2016). For SRT_{diff} (SIN recognition, corrected for auditory performance by the DIN threshold), a correlation was found with lexical-access ability when measured with the LDT. The composite measure LA was not significantly correlated with SRT_{diff}. Thus, although this finding has to be interpreted carefully because of the multiple correlations and the risk of type I errors, this suggests that lexical-access ability, measured with the LDT, does also relate to sentence recognition in noise of CI users, but the relation is not as clear as in NH listeners. After controlling for auditory factors by predicting SRT_{diff}, LA or LDT did add to the prediction model with dHL, which could be used to explain performance after implantation. Lyxell et al. (1996) also found a relation between lexical access measured pre-operatively with a lexical-decision task and subjective speech understanding of CI users. They had a study sample with a larger range of performance than the current sample. Four of the 11 participants in their study had only environmental awareness or improved speechreading with their implants, while others could understand a conversation over the telephone. When we compared the current results to the NH data of Kaandorp et al. (2016), we found that the slope of the regression line was not different between the groups, but there was a -2.7 dB worse SNR for CI users. Part of this difference may be attributed to the fact that SRT_{diff} increases for larger hearing losses due to a less steep slope of the speech information function for listeners with hearing loss (see e.g., Smits and Festen 2011). Understanding the exact meaning of this difference requires further research on this topic. To conclude, the influence of LA on CI speech recognition seems comparable to that for NH listeners but comes on top of auditory factors, thereby showing comparatively lower predictive power.

The TRT scores of our CI users were comparable with those of the NH listeners of Besser et al. (2012). A similar influence of TRT scores on SIN thresholds could thus be expected, which would suggest TRT to be a valuable predictor of CI outcome before implantation. Unlike Haumann et al. (2012), we did not find a relation between the TRT test and speech-recognition scores. The group of CI users in Haumann's study was larger than our study group; 96 participants, and still they found only a moderate correlation of $r = 0.27$ ($p = .012$), which might explain the difference in findings. Another explanation for the absence of a relation between TRT and SIN recognition can be the use of TRT_{center}, which is more difficult than the original TRT test, because words are presented one by one instead of the whole sentence at once. In future studies, the original TRT might be a better choice, also to allow for comparison of results with previous studies. VS was also not correlated with speech-recognition scores. This corresponds to findings of, for instance, Heydebrand et al. (2007), who did not find a correlation between the size of the vocabulary and improvement of word recognition at 6 months after activation. Results of other studies, on the other hand, suggest that better receptive vocabulary knowledge is related to better speech recognition of NH listeners in adverse conditions (e.g., Benard, Mensink, and Baskent 2014). We hypothesized that VS (a crystalized ability) might improve with age, and thus might show a larger range and on average better results in the current study group compared with the young listeners of Kaandorp et al. (2016), but this was not the case. In the current study, verbal working memory and lexical-decision performance (more fluid abilities) both were correlated with age, RSpan ($r = -0.54$, $p = .006$) and LDT ($r = 0.47$, $p = .019$), where older age yielded a poorer performance on both tests. Besser et al. (2012) also found a significant correlation ($r = -0.52$, $p < 0.01$) between RSpan and age for 55 NH listeners. Some studies on lexical access suggest that in lexical decision, the slower responses of older people are not a result of a lower quality of information processing with older age, but a result of factors like motor movement and degree of cautiousness in responding (Ramscar et al. 2013). Nevertheless, age alone was not correlated with speech-recognition performance in the current study. Because age can relate to both higher as well as lower performance on some cognitive tasks, the variation in age in this study might have obscured results of the influence of these abilities on speech-recognition performance. Future research should focus either on homogeneous groups of CI users with respect to CI-related factors and age, or on very large groups of listeners, to possibly find concealed relations.

To conclude, auditory factors play a major role in speech recognition of CI users. But, of the cognitive and linguistic tests that can be measured pre-operatively, RSpan in particular and LA or LDT to a lesser extent (more fluid abilities) might help to predict and understand SIN thresholds after 12 months of listening experience in CI users.

LEXICAL ACCESS – WORD NAMING AND LEXICAL DECISION

The relation between visual lexical-access ability and SIN recognition was not as clear for CI users alone compared with the NH listeners of Kaandorp et al. (2016). Examining the two lexical-access tests separately, Pearson's correlations with speech-recognition measures showed only a significant correlation for LDT with SRT_{diff} (r = 0.45 , p = .047). The WN test did not correlate with any of the speech-recognition measures, nor did the composite measure LA. In the Kaandorp et al. (2016) study both the LDT scores and WN scores were significantly correlated with SIN thresholds in the total NH group as well as in the native group. The range of WN scores was smaller in this native CI group than in that diverse NH group, which included second language users. This could render our WN test a test of less additive value to the composite measure LA. We used a very simple version of the WN test, measuring a total RT for reading 30 items in one run. A more precise WN test, that measures response times to the onset of each response with a voice key, excluding the time to produce the word, might be needed to indicate differences in lexical-access ability in this population.

In the current study, we used *visual* tests to assess lexical-access ability because we wanted to be able to predict speech-recognition outcomes with CI before implantation. A few studies on *auditory* lexical access suggest that the process of lexical access might be different for listeners with degraded auditory input. For example Farris-Trimble et al. (2014) studied the perception of degraded speech in 33 CI users and 57 age-matched NH listeners of which 16 in a CI simulation condition. They used a visual world paradigm eye tracking task in which fixations to a set of phonologically related items were monitored as they heard a target word being named. They found differences in the process of lexical access for the groups listening to degraded speech relative to a NH group listening to unfiltered speech. They also found weak evidence that the process for CI users was different from that of the NH group listening to CI simulations, suggesting they are accustomed to being uncertain and having to revise their interpretations. Also, McMurray et al. (2016) found evidence that CI users adapt their lexical access to remain flexible in situations of potential misperceptions. McQueen and Huettig (2012) also showed that young adult NH listeners adjust their strategy when there is more uncertainty in the signal. These articles suggest that auditory lexical access might be different for CI users. In the current study we assumed that visual lexical access is not changed because of CI use but is primarily a measure of linguistic ability. Thus, where *visual* lexical access is a linguistic measure that can be obtained prior to implantation, *auditory* lexical access could be different for that person and could better explain speech-recognition outcome during the rehabilitation stage. Auditory lexical access is a future topic of our research.

To conclude, although we found only weak evidence that lexical-access ability is related to SIN recognition in CI users, the findings in this study suggest to further examine the possible predictive value of these tests.

CLINICAL IMPLICATIONS

For both preoperative counseling and optimizing rehabilitation programs, it is very important to have a better understanding of the influence of cognitive and linguistic factors on CI outcome. Practical tests, that can be used in the clinic, are needed for this purpose. The current results suggest that poor speech-recognition performance with CI was likely due to auditory factors that can degrade the auditory signal, especially in noise. However, we can conclude that CI candidates with poor verbal working-memory capacity or slow lexical-access reaction times are not likely to become the best performers after implantation. The visually conducted verbal working-memory and lexical-access tests can be measured preoperatively as well as postoperatively. As in our previous study, the average RT for words of all six lists was used as LDT score. To examine the value of using only two lists, the correlation of SIN and SRT_{diff} with the average of the two lists with medium word frequency (LDT_{MF}) were examined. The analyses showed Pearson's correlations for LDT with SIN (r = 0.41; p = .076) and with SRT_{diff} (r = 0.48; p = .033), comparable with the average of 6 lists. Therefore, in future studies or in the clinic, two lists with a small range of word frequencies can be used instead of six lists, making the test quicker. For the LDT, two lists take about 5 to 8 minutes. Information about working-memory capacity and lexical-access ability can help to better inform CI candidates of speech-recognition outcome with a CI. During the rehabilitation period, this information can, combined with DIN recognition, help understand SIN recognition performance and thus performance in daily life and support a more personalized rehabilitation program. The influence of lexical-access ability on SRT_{diff} suggests that CI users with lower lexical-access skills might, despite favorable DIN scores, have significantly more problems with recognition of sentences in noise compared to their ability to recognize digit-triplets in noise. These findings suggest to use the DIN test for evaluation of fitting of the CI. However, performance in real-life situations is probably better estimated with tests that demand more cognitive and linguistic skills.

5

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a coc much better
that expl s wide range of per
needed to ove preoperative co
rehabilita is thesis aim
between clinically a ble linguis
performance with a co implant. T
skills th in people w
of current asures of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

Effects of vocoded speech or a cochlear implant on auditory lexical decision in relation to measures of speech recognition

ABSTRACT

Objectives: A major aspect of speech understanding is that the heard words need to be matched to representations in the mental lexicon (lexical access). Earlier we found that lexical access, measured with a *visual* lexical-decision test (vLDT), was related to speech recognition in noise in normal-hearing listeners. The objectives of the present study were to examine the effect of degraded input, by either CI-simulated stimuli or by CI use, on *auditory* lexical-decision (aLDT) reaction times and accuracy; to compare these aLDT variables with vLDT variables; and to examine the relation of these measures with speech-recognition performance of CI users.

Design: Eleven young adult normal-hearing (NH) listeners (Experiment 1) and 22 postlingually deafened adult CI users (Experiment 2) participated in the study. Speech-recognition performance was measured with monosyllables in quiet (CVC), sentences in noise (SIN) and digit triplets in noise (DIN). For Experiment 1, speech stimuli were processed with a 6- or 10-band noise vocoder. We hypothesized that reaction times (RTs) and number of errors would increase with more degradation of the input and that response patterns would be different for CI users compared to the NH group. Correlation analyses were done for speech recognition in CI users and several LDT variables: aLDT and vLDT accuracy and RTs, RT differences between modalities, and the quotients of accuracy and RT. These quotients reflect response efficiency, that is, an efficient balance between response accuracy and response time.

Results: Results showed lower accuracy and longer RTs when listening to vocoded speech or with CI use, but response patterns were different between listener groups. CI users were relatively slower in pseudoword responses and showed a bias toward real-word responses, which was not the case in the NH group. Correlation analyses for CI users revealed a relation between CVC scores and real-word RT of aLDT. SIN Thresholds were, however, more related to aLDT efficiency in pseudoword responses and in overall responses.

Conclusions: Degraded auditory input results in slower responses and lower accuracy in lexical-decision tasks. CI users seem to have adopted a different strategy in lexical access than NH listeners who listen to vocoded speech, possibly because CI users are accustomed to listening to degraded stimuli. Speech recognition of monosyllables in quiet was mainly related to real word response times in auditory lexical-decision. The more complex sentence recognition in noise was mostly related to response efficiency, both in pseudoword responses and overall in responses. The results showed that measures of

auditory lexical decision, together with the earlier measures of visual lexical decision, can be helpful in understanding speech-recognition outcome with CI.

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

(to be submitted to Journal of Speech, Language, and Hearing Research)

INTRODUCTION

Speech recognition can be a challenge in complex listening situations, for example when there is background noise or when speech signals are unclear. Listeners with cochlear implants (CIs) especially have difficulty understanding speech in complex situations, but performance varies greatly among CI users (e.g., Kaandorp et al. 2015; Lazard et al. 2012; Holden et al. 2013) and a large part of this variance can still not be explained (Lazard et al. 2012; Blamey et al. 2013). A major aspect of understanding spoken language is that the heard words have to be matched to representations in the mental lexicon (lexical access). Only after matching can the information stored with the mental representation become available to the listener (e.g., Grainger and Jacobs 1996). There is evidence that more efficient lexical access is related to better speech-recognition performance (Van Rooij and Plomp 1990; Lyxell et al. 1998; Ronnberg, Samuelsson, and Borg 2000). In a previous study, we found that lexical-access ability in the visual modality, i.e. with written stimuli, is a predictor of speech-in-noise recognition performance in normal-hearing listeners (NH) with varying levels of proficiency in Dutch (Kaandorp et al. 2016). In CI users, only a weak relation between visual lexical-access ability and speech recognition in noise was found. Instead, as expected, auditory factors showed to be the main predictor of speech recognition with a CI (Kaandorp et al. 2017). In these studies, lexical access was assessed in the visual modality, to allow for examination of one aspect of linguistic abilities prior to implantation. As a result, this measure can be used in pre-operative counseling. Furthermore, these visual measures were chosen because they might help in identifying areas in which rehabilitation programs could be optimized after implantation. Efficient *visual* lexical access, however, does not necessarily imply efficient *auditory* lexical access. In the auditory modality, the input signal is degraded for CI users, which affects their performance. Some studies suggest that CI users might have adapted their lexical access strategy to deal with this degraded input (e.g., Farris-Trimble et al. 2014; McMurray et al. 2016). Knowledge of such an adaptation could be important for understanding speech-recognition outcomes of an individual CI user. The current study focused on the effect of degraded auditory input on *auditory* lexical access, in Experiment 1 for NH listeners listening to vocoded stimuli and in Experiment 2 for CI users for whom listening via electrical stimulation automatically implies listening to a degraded signal. In addition, the relation of several auditory lexical-access measures with speech recognition in quiet and in noise was examined.

MODELS OF AUDITORY AND VISUAL LEXICAL ACCESS

Established models of lexical access in spoken-word recognition (the Cohort Model by Marslen-Wilson 1987; the Shortlist B by Norris and McQueen 2008; the TRACE model by McClelland and Elman 1986; e.g., the Neighbourhood Activation Model by Luce and Pisoni 1998) assume that with the onset of input, multiple lexical candidates are activated simultaneously and form a cohort. When the stimulus unfolds and more information becomes available, competitor words in the cohort that no longer match with the input are deactivated (inhibition). Discrimination is then possible when a word's level of activation reaches a critical difference from that of its competitors (Marslen-Wilson 2013). Visual lexical-access models like the *Multiple Read-out Model* (Grainger and Jacobs 1996) similarly assume that orthographically similar words are activated and evaluated simultaneously. Models of auditory and visual lexical access largely overlap. An important difference is that in the auditory modality the stimulus is not presented and interpreted instantaneously, as in the visual modality, but sequentially as the stimulus unfolds. In auditory lexical access, a so-called uniqueness point (Marslen-Wilson 1987) determines when sufficient information is available for the target word to be discriminated from its competitors. For short words, often the whole word is needed to assemble all necessary information. Longer words can sometimes be discriminated before they are completely pronounced. Another difference is the presence of cues to morphological structure in the auditory modality but not in the visual modality, for example, the prefix in *return* and *rebuild* is pronounced differently (Tyler, Waksler, and Marslen-Wilson 2013).

Lexical-access ability is often assessed by means of lexical-decision tests (LDT). In these tests a person has to discriminate between real words and pseudowords. Pseudowords are letter strings that are in line with the orthography and phonology of the test language, but lack meaning. They need to be well constructed to ensure that the response is based on lexical access and not simply on a process that determines whether or not the stimulus looks word-like. Most models of lexical decision assume that the participant decides that the stimulus is a pseudoword if no match with a representation in the mental lexicon has been found at the moment a preset time deadline is reached (e.g., Grainger and Jacobs 1996; Ratcliff, Gomez, and McKoon 2004). In the current study, we used lexical-decision tests to measure lexical access because they are easy to use and can be used both in the auditory and visual modality.

When studying the effect of degraded input on lexical access by means of a LDT, the recorded reaction times and accuracy (error percentages) for responses to real words and pseudowords provide information on the effect of degradation on overall lexical processing. However, they do not give information about differences in specific stages of lexical access among participants. The reaction time in lexical decision comprises the total lexical-access time (activation of the cohort, evaluation of the cohort, retrieving information from the lexicon), the decision time, and the time it takes to execute the response. We hypothesized that comparisons of results of the *auditory* lexical-decision test (aLDT) with the *visual* lexical-decision test (vLDT) within a person could highlight the parts of the process that differ between the two modalities. In addition, different effects for responses to real words and to pseudowords could inform on different parts of the process.

Reaction times are influenced by several variables. For instance, word frequency (often described by occurrences per million words in a language) is thought to influence activation levels (e.g., Luce and Pisoni 1998). Highly-frequent words are more strongly activated than words with a lower frequency of occurrence, resulting in shorter LDT reaction times (Kaandorp et al. 2016). Meyer et al. (2003) used the Neighborhood Activation Model to predict speech recognition in CI users and concluded that CI users use this word-frequency information to identify spoken words similarly to the way NH listeners do. Pseudoword RTs are also influenced by test parameters. If, for instance, pseudowords are generally very different from the real words in the test (e.g., uncommon letter combinations or very different in length), discrimination between words and pseudowords is relatively easy and a shorter deadline may be used so that RTs for both words and pseudowords will be short (Ratcliff, Gomez, and McKoon 2004). Actually, under those circumstances lexical access may not take place at all, instead the decisions may be based on shallow perceptual processing of the stimuli.

Next to fixed test parameters, individual differences influence LDT reaction times as well. For instance, different levels of proficiency in a language are likely associated with different baseline levels of activation in lexical representations and, hence, with differences in LDT reaction times. Furthermore, LDT reaction times (and LDT errors) depend on the amount of so called 'spurious activation' in the mental lexicon, that is, the number of lexical elements that, in addition to the element representing the input stimulus, are initially activated (i.e., the size of the cohort). It has been shown that the amount of spurious activation during auditory word recognition can be excessively large in nonnative speakers, particularly when the nonnative language contains pairs of closely similar, confusable, phonemes that are not distinctive in the native language (e.g., the /ɛ/ and /æ/, as in *flesh* and *flash*, in the English L2 lexicon of a native speaker of Dutch; Broersma (2007), Broersma and Cutler (2011)). This finding is of particular relevance in the present context because, for listeners with hearing loss, it may be expected that relatively many phonemes are confusable, even in the native language. In other words, the amount of spurious activation during lexical access may be excessively large in people with hearing loss, resulting in relatively long response times and lower accuracy. Another factor that influences RTs and accuracy is age. RTs tend to get longer with older age. Some studies on the effect of age on lexical-decision RT, however, suggest that slower responses are not a result of a lower quality of information processing with older age, but a result of factors like slowed motor movement and a higher degree of cautiousness in responding (Ramscar et al. 2013).

SIGNAL DEGRADATION AND LEXICAL ACCESS

Lexical access has been subject of investigation for many years, however mostly with clear input signals. A few recent studies used a *visual-world* paradigm eye-tracking task to investigate the time-course of spoken word recognition with several degraded auditory stimuli by tracking eye-movements to pictures of target words and competitor words (e.g., McMurray et al. 2016; McQueen and Huettig 2012; Farris-Trimble et al. 2014; Wagner, Toffanin, and Baskent 2016). McQueen and Huettig (2012) for example created uncertainty in the input by replacing a certain number of phonemes in the sentence, but not the target word, by noise. They showed that young adult NH listeners adjust their strategy in competitor evaluation when there is more uncertainty in the input sentence, even if the target word itself is clear speech.

Farris-Trimble et al. (2014) studied the process of spoken word recognition in 33 CI users and 57 age-matched NH listeners, 16 of which listened to degraded signals produced by CI simulation software. Their eye-tracking task showed that both CI users and CI-simulated listeners were overall slower and less accurate in identifying the target word than their age-matched controls. They argued that the poorer signal quality may cause uncertainty, resulting in later activation and different evaluation of the cohort. They also found weak evidence that CI users showed a difference in the degree of peak and late competitor activation, suggesting that CI users maintain more activation for competitors for later revision of interpretation, which might be the result of long-term adaptation to the degraded signal. McMurray et al. (2016) also found evidence that CI users adapt to their degraded input by amplifying competitor activation to preserve their flexibility in the face of potential misperceptions. To sum up, these studies showed that people adjust their lexical processing when input is degraded.

The general goal of the current study was to examine the effect of degraded input on auditory lexical-decision reaction times and accuracy, to compare aLDT with vLDT variables, and to examine the relation of these measures with speech-recognition performance in CI users. We used a simple model for lexical decision, based on accepted models (see above), with the following stages: 1. Activation: orthographically similar lexical elements (in visual lexical decision; e.g., mouse, house, mouth, muse, morse, …) or phonologically similar lexical elements (in auditory lexical decision; e.g., like, light, lime, …) are simultaneously activated and form a cohort. 2. Cohort evaluation: When more information becomes available, competitor elements in the cohort are deactivated. 3. Decision: When a critical activation threshold for the target word is reached, a YES-response follows, or if a time deadline is passed before a match is found, a NO-response follows. In Experiment 1, we aimed to determine aLDT scores for young adult normal-hearing listeners (university students; norm data) and to investigate the effect of degraded input in NH listeners on aLDT RT and accuracy. We hypothesized that for vocoded stimuli compared to original stimuli, the RT for real words would be longer, because the process of activation of candidates and evaluation of competitors takes longer (Farris-Trimble et al. 2014) due to uncertainty in the signal and, probably, spurious activation. We also expected the RT for pseudowords to be longer because the deadline is usually set higher for more difficult conditions (e.g., Ratcliff, Gomez, and McKoon 2004). To examine whether more degradation of the stimulus results in even slower responses, we used two different vocoder settings to produce degraded stimuli. We expected that more degradation would result in more guessing, thus lower accuracy, both for responses to real words as to pseudowords. In Experiment 2 we aimed to compare lexical-decision response patterns of CI users with the results of the NH listeners of Experiment 1. We expected to find differences in relative RTs and accuracy for CI users compared to listeners that were not accustomed to listening to unclear speech (cf., Farris-Trimble et al. 2014). Because they are used to guessing, CI users might have an accuracy bias toward words, more likely identifying pseudowords as words when they need to guess. We also expected that, compared to NH university students listening to vocoded stimuli, CI users would have a larger range in RTs because of a larger range in age and a larger variance in linguistic skills. We assumed that the effect of age and linguistic skills would be similar for aLDT and vLDT. Therefore, a ΔLDT measure, between aLDT in the unprocessed speech condition and vLDT, could reflect the effect of uncertainty caused by the degraded auditory input signal on the lexical-access process, because individual differences in linguistic skills or motor movement were incorporated in both measures and thus compensated in this difference measure.

EXPERIMENT 1

MATERIALS AND METHODS

Participants. Eleven young normal-hearing university students participated (two male, 9 female; age *M* = 22, *SD* = 4). No dyslexia, reading problems, or hearing problems were reported in an interview prior to participation. Hearing thresholds were 20 dB HL or less for octave frequencies between 500 and 8000 Hz, except for one person who had a threshold of 35 dB HL at 8 kHz for one ear and good thresholds for the other ear. All were native speakers of Dutch. Participants' vision, with corrective eyewear if needed, was checked with a near-vision screening test (Bailey and Lovie 1980). All participants were able to read the words of a chart with different print sizes down to a size of at least 16 points at approximately 50 cm from the screen. Participants provided informed consent before participating. The study was approved by the Medical Ethics Committee of VU University Medical Center.

CI simulation. All auditory speech stimuli were administered unprocessed and in two processed conditions to simulate CI use. Based on results of pilot tests, 6-band vocoded speech stimuli (6chVOC, harder condition) and 10-band vocoded speech stimuli (10chVOC, milder condition) were produced by a noise vocoder (Rosen, Faulkner, and Wilkinson 1999) with an envelope low-pass filter cut-off frequency of 320 Hz. These settings resulted in phoneme identification scores comparable to, respectively, reasonable CI-outcome (around 70% phonemes correct for a standard CVC word test, see below), or good CI-outcome (around 90% phonemes correct). The participants of the current study had never listened to vocoded speech stimuli before.

Speech-recognition measures. Speech recognition was measured using monosyllables, sentences, and digit triplets.

Monosyllabic word lists (NVA, Bosman and Smoorenburg 1995) consisting of 12 meaningful monosyllables of the consonant-vowel-consonant (CVC) type were used. The first word was used to focus the listener's attention. CVCs were recorded by a female speaker and presented in quiet at 65 dB SPL. The test score was defined as the percentage of phonemes correctly reproduced from the last 11 words.

Recognition of sentences in noise (SIN) was measured with sentence lists (VU98, Versfeld et al. 2000) consisting of 13 short meaningful sentences, recorded by a female speaker, that were 8 or 9 syllables in length. They were presented in steady-state long term average speech spectrum (LTASS) masking noise (Versfeld et al. 2000). The speech reception threshold (SRT) in noise, defined as the signal-to-noise ratio (SNR) at which on average 50% of the sentences was repeated correctly, was measured by the adaptive procedure described by Plomp and Mimpen (1979). The speech level was kept constant at 65 dB A and the noise level was varied. The first sentence of each trial was presented at 0 dB SNR 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. A response was considered correct if all 2 to 7 pre-defined keywords in the sentence were repeated correctly in the order presented (Kaandorp et al. 2016). After a correct response the SNR was lowered by 2 dB and after an incorrect response the SNR was raised by 2 dB. The SIN score was calculated by taking the average SNR for Sentences 5 to 14 (where Sentence 14 was actually not presented, but its SNR was calculated based on the response to Sentence 13, Plomp and Mimpen 1979).

Recognition of digit triplets in noise was measured using the DIN test (Smits, Goverts, and Festen 2013) that uses digit-triplet lists containing 24 broadband, homogeneous digit triplets. The digits were recorded by a male speaker and were presented in steady-state LTASS masking noise (Smits, Goverts, and Festen 2013). The same adaptive procedure was used for the DIN test as for the SIN test. 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 triplet to be considered correct.

All speech-recognition tests were administered two times in each condition. For each test, participants were assessed in order of decreasing stimulus quality: first with original speech stimuli, then with 'milder' 10chVOC stimuli, and finally with 'harder' 6chVOC stimuli. Scores on the test and retest list were averaged for the analyses.

Lexical-access measures. For the *visual* lexical-decision test (Rubenstein, Lewis, and Rubenstein 1971), words were used from a Dutch list that was earlier used by Groot et al. (2002). These words were derived from the CELEX database (Baayen, Piepenbrock, and Van Rijn 1993). The pseudowords were constructed from words of the CELEX database that were altered by changing at least one letter in such a way that they represent orthographically correct, but meaningless letter strings. All real words and pseudowords were 4 to 7 letters long and were presented in the middle of the screen, one by one. Two fixed lists with the same average word log frequency (3.25 per million) from the measurement protocol of the Kaandorp et al. (2016) study were used, each containing 30 real words and 20 pseudowords. The average subjective concreteness was 4.88, measured on a 7-point scale by Groot et al. (2002). It is important to note that these were all highly frequent words and thus known to all participants.

Participants were instructed to press a green button with their right hand for each word or a red button with their left hand for each pseudoword and to respond as quickly and accurately as possible. Reaction times (RT) were measured from stimulus onset to response. RTs for correct responses to real words and pseudowords as well as errors were registered to be used in the analyses. Accuracy was expected to be near ceiling in vLDT, since all words were highly familiar words. RTs under 300 ms or above 1500 ms (e.g., Carreiras, Perea, and Grainger 1997) as well as RTs that deviated more than 2.5 *SD*s from the participants' resulting list average were not included in the analyses.

The *auditory* lexical-decision test (aLDT) was built with the Psychophysics toolbox in Matlab using Audio Stream Input/output (ASIO) driver. The real words were selected from the same list of Groot et al. (2002) and the same pseudowords as in the vLDT were used. All 4 to 7 letters words (330) and 330 pseudowords were recorded by a female speaker in a sound-treated booth using a flat-panel microphone. Silent periods at the beginning and the end of the audio file were removed manually to obtain precise word length information. Each audio file was RMS normalized. For all stimuli, the duration of the audio file (in ms), the frequency of occurrence, and concreteness values of the words were stored in a database. In each condition, two lists of 15 real words and 15 pseudowords each were used and preceded by 4 practice stimuli. The duration of the stimuli was 380 to 1200 ms with an average of 750-800 ms for each list. Word log frequency of the real words (per million) in each list was on average 3.26 and their average concreteness value was 4.65 (measured on a 7 point scale by Groot et al. 2002). A one way ANOVA showed that the frequency and concreteness values were not significantly different between lists.

Before each stimulus, a 500 ms warning signal sounded, followed by 500 ms of silence. Simultaneously with the warning signal, a loudspeaker icon was presented on the screen. The inter-trial interval was 1500 ms. Participants were again instructed to press a green button with their right hand for each word or a red button with their left hand for each pseudoword and to respond as quickly and accurately as possible. RTs were measured from stimulus onset to response. RTs for correct responses to real words and to pseudowords as well as the error percentage were registered to be used in the analyses. RTs under 100 ms or above 3000 ms as well as RTs that deviated more than 2.5 *SD*s from the participants' resulting list average were not included in the analyses. Like the speech-recognition tests, all LDT tests were administered two times and scores on the test and retest list were averaged for the analyses.

General Procedures. All tests were presented in a fixed order (see Table 1). A pure tone audiogram was made using a clinical audiometer (Decos Audiology Workstation, Decos Systems, Noordwijk, The Netherlands). Tests were performed in a sound-treated booth by a trained experimenter. Speech-in-noise tests were measured using a Soundblaster Audigy soundcard and a Soundblaster T20 loudspeaker. The signal was calibrated with a sound level meter at the expected position of the participants' heads. Participants were seated facing the loudspeaker at a distance of approximately 70 cm or at a comfortable distance from the display monitor.

RESULTS

Speech recognition. Average CVC, SIN and DIN scores for each condition are listed in Table 2. For the tests with original stimuli, the DIN and SIN thresholds were comparable to those reported before for young NH listeners (-9.3 dB SNR and -4.2 dB SNR respectively, Kaandorp et al. 2015). CVC scores for original stimuli were expected to be 100% correct, but two participants misheard 1 phoneme. Results for 10chVOC and 6chVOC stimuli were, as expected, in the same range of speech-recognition scores that are typically found in CI users with reasonable to good speech-recognition performance (-1.8 dB SNR for DIN and +8.0 dB SNR for SIN, Kaandorp et al. 2015).

Auditory lexical decision. Table 3 shows means for the aLDT response times and error percentages for each condition. The RTs for the original stimuli can be regarded as baseline data for those in the other conditions. For the separate conditions, aLDT RT for real words in the unprocessed condition was a bit left skewed and error percentages for the original stimuli were at floor level. For the other aLDT variables RTs and error percentages were approximately normally distributed (checked by visual inspection of variable histograms). Error percentages were transformed into rationalized arcsine units (RAU) to linearize the data in relation to the variance (Studebaker 1985).

TABLE 1. Measurement protocol.

Test	Condition 1 Original stimuli	Condition 2 10chVOC	Condition 3 6chVOC
1. CVC	2 lists	2 lists	2 lists
2. vLDT	2 lists		
3. aLDT	2 lists	2 lists	2 lists
4. SIN	2 lists	2 lists	2 lists
5. DIN	2 lists	2 lists	2 lists

Note. The tests were presented in a fixed order. All auditory tests were presented two times in each of the three conditions in order of increasing difficulty. CVC = monosyllable test; vLDT = visual lexical decision test; aLDT = auditory lexical decision test; SIN = sentences-in-noise test; DIN = digits-in-noise test; 10chVOC = 10 band noise vocoded speech condition; 6chVOC = 6 band noise vocoded speech condition.

TABLE 2. Means and ranges of speech-recognition scores for three conditions of speech stimuli: Original, 10 band vocoded speech (10chVOC), and 6 band vocoded speech (6chVOC).

Variable	Original mean (range)	10chVOC mean (range)	6chVOC mean (range)
CVC [%]	99.7 (98.5 – 100)	88.3 (82.0 – 92.5)	76.4 (62.5 – 85.0)
DIN [dB SNR]	-9.3 (-10.9 – -7.8)	-3.4 (-5.1 – 0.1)	-0.2 (-1.7 – 1.25)
SIN [dB SNR]	-4.3 (-5.2 – -3.4)	4.5 (3.0 – 7.0)	9.2 (6.8 – 10.8)

Note. CVC = monosyllable test, in % correct phonemes; SIN = sentences-in-noise test, in dB SNR; DIN = digits-in-noise test, in dB SNR. SNR = signal-to-noise ratio.

TABLE 3. Means of reaction times (RT) and error-percentages for real words and pseudowords in auditory lexical decision (aLDT) for three conditions of speech stimuli: Original, 10 band vocoded speech (10chVOC), and 6 band vocoded speech (6chVOC) in Experiment 1 and for CI users of Experiment 2. Also RTs and % errors for visual lexical decision (vLDT) are given.

	vLDT				aLDT			
	Real words		Pseudowords		Real words		Pseudowords	
	RT [ms] *M* (SD)	%errors *M*	RT [ms] *M* (SD)	% errors *M*	RT [ms] *M* (SD)	% errors *M*	RT [ms] *M* (SD)	% errors *M*
NH								
Original	473 (49)	3.0	567 (47)	11	879 (71)	1.8	963 (82)	1.8
10chVOC					1013 (65)	8.8	1113 (93)	8.8
6chVOC					1003 (49)	17.6	1152 (112)	17.6
CI	579 (73)	1.3	764 (133)	6.5	1174 (126)	12.7	1618 (288)	23.2

Note. vLDT = visual lexical-decision test; aLDT = auditory lexical-decision test; NH = normal hearing listeners; CI = cochlear implant users;

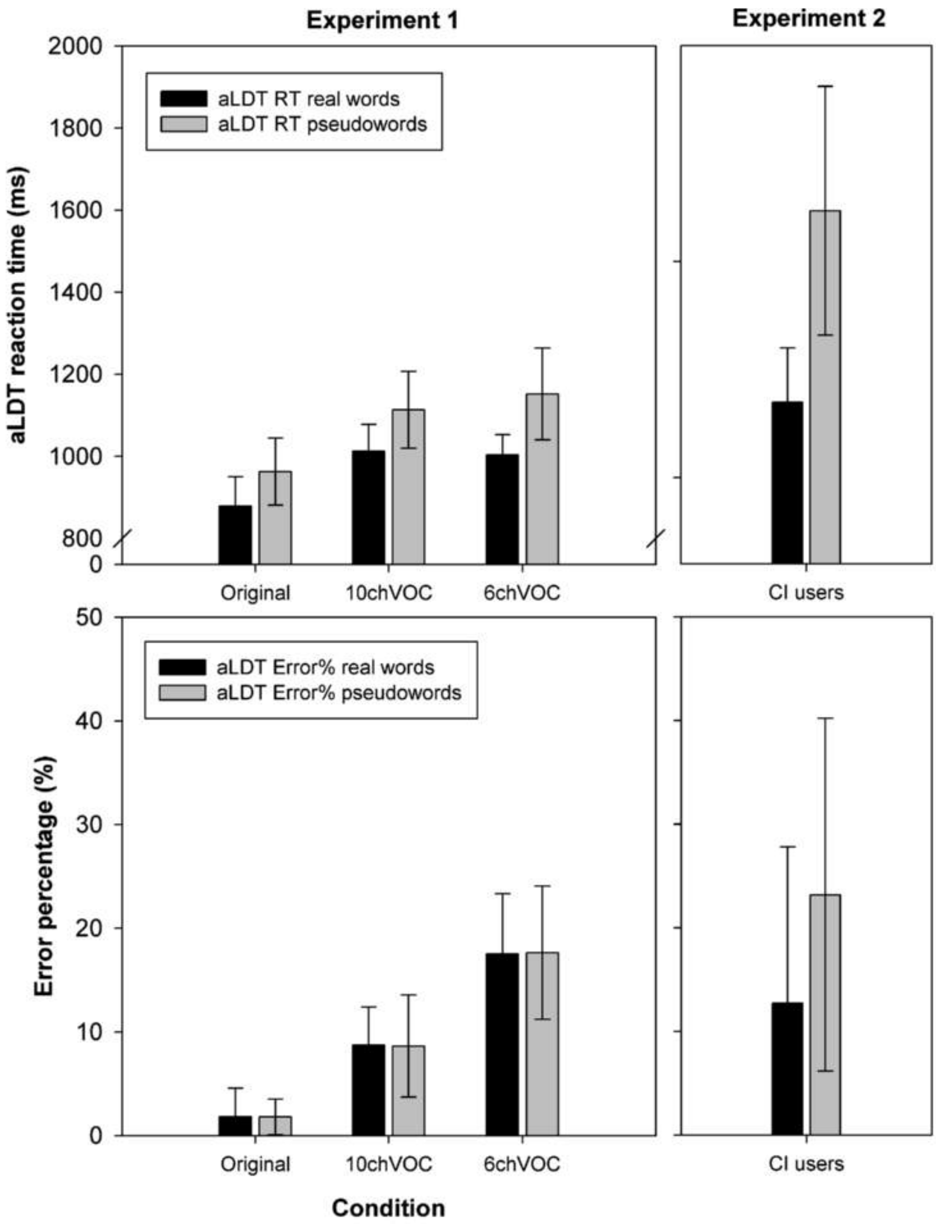
5

aLDT RTs (Figure 1a, Experiment 1) were compared with a mixed model with fixed effects for Condition (original, 10chVOC, 6chVOC) and for Stimulus type (real words, pseudowords) and two-way interaction effects. A main effect was found for Condition [Wald χ^2 = 101.0, df = 2, p < 0.001] and for Stimulus type [Wald χ^2 = 42.1, df = 1, p < 0.001], and an interaction for Condition × Stimulus type [Wald χ^2 = 6.6, df = 2, p = 0.037]. Post-hoc comparisons with Bonferroni correction for multiple comparisons showed that RTs for real words were, compared to original stimuli, significantly longer for 10chVOC and 6chVOC stimuli (respectively 134 ms and 125 ms, p < 0.001). RTs for real words were not significantly different between the two vocoded conditions (p = 1.0). Also, RTs for pseudowords were, compared to original stimuli, longer for 10chVOC and 6chVOC stimuli (respectively 150 ms and 189 ms, p < 0.001). RTs for pseudowords were also not significantly different between the two vocoded conditions (p = 0.052). The interaction between Condition and Stimulus type, however, indicated that the difference between RTs for real words and pseudowords increased when degradation increased.

aLDT Error% after transformation to RAU (Figure 1b, Experiment 1) were also compared with a mixed model with fixed effects for Condition (original, 10chVOC, 6chVOC) and for Stimulus type (real words, pseudowords) and two-way interaction effects. A main effect was found only for Condition [Wald χ^2 = 176.3, df = 2, p < 0.001]. Post-hoc comparisons with Bonferroni correction for multiple comparisons showed that the Error% for both real words and pseudowords were worse, compared to original stimuli, for 10chVOC and 6chVOC stimuli (respectively 7.0% and 15.8%, p < 0.001). The difference of Error% between 10chVOC and 6chVOC stimuli was also significant (p < 0.001).

To summarize, when stimuli were degraded to an extent resulting in speech-recognition scores that are comparable to those of good performers with a CI, response times increased and more decision errors were made. However, when the degree of degradation was further increased (comparable to reasonable performers with a CI) response times remained at the same level, while the error percentage almost doubled.

Visual lexical decision and ΔLDT. Mean vLDT RT for real words was 473 ms (SD = 49 ms) and for pseudowords it was 567 ms (SD = 47 ms). The vLDT Error% for real words was 3% (range 0 – 7.5%) and for pseudowords it was 11% (SD 0 – 26.9%). *T*-tests showed that the RTs for real words were in line with earlier results, while RTs for pseudoword were slightly shorter.



5

FIGURE 1. aLDT RT in ms (upper panels) and aLDT Error % (lower panels) for real words (black) and pseudowords (grey) for three conditions of speech stimuli in Experiment 1: unprocessed speech (US), 10 band vocoded speech (10chVOC) and 6 band vocoded speech (6chVOC) and for cochlear implant (CI) users in Experiment 2. Error bars represent 1 SD.

The RT differences between the visual and auditory modality, ΔLDT RT, for words (*M* = 406 ms, *SD* = 88 ms) and for pseudowords (*M* = 395 ms, *SD* = 81 ms) were not significantly different (t = 0.747, *p* = 0.472). Thus, with original stimuli, these young adult NH participants were on average 400 ms slower in the auditory modality than in the visual modality.

DISCUSSION EXPERIMENT 1

In this experiment we degraded the auditory input to the extent where speech-recognition scores resemble those of reasonable to good CI performance. The results of this experiment confirmed that this degradation results in slower responses in auditory lexical decision. In addition, it showed that participants made more errors when the amount of degradation increased. The RTs for the two vocoded conditions, however, were not significantly

TABLE 4. Characteristics of the CI users.

Participant nr.	Age [years]	Age at onset of SHL [years]	Duration of HL [years]	CI experience [years]	FF threshold [dB HL]
1	49	0	49	0.6	24
2	60	28	59	0.3	36
3	58	54	29	0.6	25
4	28	3	26	4	28
5	25	13	16	9	31
6	59	57	50	2	33
7	56	55	37	1	23
8	80	70	46	1.2	36
9	79	47	32	4	31
10	51	24	28	0.2	28
11	24	1	23	9	25
12	50	21	30	6	38
13	63	56	14	7	34
14	65	59	39	6	19
15	53	51	36	3	23
16	75	58	65	12	36
17	65	60	51	4	31
18	70	69	14	0.2	23
19	74	73	56	0.2	25
20	68	49	43	10	31
21	72	72	10	0.7	25
22	72	69	13	0.6	30

Note. Age at onset of severe hearing loss (SHL) and duration of hearing loss (HL) were obtained from questionnaires and medical records. SHL was defined as either not being able to use the telephone or aided monosyllable (CVC) recognition of less than 50% phonemes correct at 65 dB SPL. CI = cochlear implant different, suggesting that the degree of degradation does not relate to how much the responses are slowed down. Accuracy was, as expected, lower for the harder vocoder condition. This pattern of results suggests that participants changed their strategy for lexical decision when more degraded signals were presented. This was also suggested by the interaction between vocoder condition and stimulus type; that is, for the harder vocoder condition, the difference between the RTs for real words and pseudowords was larger than for the milder vocoder condition. Accuracy was the same for real words and pseudowords in both vocoder conditions. Thus, when examining the effect of signal degradation on auditory lexical access in listeners with hearing loss, the outcomes of this experiment indicate that RT and accuracy are both relevant. In some other studies, the percentage of correct answers divided by the RT is used as outcome measure for aLDT (i.e., Larsby, Hällgren, and Lyxell 2012). Such a quotient that reflects the trade-off between accuracy and RT, and perhaps reflects efficiency, might possibly be more related to speech-recognition performance in listeners with hearing loss. Therefore, this variable was included in Experiment 2.

EXPERIMENT 2

MATERIALS AND METHODS

The speech-recognition measures, lexical-access measures and general procedures were identical to the ones used in Experiment 1. In this experiment, only original unprocessed speech stimuli were used. All tests were administered two times.

Participants. Twenty-two postlingually deafened CI users (11 men, 11 women) participated. They were aged 24 to 80 years (M = 59 yrs). Table 4 lists characteristics of the participants. No dyslexia, reading problems, or relevant medical problems were reported in an interview prior to participation.

Participants were all patients of the Otolaryngology department, Section Ear & Hearing of the VU University Medical Center (VUMC), Amsterdam, The Netherlands. CI users with known relevant medical problems or prelingually deaf patients were not invited to participate. All participants used Cochlear devices (Sydney, Australia). During the tests, they used their device with the program and setting they preferred for everyday use. Participants using a contralateral hearing aid were asked to turn it off during the experiment. Audibility was checked by measuring sound field thresholds at 500, 1000, 2000 and 4000 Hz (M = 29 dB HL, SD 5.4 dB HL). All were native speakers of Dutch. Participants' vision, with corrective eyewear if needed, was also checked with a near-vision screening test (Bailey and Lovie 1980).

RESULTS

Speech recognition. Figure 2 shows the speech-recognition scores of the CI users. The mean CVCs score was 85% (SD 10%), the mean DIN threshold was 1.2 dB SNR (SD 3.8 dB SNR).

In line with Kaandorp et al. (2015) only SRTs of 15 dB SNR or lower were considered reliable. Some participants had only for one of their SIN test/retest a threshold of 15 dB SNR or lower, which was then taken as SIN result. Six participants had no SIN SRT of 15 dB SNR or lower, and were not included in the analyses. The mean SIN threshold for the 16 remaining participants was 10.6 dB SNR (SD 3.6 dB SNR).

Auditory lexical decision: comparison between groups. aLDT RTs (Figure 1a) of the CI users and NH listeners of Experiment 1 were compared (NH-VOC: 10chVOC and 6chVOC combined). A repeated measures analysis of variance with Stimulus type (real words vs. pseudowords) as within subject variable and Group (CI vs. NH-VOC) as between subject variable showed a main effect for Stimulus type [$F(1,42) = 86.1$, $p < 0.001$]. Responses to real words were faster than to pseudowords. An interaction was found between Group and Stimulus type [$F(1,42) = 27.1$, $p < 0.001$]. The difference between RTs for real words and pseudowords was much larger for CI users (444 ms) than for the NH group listening to vocoded stimuli (41 ms). Because CI users were generally slower, this difference between real-word and pseudowords RTs was also compared between groups after relating it to the overall RT for real words. An independent samples *t*-test showed that CI users also had a larger relative difference between real words and pseudowords than the vocoded group (respectively 0.31 and 0.11, $t = 4.987$, $p < 0.001$).

aLDT error percentages converted to RAU units were also compared between the two groups with a repeated measures analysis and showed a main effect for Stimulus type [$F(1,42) = 9.2$, $p = 0.004$]. More responses to real words were correct than to pseudowords. There was, however, also an interaction between Group and Stimulus type [$F(1,42) = 10.1$, $p = 0.003$]. For the NH group, there was no difference in error percentages between real words and pseudowords in the vocoder conditions, while CI users were 10.4% more accurate in real-word responses than in pseudoword responses.

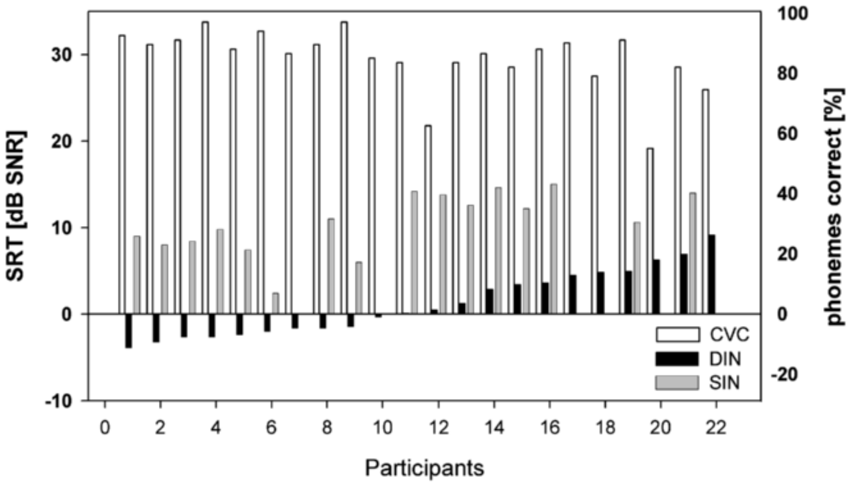


FIGURE 2. Speech-recognition scores per CI user in order of DIN performance (best performance on the left side). SIN and DIN SRTs in dB SNR and CVC scores in % correct phonemes.

Auditory versus visual lexical-decision. LDT RTs in the two modalities were compared**,** as we expected that effects of age and linguistic skills would be similar in both modalities and a difference measure would reflect the effect of degradation of the auditory signal. Mean vLDT RTs for the CI users were 579 ms (range 457 – 726 ms) for real words and 764 ms (range 564 – 1041 ms) for pseudowords. Their vLDT error percentage was less than 10% for both real words (M = 1.3%, *range* = 0 – 6%) and pseudowords (M = 6.5%, *range* = 0 – 15.6%). Figure 3 shows aLDT- against vLDT RTs for real words (upper panel) and pseudowords (lower panel) for both CI users and the NH listeners of Experiment 1. For the CI group, the Pearson correlation between aLDT- and vLDT RT was not significant (r = 0.39, p = 0.074) for real words, but it was significant for pseudowords (r = 0.54, p = 0.01).

Figure 4 shows the RT difference between modalities (ΔLDT RT) for real words and pseudowords. It shows that ΔLDT RTs for NH listeners for real words and pseudowords were equal, as was also reported in Experiment 1. For CI users ΔLDT RT was larger compared to the NH group and it differed between real words and pseudowords.

Correlations between speech recognition and LDT. As mentioned above, only SRTs of 15 dB SNR or lower were included in the analyses. This resulted in a group of 16 participants with speech-recognition data for all three tests. Based on the results of Experiment 1, aLDT efficiency measures (the quotient between percentage correct answers and RTs for real words, pseudowords, or overall responses) were also included as measures of lexical-access ability. Pearson correlations were calculated between the three speech-recognition measures and the aLDT, vLDT, ΔLDT, and aLDT efficiency variables (Table 5). Because a high number of correlations were calculated, the results have to be interpreted carefully as Type I errors might occur. Therefore, only correlations with a more conservative significance level of p < 0.01 are discussed. Significant correlations were found for CVC scores with aLDT RT for real words (r = -0.75, p = 0.001) and for SIN thresholds with the aLDT efficiency in pseudoword responses (r = -0.62, p = 0.01) and efficiency in all responses (r = -0.66, p = 0.005).

To summarize, first, response times to real words in aLDT were related to word recognition in quiet. Thus, CI users with poorer spoken-word recognition performance in quiet were slower in auditory lexical decision to words. Second, the auditory lexical-access efficiency in pseudoword responses and overall responses seemed to be more related to the more complex speech-in-noise performance. Thus, persons who have more problems understanding speech in noise are either more inclined to accept a pseudoword as a word, suggesting that they are used to guessing relatively often, or they need more time before making a decision.

TABLE 5. Pearson correlations between speech-recognition measures and measures of auditory lexical decision (aLDT) and visual lexical decision (vLDT) in CI users.

Variables (n = 16)		CVC	DIN	SIN
aLDT	RT [ms] real words	-0.75**	0.22	0.38
	RT [ms] pseudowords	-0.15	0.38	0.46
	RAU real words	0.33	-0.60*	-0.61*
	RAU pseudowords	0.14	-0.56*	-0.58*
aLDT efficiency	%correct/RT real words	0.55*	-0.52*	-0.56*
	%correct/RT pseudowords	0.22	-0.52*	-0.62**
	%correct/RT overall	0.39	-0.58*	-0.66**
vLDT	RT [ms] real words	-0.41	0.49	0.40
	RT [ms] pseudowords	-0.05	0.35	0.14
ΔLDT	ΔRT [ms] real words	-0.59*	-0.07	0.17
	ΔRT [ms] pseudowords	-0.17	0.29	0.54*

Note. RT, reactions time; RAU, rationalized arcsine units for accuracy; CVC, monosyllables DIN, digits in noise; SIN, sentences in noise. ΔLDT, RT difference measures between aLDT and vLDT variables. aLDT efficiency variables reflect the aLDT quotient between accuracy and response time.

* p < 0.05; ** p < 0.01

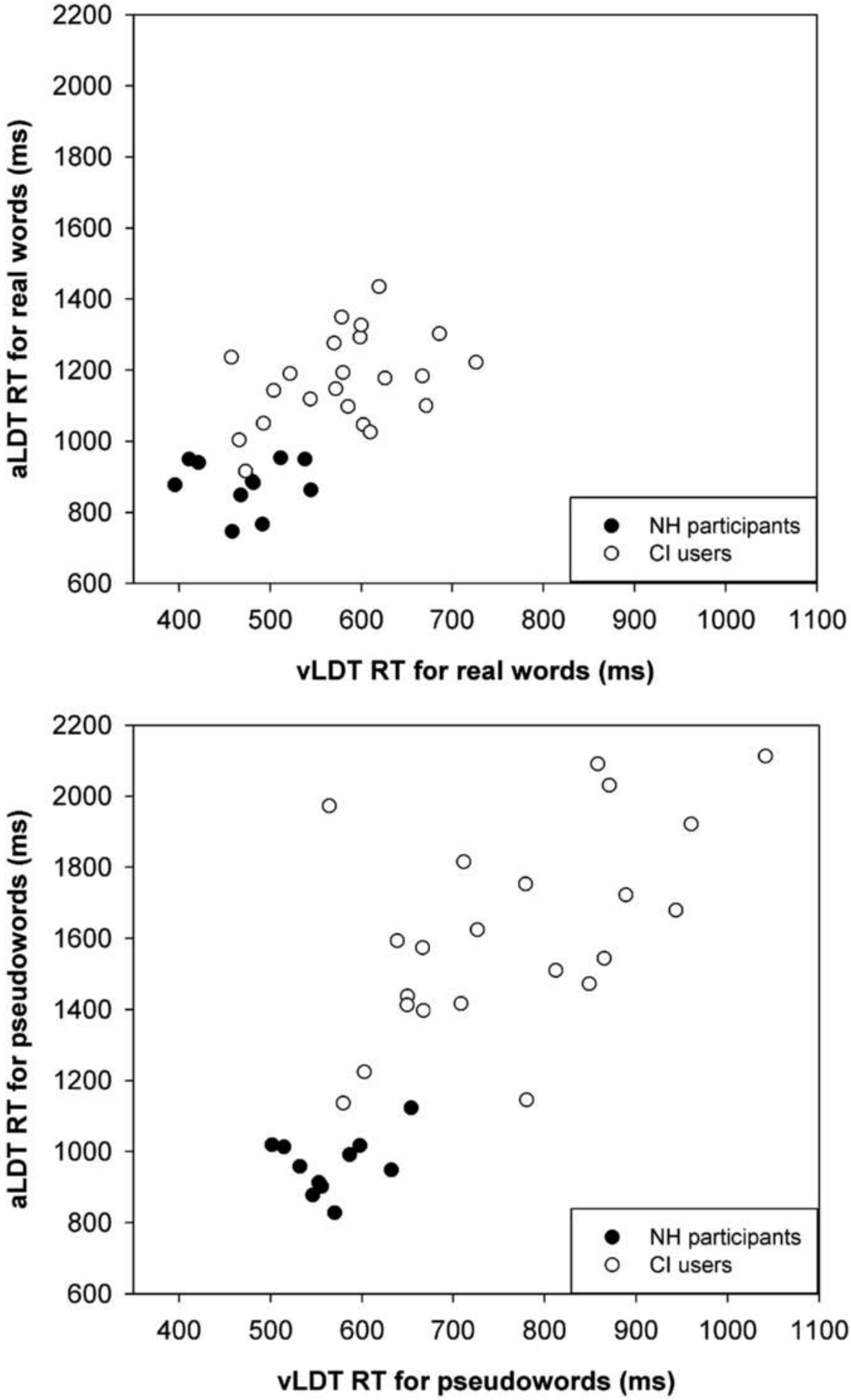


FIGURE 3. aLDT RT versus vLDT RT for real words (upper panel) and pseudowords (lower panel) for normal-hearing listeners with original stimuli, and for CI users.

5

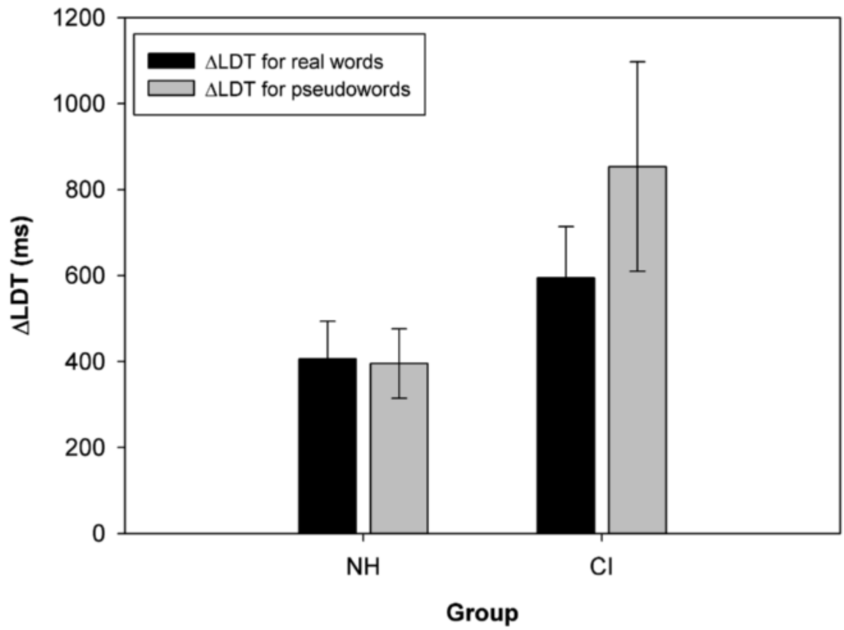


FIGURE 4 ΔLDT (aLDT-vLDT) for real words and pseudowords for two groups: normal-hearing listeners with original stimuli (NH) and CI users (CI).

DISCUSSION EXPERIMENT 2

RTs for auditory lexical decision were longer for CI users and response patterns were different than for the NH group listening to vocoded speech. The RT difference between pseudowords and words was much larger for CI users than for the vocoder group. This difference between groups plausibly resulted from the CI users' adaptation to the degraded input. CI users may always apply a longer time-deadline to give themselves more time to evaluate the uncertain input before they make a decision. Also, CI users were, as expected, more likely to accept pseudowords as real words. The underlying reason might be that they are used to guess and select the most probable target word when they mishear parts of the stimulus.

GENERAL DISCUSSION

The results of this study showed that young normal-hearing adults listening to vocoded stimuli and older CI users both had longer RTs and were less accurate in aLDT than the normal-hearing listeners listening to original stimuli. The CI users, however, showed different response patterns relative to normal-hearing listeners with respect to real word versus pseudoword RTs and accuracy. Above all, CI users were relatively slower in pseudoword responses and had a large bias toward real-word responses in contrast to the group listening to vocoded stimuli, that showed equal accuracy for both stimulus types. In other words, CI users were strongly inclined to accept the stimulus as a real word. Furthermore, we found that measures of speech in quiet and in noise were related to aLDT variables.

The effect of signal degradation on accuracy and RTs of aLDT is in line with previous studies, (e.g., Hallgren et al. 2001) that examined the effect of degraded input by adding noise to the stimuli. These results agree with our hypothesis that degraded input causes changes in the activation stage and in the cohort-evaluation stage because of spurious activation, leading to longer RTs for real words. They also agree with the hypothesis that the RT for pseudowords gets longer because listeners adopt a longer time deadline.

In Experiment 1, RTs were similar for the two vocoder conditions, which contrasts with the hypothesis that the RT would increase between the mild and harder condition. The study of Goy et al. (2013), who found longer RTs in aLDT in conditions with distorted primes than in conditions with intact primes, showed similar results. For 48 young adult listeners they found that RTs were similar for different amounts of distortion when the context was neutral, which could be compared to the current study without priming. In the current study, in contrast to the Goy et al. (2013) study with clear target words, accuracy was lower for the harder vocoder condition, indicating a difference in processing between conditions. This can have various reasons. Participants could have adopted a different strategy in the two vocoder conditions, because the instructions to the test were to respond as fast and accurately as possible. Perhaps, the increase in errors in the harder condition with equal RT is caused by the so-called speed-accuracy trade-off. In this case, more errors are made because listeners try to maintain their speed. Another reason for this difference in processing could be that putting more time in the lexical-decision process would not result in more accuracy. Norman and Bobrow (1975) state that a process can be limited in its performance either by limits in the amount of available processing resources (resource-limited; attention, working memory etc.) or by limits in the quality of the available data (data-limited). Most tasks are resource-limited until the test conditions are such that more processing will not improve performance, and from there on become data-limited. Applying this theory to the results of the vocoder conditions, we could say that more resources were applied in the mild vocoder condition than in the unprocessed condition, which was reflected by longer RTs. In the harder vocoder condition, however, RTs remained similar but accuracy dropped further, thus, all extra resources were already in use or applying more resources would not help to maintain the same accuracy level because processing was data-limited. Given that speech recognition of monosyllables was about 76% in the harder vocoder condition, this latter theory is more applicable to the results than an interpretation in terms of speed-accuracy trade-off.

In Experiment 1, average RTs were not significantly different between the two vocoder conditions for both real words and pseudowords. However, there was a significant interaction between stimulus type and vocoder condition. The difference between real word and pseudoword RT was larger for the harder vocoder condition. There was no difference in accuracy between stimulus types. When input is degraded, longer RTs for real words are plausibly caused by other changes in processing than longer RTs for pseudowords. We assumed that the RTs for real words were longer because of spurious activation of lexical elements that needed to be evaluated, while the RTs for pseudowords were related to the adopted time-deadline. Several studies have shown that listeners adjust this deadline to the test difficulty. When pseudowords deviate relatively strongly from the real words in a list, pseudoword RTs are generally shorter than when the words and pseudowords in a list are more similar (e.g., Ratcliff, Gomez, and McKoon 2004). When input is less certain, the perceptual deviation of pseudowords from real words will be reduced, and it can be expected that pseudoword RTs are longer compared to test runs with more certain input.

To conclude, first, for unexperienced listeners listening to degraded stimuli, RTs were longer for mildly degraded input, but the accuracy only decreased further for more degraded input, while RTs remained the same. We argued that in the latter condition, processing became data-limited. This change in strategy suggested that a quotient of accuracy and RT might give more information about the effect of the degree of degradation of the input signal and about efficiency in using top-down and bottom-up information. This aLDT efficiency measure was therefore included as an outcome measure in Experiment 2 when examining the relation between aLDT and speech recognition. Second, the interaction between real-word RT and pseudoword RT indicated that RTs for both word types were affected differently by signal degradation. Responses to pseudowords slowed down more with increasing degradation of the signal than responses to real words.

The aLDT results of the CI users in Experiment 2 correspond to the findings of Farris-Trimble et al. (2014), who found slower and less accurate responses in identifying the target word with their *visual-world* paradigm task for both CI users and age-matched CI-simulated listeners. They also found a difference in response pattern between these groups, which, as they suggested, was caused by maintaining activation for competitors for later revision of interpretation. This might lead to the observed longer RTs for identifying the target word (real word RTs). One could argue that if CI users adopt this strategy, they probably also adopt a longer time deadline in the decision stage to allow more time analyzing the cohort. This latter hypothesis is supported by the finding of the present study that CI users showed a larger difference between RTs for real words and pseudowords than NH listeners listening to vocoded stimuli. This indeed suggests that the long-term adaptation to the degraded input results in a different processing strategy in CI users as compared with naive listeners.

The CI users were less accurate in responses to pseudowords than to real words, which confirmed our hypothesis that CI users have an accuracy bias toward words. These results correspond to the study of Hallgren et al. (2001), who found that listeners with hearing loss made more errors in non-word responses. They argued that in the aLDT, when receiving incomplete information, subjects with hearing loss favor word decisions because those are closer to their real life situation. In a study with children, Pittman and Rash (2016) used an auditory lexical-decision task to investigate their ability to identify words they did or did not know. They found that, when listening in noise, both children with normal hearing and those with hearing loss often misperceived nonwords and mistakenly accepted nonwords as real words. Only the children with hearing loss also made these errors in quiet. In another study, Pittman and Schuett (2013) found that children with hearing loss were more likely to underestimate the number of nonwords in a sentence compared to children with normal hearing, suggesting that children with hearing loss apply strong repair strategies during speech perception. Thus, with degraded input, children tend to change words they do not know into words they do know, which likely also applies to adults that are used to receiving uncertain input.

Because the CI users in the current study had a larger age range and were on average older than the young NH group, age might have had an effect on the results. For instance, Goy et al. (2013) found that older listener did have significantly longer aLDT RTs with more distortion of the primes, in contrast to the younger listeners. They argued that their young participants might have adopted a more automatic listening strategy, where the older participants might have adopted more controlled processing strategies. Other studies, however, have shown differences in processing between CI users and age-matched listeners to CI-simulations (e.g., Farris-Trimble et al. 2014), suggesting that differences in processing are the result from adaptations to long-term degraded input. To examine a possible effect of age in the present results, we split our CI group in the youngest half and oldest half. A comparison of results between these age groups showed that the youngest half had slightly shorter vLDT RTs, but their aLDT RTs were similar to the oldest half for responses to words and even longer for responses to pseudowords. Also, the youngest group made more errors and had a larger bias toward words. Thus, age does not explain the differences in aLDT responses between the CI users and NH listeners in the present study, but an interaction between age and effect of degradation cannot be excluded.

Thus, CI users seem to use a different processing strategy in auditory lexical decision than naive NH listeners listening to clear input. For CI users, both the RT and accuracy of pseudoword responses were mostly affected by signal degradation. These variables can, therefore, give other information about the degree of experienced difficulty in listening and how CI users deal with it, than the RT and accuracy of the real-word responses.

COMPARISON OF AUDITORY AND VISUAL LDT RTS

Lexical access was measured and compared in the visual and auditory modality to gain more insight in the effect of degraded input on the lexical-access process. For NH listeners (Experiment 1), overall RTs for unprocessed real words and pseudowords were 400 ms shorter in the visual than in the auditory modality. Hallgren et al. (2001) also used a LDT in both the visual and auditory modality and found a difference of 541 ms for young NH listeners and 483 ms for elderly NH listeners. This difference between modalities is at least partly inherent to the different test procedures and setup. RTs were measured from the moment the word appeared on the screen (visual) or from word onset (auditory) until the response button was pressed. Due to the sequential nature of the auditory stimulus, not all information is available immediately from word onset. In contrast, in visual word presentation, all information is available instantly. Thus, RTs will evidently be longer in aLDT compared to vLDT. Some studies therefore measure aLDT RTs starting from the uniqueness point. For instance, in a study using the *visual-world* eye-tracking paradigm, Dahan and Tanenhaus (2004) found that participants needed on average about 350 ms from the auditory word onset before eye fixations to a cohort competitor started deviating from fixations to the target. Thus, the uniqueness point of their auditory target words was reached about 300-400 ms after word onset. Similarly, Grosjean (1980) used an auditory gating technique in which words were presented repeatedly in fragments of increasing size from word onset. He identified mean uniqueness points for higher- and lower-frequency words of several lengths, with a global mean of 333 ms. Using the uniqueness point is especially important when investigating effects of structural differences among stimulus items (Goldinger 1996). In the current study, however, we were interested in differences between listeners and between conditions. For the comparison of vocoder conditions, balancing of the uniqueness points of the stimuli might have given a more precise result than the used balancing on word length and concreteness. Such an approach would, evidently, annihilate a large part of the time difference that we found between modalities. The factors mentioned above explain the smaller differences between visual and auditory LDT RTs that are found in other studies. For instance, measuring RTs from the earliest point at which the final phoneme could be identified, Turner, Valentine, and Ellis (1998) found a difference of only 23 ms between the two modalities. When applying the average uniqueness point of Dahan and Tanenhaus (2004) and Grosjean (1980) to the results of the present study, a difference of approximately 50-60 ms between modalities would remain.

Another relevant difference between the used test procedures in vLDT and aLDT is that more real words than pseudowords were used in the vLDT. This choice was based on the assumption that pseudowords are perceptually more salient than real words, and therefore a list with equal distribution might result in a bias toward pseudoword responses. The current results, however, showed a bias towards word responses in the vLDT. This finding suggests that a smaller difference between word types is probably better. In the aLDT, equally large sets of real words and pseudowords were used, because we assumed that this saliency of pseudowords would be less prominent with degraded stimuli. Since the difference between pseudoword and real-word responses within vLDT was not included in the analyses, we expected this difference in setup had a negligible impact on the current results.

The study of Hallgren et al. (2001) suggested that there was a modality-independent decline in the speed of performance with age. This would imply that the age effect on RTs is similar in vLDT and aLDT and is thus eliminated in the ΔLDT RT variables. We expected that linguistic abilities would also have similar effects on aLDT and vLDT scores. The significant correlation between aLDT RT and vLDT RT for pseudowords (see Figure 3) showed that the time-deadline used for pseudoword decisions in vLDT is a predictor of that in aLDT. The modality differences in ΔLDT that were present between the, on average, older CI users with various levels of education and the high educated young adult group listening to unprocessed speech, could thus be attributed primarily to the effect of degraded input and not to effects of age or linguistic skills.

LEXICAL DECISION AND SPEECH RECOGNITION OF CI USERS

Correlation analyses only revealed significant correlations (with p < 0.01) between, first, aLDT RT for real words and recognition of monosyllables in quiet (r = -0.75, p = 0.001). Poorer CVC recognition in quiet was, thus, mainly related to longer response times to words. This could be expected, because for listeners with poorer CVC recognition ability there is probably more spurious activation during lexical access, resulting in a larger cohort of competitors. Consequently, more time is needed for evaluation of the cohort, resulting in a longer period before the critical activation threshold for the target word is reached. Secondly, lexical access efficiency in pseudoword responses and in overall responses, reflected by the quotient variables, were related to sentence-in-noise recognition (r = -0.62, p = 0.01 and r = -0.66, p = 0.005 respectively). Although the level of significance was lower for other variables, a similar pattern was found for the separate RT and accuracy measures and the modality difference ΔLDT RT. In the current study, vLDT measures, reflecting general processing speed and linguistic skills, were not correlated to speech-recognition measures.

Larsby et al. (2012) also compared the overall quotient as the outcome measure of the lexical-decision test with speech-recognition scores in 40 unaided listeners with mild to moderate hearing loss, ranging from 46 to 86 years. After correcting for pure-tone hearing loss, they found a significant correlation of -0.37 between this overall aLDT quotient and the outcome on the Swedish HINT test. In the CI users of the current study, audibility was ensured by measuring sound field thresholds. Sentence-in-noise thresholds in the present study seemed more related to (efficiency of) pseudowords responses than to real word responses. Thus, participants who used a relatively long time deadline, or relatively often accepted pseudowords as real words, had more problems recognizing speech in noise (higher SRTs) than participants with a shorter deadline or fewer false positives to pseudowords. Possibly, listeners who need a longer time deadline in lexical decision because of the degraded input, have more problems in recognizing longer segments of speech (as in the sentence-in-noise test), where time is more limited. Furthermore, many parts of the stimulus are uncertain when presented in noise, which might be reflected in a larger bias toward words as well as in higher SRTs. That is, listeners who are accustomed to poor quality of speech input are probably more used to guessing than listeners without hearing loss, which is reflected by a higher frequency of accepting a pseudoword as a word (lower accuracy for pseudowords responses and a lower quotient).

To sum up, the present results suggest that aLDT real-word responses and pseudoword responses relate to different aspects of speech recognition. It would be interesting to study the LDT measures in a larger study group to gain more insight into the value of these measures in personalizing rehabilitation programs.

CLINICAL IMPLICATIONS

In this study, we used lexical decision as a measure of lexical access for two reasons: First, we intended to compare the outcome of this test in the auditory and visual modality (which can also be measured pre-operatively in CI users) and, second, we wanted to use a test that is easy to use in the clinic. A drawback of the LDT is that it captures not only lexical-access ability, but also other processes that are needed to respond to the given input, like pressing the button. Comparison of LDT in the auditory and visual modality could partly eliminate these factors. Another way to gain a more pure measure of lexical access is by looking at the shared variance of multiple tests that share a common lexical-access component while all are measuring other task specific processes (Kaandorp et al. 2016).

Some studies suggest that LDT RT measures may provide a measure of the amount of listening effort even when accuracy of word recognition is high (e.g., Goy et al. 2013). Longer RTs are assumed to reflect the use of more resources by the listener. For instance, Pals, Sarampalis, and Baskent (2013) used RTs on a secondary task to measure listening effort. They found constant speech-recognition performance between 6- and 8-channel vocoded speech, while RTs, improved for the 8-channel condition, reflecting less listening effort. In contrast, in the current study, we found that accuracy improved for the 10-channel noise-vocoded speech condition, while RTs were not significantly different between the 6- and 10-channel conditions. More research of these measures in combination with established measures of listening effort, such as pupillometry, should be done to examine the suitability of aLDT for evaluation of differences in listening effort in different settings of CI's or hearing aids.

Although more research is needed in larger participant groups to obtain a better understanding of the relation of the separate LDT measures with speech recognition, the current results already provide some information for use in the clinic. vLDT was not correlated to speech recognition in the present study sample, but previous findings suggest that poor vLDT scores do limit CI performance (Kaandorp et al. 2017). Poor aLDT RT for real words might point to a poor auditory input, especially when vLDT RT are normal and DIN scores are also poor. For these listeners processing is probably data limited and rehabilitation should focus on enhancing the quality of the bottom-up stimulus, either by fitting or additional technology. Poor aLDT RT for pseudowords, especially combined with accuracy (aLDT efficiency for pseudowords), and overall aLDT efficiency seem to provide information about dealing with the poor input in longer speech segments and thus auditory performance in daily life. For these listeners processing may be resource limited and rehabilitation should focus on strengthening the use of linguistic and cognitive resources.

CONCLUSIONS

Degraded auditory input resulted in slower responses and lower accuracy in the lexical-decision test. CI users seem to have adopted a different strategy in lexical decision than normal-hearing participants listening to vocoded speech, possibly because they are used to listening to degraded stimuli. For CI users, speech recognition in quiet was mainly related to auditory response times for real words. More complex recognition in noise was more related to the efficiency in auditory pseudoword responses and overall response efficiency. These results showed that measures of auditory lexical decision combined with measures of visual lexical decision can be helpful in identifying whether speech-recognition performance is data limited or resource limited and hence in determining the rehabilitative approach.

6

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a cochle … uch better
that expl … s wide … ange of perf
needed t … rove preoperative co
rehabilit … is thesis aim
betwee … sally ap … le linguis
performa … ith a coc … implant. T
skills that … n people w
of current clini … asures of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

General discussion

Speech-recognition abilities of cochlear-implant (CI) users are determined by auditory, linguistic/cognitive, and biomedical factors. Within this framework, the studies described in this thesis aimed to gain a better understanding of the influence of linguistic skills, specifically vocabulary size and lexical-access abilities, on speech-recognition performance in CI users. Clinical measures of speech-recognition performance were evaluated first in a validation study of the Dutch digits-in-noise test (DIN test). Next, the influence of linguistic skills on speech-recognition abilities was examined with visual tests of lexical-access ability and vocabulary size in NH listeners, with a large variation in linguistic/cognitive factors, and subsequently in CI users. The effect of linguistic skills on speech recognition of CI users was compared to the effect of other, more commonly, visually measured verbal cognitive abilities. The conclusions from these studies led to the study described in Chapter 5 on the effect of signal degradation on auditory lexical-access ability either with vocoded speech or by CI-use. In this final chapter, the main findings of the studies are brought together and discussed, and suggestions for further research and clinical implications are provided.

OUTCOME MEASURES OF SPEECH RECOGNITION IN CI USERS

Standard clinical measures of speech recognition are important for comparisons between listeners and for evaluation or follow-up of hearing device fitting in clinical practice. These tests should be easy, quick, have good reproducibility, minimal floor or ceiling effects and should measure clinically relevant factors. In all four studies of this thesis tests that are currently used in Dutch clinical practice were utilized to assess the ability to recognize speech in quiet and in noise. Based on clinical experience, the word recognition test with phoneme scoring (CVC, Bosman and Smoorenburg 1995), currently the Dutch standard test for evaluation of CI outcome, was expected to show ceiling effects for high performing CI users. The results confirmed this (Chapters 2, 4, and 5) and showed, at the same time, that high CVC scores not necessarily indicated good speech-in-noise recognition. The use of CVC words as CI outcome measure, thus, limits evaluation of functioning in daily life listening situations, comparisons between listeners, and follow-up of rehabilitation progress. Therefore, it is important to also measure speech recognition in noise performance. Adaptive speech-in-noise tests have proven valuable tests for listeners with normal hearing or mild hearing loss without floor- or ceiling effects and with good test-retest reliability (e.g., Plomp and Mimpen 1979; Nilsson, Soli, and Sullivan 1994). The standard adaptive sentences-in-noise test, however, showed to be difficult for a large part of the CI users and for some hearing-aid users with a severe hearing loss (Chapters 2, 4, and 5). A first reason is that it requires good recognition of sentences in quiet, that is, >about 75%, to enable adaptive testing in noise, a performance level that is not always reached in this population. Secondly, adding noise to the speech stimuli further reduces the quality of the input signal, further increasing the demand on good auditory processing and on linguistic and cognitive abilities. For high signal-to-noise ratios (SNRs), the adaptive procedure does not work properly and scores do not reflect the ability to understand speech in noise. There is no agreement on a criterion to judge reliability of SRTs. Therefore, we introduced a criterion based on the SII (ANSI 1997). Scores were considered unreliable when SRTs were > +15 dB SNR and were excluded from some of the analyses for part of the participants. In this thesis we evaluated two methods to broaden the range of patients that can be tested with adaptive speech-in-noise testing: we introduced a different scoring method (keyword scoring) in the SIN test to allow for small mistakes and to improve SIN scores (Chapter 3) and we used a second test with speech material that puts a lower demand on linguistic and cognitive ability, the DIN test.

The comparison of keyword- and sentence-scoring in NH listeners showed that, on average, listeners obtained a 1-dB better SRT with keyword scoring. The effect was largest for the non-native normal-hearing listeners (1.8 dB). Nevertheless, the SIN test with keyword scoring is applicable to a wider range of listeners, likely enabling reliable test results for more CI users, and it is less sensitive to differences in linguistic ability.

The relatively new DIN test (Smits et al., 2013) proved suitable for measuring speech recognition in noise in the clinic (Chapter 2) in hearing-impaired listeners using hearing aids or CIs. The high correlation between the DIN and SIN tests showed that they largely measure the same auditory abilities. The relation between DIN and SIN was slightly different for the CI users with reliable SIN thresholds than for HI users and NH listeners. This could be the result of differences in signal processing between CIs and hearing aids, but might also be the result of differences in personal characteristics or hearing loss.

Taken together, the results showed that it is important to add speech-in-noise tests to the test battery that measure relevant capacities for listening in daily life situations. The DIN and SIN tests can further differentiate between listeners that perform at ceiling level on the CVC test and can be suitable to follow-up progress within rehabilitation periods. This is further discussed in the Clinical Implications Section.

MEASURES OF LINGUISTIC SKILLS

In this thesis lexical access and vocabulary size were considered to be aspects of linguistic ability that are important for speech recognition in noise. The motivation for this choice was that words convey most of the meaning of a language utterance.

In the following paragraphs, the results of the studies are discussed in view of a simple model of lexical access and lexical decision (see introduction). This model assumes that with the onset of word input multiple candidates are activated. The cohort of activated candidates that results is then evaluated until a decision about the target item can be made. In lexical decision a word response follows when a word's level of activation has reached a critical level of difference to that of its competitors and a pseudoword response follows when a preset time-deadline has passed.

Lexical access in the visual modality. In line with the used model, lexical-access RT was influenced by word frequency, a factor that determines baseline activation levels of (competitor) words in the cohort. The used vLDT test proved sensitive to word frequency as provided in a Dutch language corpus, as measured within listeners (Chapter 3). The vLDT test could therefore be expected to be sensitive to differences in lexical-access ability among persons for whom word frequency (familiarity with a word) might differ. Indeed, in the study that tested listeners with normal hearing but differing widely in language proficiency (Chapter 3), average lexical-decision RTs were shorter for high-educated participants than for participants that were expected to have lower linguistic skills, that is, lower-educated native- and high-educated non-native participants. Although it might seem plausible that a larger vocabulary size requires a longer evaluation phase, RTs were shortest in the highly-educated group with the largest vocabulary size. Furthermore, the results showed that although the lower-educated native- and high-educated non-native participants had the same vocabulary size, the latter had longer vLDT RTs. Both results correspond to the assumption that for listeners with more access to a language, pre-set activation levels of familiar words are higher than for listeners with less access to a language, resulting in a quicker match to a lexical representation in the mental lexicon. Thus, the average vLDT RT proved suitable to indicate differences in linguistic abilities and showed to measure different aspects of linguistic ability than vocabulary size.

The selection of young participants in the normal-hearing study aimed to cover the wide range of linguistic skills as might be found in postlingually deafened CI users. vLDT scores of CI users were, indeed, in the same range as those of the NH listeners. On the other hand, a drawback of selecting only young participants is that the effect of age cannot be quantified. For CI users, who were on average older than the NH listeners, part of the slower responses are not a result of a lower quality of information processing with older age, but likely originate from slowed motor movement while pressing the button (Ramscar et al. 2013). A composite measure of vLDT and word naming (LA), was assumed to be a more pure measure of lexical access, with less influence of such task-specific processes (see Chapter 3). For the participants of Chapter 3 with a large range of language proficiency, the correlation between vLDT and word naming was indeed high, indicating a large shared variance. However, for a sample with only native listeners the word-naming test was less discriminative. This will be further discussed in the Methodological considerations section.

Lexical access in the auditory modality. After examining visual tests of lexical access to avoid interaction with hearing status, this interaction itself was subject of investigation in the final study (Chapter 5) of this thesis. In aLDT, in contrast to vLDT, not only linguistic skills determine RTs but also uncertainty caused by degraded input. The effect of degraded input on lexical decision was examined first in young university students with normal hearing, to avoid interaction with biomedical or linguistic/cognitive factors. RTs for both words and pseudowords in aLDT were longer for two vocoder conditions compared to the original stimulus. In view of the adopted model of lexical access, this confirmed the hypothesis that uncertainty in the input signal might cause more spurious activation of lexical items in the cohort and result in both a longer evaluation phase and, consequently, the use of a longer time-deadline. Accuracy was expected to be near ceiling in vLDT, since all words were highly familiar words. In contrast, in aLDT, accuracy was also determined by perception of the input, as reflected by a lower accuracy for the vocoder conditions compared to the original stimulus, and should reflect wrong guesses when the input was not clearly perceived. Examination of the effect of degree of degradation of the input signals on accuracy and RT, revealed more errors for the harder vocoder condition compared to the milder condition, but not slower responses. We discussed in Chapter 5, that possibly, in the harder vocoder condition, adopting a longer evaluation phase would not result in more accuracy, because there is not enough information available to use more resources on. Norman and Bobrow (1975) state that cognitive processing can be limited in its performance either by limits in the amount of available processing resources (resource-limited) or by limits in the quality of the available data (data-limited). This difference in limitations of processing seems important, because hearing- impaired listeners might try to use more resources when listening situations increase in difficulty. However, from some point onward the quality of perceived information is limiting and the use of more resources might not result in better performance. Thus, in the auditory modality, in contrast to the visual modality, accuracy of aLDT responses is mostly determined by the quality of speech input. Additionally, the balance between accuracy and RT seems important to describe the effect of degraded input on efficiency in auditory lexical access.

CI users showed a different pattern of responses than NH listeners listening to vocoded speech. Interestingly, they had a bias toward word responses, which was not present in the responses of listeners with normal hearing listening to degraded speech. This is in line with findings of previous studies on aLDT in listeners with hearing impairment (e.g., Hallgren et al. 2001). CI users were also slower in pseudoword responses relative to word responses than the normal-hearing listeners. This suggests that CI users are more inclined to accept a pseudoword as a word, suggesting that they are used to guessing relatively often or that they need more time before they make a decision. This different pattern in processing might involve an adaptation to degraded input caused by long term hearing loss.

THE INFLUENCE OF LINGUISTIC SKILLS ON SPEECH RECOGNITION

The main question of the studies described in this thesis was: Is speech-recognition ability with a cochlear implant influenced by linguistic abilities as measured with vocabulary size and lexical-access measures and how do they relate to other cognitive measures?

The effect of visually measured linguistic abilities on speech recognition. The effect of linguistic abilities on speech recognition was examined first with visually measured tests in Chapters 2, 3, and 4 to avoid interaction with auditory capacity and to evaluate suitability for pre-operative testing. The results of the study with three groups of NH listeners with a wide range of linguistic skills, including non-native listeners (Chapter 3), showed that lexical access, a more fluid cognitive ability, and, to a lesser extent, vocabulary size, a crystallized cognitive ability, were predictors of speech recognition in noise.

For CI users the relation between lexical access and speech recognition in noise was less clear, because auditory factors were the main predictors of the results. A comparison of the role of lexical access and vocabulary size with the role of other cognitive measures showed a relatively stronger impact of working-memory capacity on speech recognition than of the measured linguistic abilities. This influence was actually quite large and corresponds to the results of several studies on the role of working memory in speech recognition in listeners with hearing loss (for a review see Besser et al. 2013). Although widely accepted as important for speech recognition, a meta-analysis by (Fullgrabe and Rosen 2016), recently pointed out that the relation between working-memory capacity and speech recognition in noise seems to be dependent of hearing status and age of the listener. Their analysis revealed that for young NH listeners working-memory capacity predicted less than 2% of the variance in speech recognition in noise. Nonetheless, they suggest that for older hearing-impaired listeners, like most CI users in this thesis, this measure may be helpful in improving the prediction of speech recognition in noise. Working-memory capacity is probably important in sentence-in-noise recognition because with degraded input a listener needs to be able to reconsider speech segments in memory when more information becomes available. For CI users the speech signal is degraded even in quiet environments. The ability to recognize visually masked sentences with the TRT test, was no relevant predictor in this study in contrast to the study of Haumann et al. (2012) and a recent study of Rosemann et al. (2017) with NH listeners listening to CI simulations. More research with different test procedures and larger sample sizes should be done to further investigate the predictive value of the TRT.

Importantly, some CI participants were very poor performers on speech-in-noise tasks, although they mostly had good linguistic and cognitive skills. SIN performance was not evidently related to their CVC scores, but DIN performance was also very poor for most of these listeners. Furthermore, inspection of biomedical and linguistic or cognitive abilities of these cases showed no clear limiting factors. For these listeners the quality of the input might be too poor to enable the use of top-down processing (i.e., for them processing is data-limited). Results of some other studies also suggest that the relation between linguistic and cognitive ability, on the one hand and speech recognition on the other depends on the quality of the input (Baskent, Eiler, and Edwards 2010; Collison, Munson, and Carney 2004).

The results of the present studies indicate that visually measured linguistic skills largely explain sentence-in-noise performance of listeners with normal hearing. In addition to auditory factors, these skills are also relevant for sentence-in-noise performance of CI users. Working-memory capacity was the strongest predictor of post-operative sentence-in-noise recognition that can be used pre-operatively. These measures may thus provide information for counselling pre-operatively and can help explain performance post-operatively.

The effect of auditorily measured linguistic abilities on speech recognition. The relation between linguistic skills and speech recognition is complex. The results of Chapter 3 replicated that for normal-hearing young listeners linguistic skills have a large influence on sentence-in-noise recognition. However, the discussed results indicate that when the quality of the auditory input decreases, the extent to which linguistic resources can be used might change. Also, long term adaptation to degraded speech signals seems to change auditory lexical access in CI users.

In the auditory modality the *efficiency* of processing in lexical decision (in terms of the quotient between accuracy and RT) seemed relevant and was examined in addition to RT and accuracy variables for vLDT and aLDT. For CI users, speech recognition of words in quiet was mostly related to aLDT RTs for real words. The more complex sentence recognition in noise was mostly related to aLDT efficiency in processing pseudowords and overall aLDT efficiency. Thus, an efficient balance between response accuracy and RT helps to better understand sentences in background noise. This seems plausible since running speech does not allow for long processing of separate words.

The modality difference measures ΔLDT (aLDT-vLDT) were also significantly related to CVCs and SIN thresholds, although with a less strict level of significance ($p < 0.05$). By using the delta measures effects of factors that influence LDT RTs in both modalities (e.g., age) are excluded. Interestingly, ΔLDT RT for words was related to CVC scores, and ΔLDT RT for pseudowords to SIN scores. Possibly, word responses are more related to bottom-up effects of degradation, whereas pseudoword responses are more related to top-down processing. Listeners who in a lexical-decision test adopt a longer time deadline because of the degraded input, have more problems in longer segments of speech (as in the sentence-in-noise test), where time is more limited.

Thus, measures of auditory lexical decision, together with the earlier measures of visual lexical decision, can help in understanding speech-recognition outcome with CI use, especially in differentiating between bottom-up and top-down factors. Although, more research is needed in larger participant groups to obtain a better understanding of the exact relation between the separate LDT measures and how they relate to speech recognition.

METHODOLOGICAL CONSIDERATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In all studies the SIN test was used, because it is the standard speech-in-noise test in Dutch clinical practice. In three of these studies some of the participants with hearing loss were not able to perform the SIN test or had unreliable thresholds. It proved difficult to deal with these unreliable, and always poor thresholds. Although the used upper limit of +15 dB SNR may be a too strict criterion for hearing-aid or CI users, there is no commonly used alternative, therefore it seemed the most reasonable one to use on our data. Recently, other studies have used the same criterion, indicating that they encounter similar problems for which they did not find a better approach (Plant et al. 2016; Devocht et al. 2016). To overcome these difficulties in future research on listeners with a wide range of auditory acuities, other speech-in-noise tests or test methods could be considered, for instance, measuring at two or three predefined and fixed SNR's. Alternatives, however, also have their own drawbacks, like lower test-retest reliability, floor and ceiling effects, and difficulty to compare participants over a wide range of performance or for follow-up of rehabilitation progress. Future research should also examine the exact SNR limit for the SIN test where the noise still influences speech-recognition performance of CI users.

The vLDT test and word-naming test both measure different test-specific factors besides lexical access and, therefore, converging evidence is preferred to obtain a more accurate estimate of lexical access. In Chapters 3 and 4 a quick and easy word-naming test was used next to the vLDT and results were combined in the composite measure LA. The simple word-naming test was chosen for its ease and quick application, but it appeared to be less discriminative among native listeners than the vLDT. For future research a more precise word-naming test would probably give more representative results, for instance a test where RTs are measured for each test-item separately by means of a voice-key, that measures RT until a voice is detected, that way partly excluding production time. Also, other tests of lexical access could be evaluated (such as progressive demasking, Grainger and Segui 1990; rhyme judgment, e.g., Lyxell et al. 1996) for their suitability and clinical applicability together with lexical decision.

In the studies in this thesis visual lexical access was assumed to be more or less stable before and after implantation. However, lexical-access ability is influenced by word frequency. This effect is not only determined by frequency of occurrence in a language, but also by frequency of word exposure to a reader or listener. For some listeners with a long duration of hearing loss, the effective frequency of hearing everyday words increases after implantation. It would be interesting to measure lexical-decision performance before implantation and after implantation (in both modalities) at several intervals to examine changes in lexical-access efficiency. Following this line of thinking, it would also be very interesting to examine whether lexical-access efficiency could be trained, for instance with special reading or listening programs, and if it could be trained, whether this would improve speech-recognition performance.

In Chapters 3 and 5 different groups of participants were included to obtain variation in one area, but keep other areas constant. In Chapter 3 these were the three groups of listeners with a wide range of variability in linguistic/cognitive skills, but who were homogeneous in hearing and biomedical factors. In Chapter 5 these were the normal-hearing listeners listening to vocoded speech, to control auditory input and keep linguistic/cognitive and biomedical factors constant. This approach was successful in examination of the separate impact of linguistic/cognitive factors and auditory factors on speech recognition. Nevertheless, this approach also has its drawbacks. With these homogeneous groups not all questions can be answered. In addition to the present studies, it would be interesting to investigate the role of auditory and visual lexical decision on speech recognition in a group of NH listeners that are matched with respect to age and education to the CI users. To better understand the interaction between linguistic and cognitive ability with speech-recognition performance also more groups of participants with different auditory capacity and/or speech-in-noise tests with different demands could be examined. When the specific effects of age and linguistic/cognitive skills need to be disentangled, groups of participants with different ages with similar auditory and linguistic/cognitive abilities should be compared.

In the previous section we discussed that the relation between linguistic ability and speech recognition is complex. In some listening situations the role of linguistic/cognitive skills is less clear, because processing might be data-limited, and the use of linguistic and cognitive skills might therefore not be effective. Some studies have used measures of speech recognition in noise at a more favorable SNR than the common SRT for 50% correct responses. For instance Larsby, Hällgren, and Lyxell (2012) also measured the SRT for 80% correct responses. They concluded that this easier setting seemed more sensitive to cognitive capacity because in the difficult setting not enough auditory information is available to use linguistic or cognitive skills. Thus, including speech-in-noise test settings that are less difficult presumably will give a more complete view on the role of linguistic/cognitive abilities.

CLINICAL IMPLICATIONS

The studies described here give valuable information to consider when measuring speech-recognition performance and relevant linguistic skills in clinical otological and audiological practice. The data confirm and quantify that linguistic and cognitive abilities are important factors in speech-in-noise performance, although auditory capacity remains the main predictor of results in listeners with hearing loss or using a CI.

When CI candidacy is considered, visual measures of linguistic and cognitive ability (lexical-access ability and working-memory capacity) could help in managing expectations. Slow visual lexical-access ability or poor working-memory capacity potentially limit performance with CI. On the other hand, efficient lexical-access ability and good working-memory capacity in cases of poor speech-recognition performance after implantation, indicate that the poor performance is likely caused by poor auditory processing. Thus, these tests could help to interpret the results post-operatively and individualize rehabilitation programs. In addition, measures of auditory lexical decision can help in understanding the role of bottom-up and top-down factors in speech-recognition outcome with CI use.

Using linguistic tests in clinical practice. When using a lexical-decision test in clinical practice or in research to measure lexical-access ability it is important to be aware of the factors that influence the test results. This holds especially for studies where participants of different centers from different countries are compared that are tested in their own language. The LDT tests utilized in the current studies used only highly frequent, thus known, words to test lexical-access ability, not the vocabulary size. Secondly, it is important to use well formed pseudowords to ensure that lexical access actually takes place. Our results also support the assumptions of the model that response times for words and pseudowords reflect different aspects of lexical decision and should therefore be analyzed separately.

Preferably a second test of lexical access is used to obtain a more accurate estimate of lexical access by looking at the functional overlap between tests. Although the simple word-naming test proved suitable for this purpose in Chapter 3, a more precise test would probably give more representative results for groups of native listeners, like those tested in Chapter 4.

Measuring speech recognition in clinical practice. The results of Chapters 3, 4 and 5 showed that linguistic skills and working-memory capacity are important factors in speech recognition of sentences in adverse listening conditions and that the DIN test is less influenced by linguistic and cognitive skills (Chapters 3 and 4) than the SIN test. This corresponds with the results of Heinrich, Henshaw, and Ferguson (2015), who found, in a group of 44 adults with mild hearing loss, that performance on the sentence-in-noise task was associated with cognition, while performance on the digits-in-noise task was not. The DIN test is more often used as screening instrument and recently suggested as a tool for diagnostics in home-measurements (Cullington and Aidi 2017; De Graaff et al. 2016). The results also confirm that the DIN test could be used to compare performance of a wide range of listeners with normal hearing, mild hearing losses, hearing-aid users, and CI users. Furthermore, for non-native listeners the DIN test gives a better estimate of auditory aspects of speech-in-noise recognition while the SIN test provides a better estimate of performance in the non-native language. The lower demand on linguistic and cognitive ability suggests that the DIN test is also a speech-in-noise test that could be more applicable to young children, providing an indication of mostly bottom-up processing of speech-in-noise. This information is useful for evaluation of CI- or hearing-aid fitting. On the other hand, for evaluation of daily life speech-recognition performance, where context and linguistic knowledge can be used, the DIN test might be less suitable. Heinrich et al. (2015), nevertheless, found that their digits-in-noise test correlated more with several questionnaires on aspects of auditory functioning than phoneme discrimination or their sentence-in-noise test.

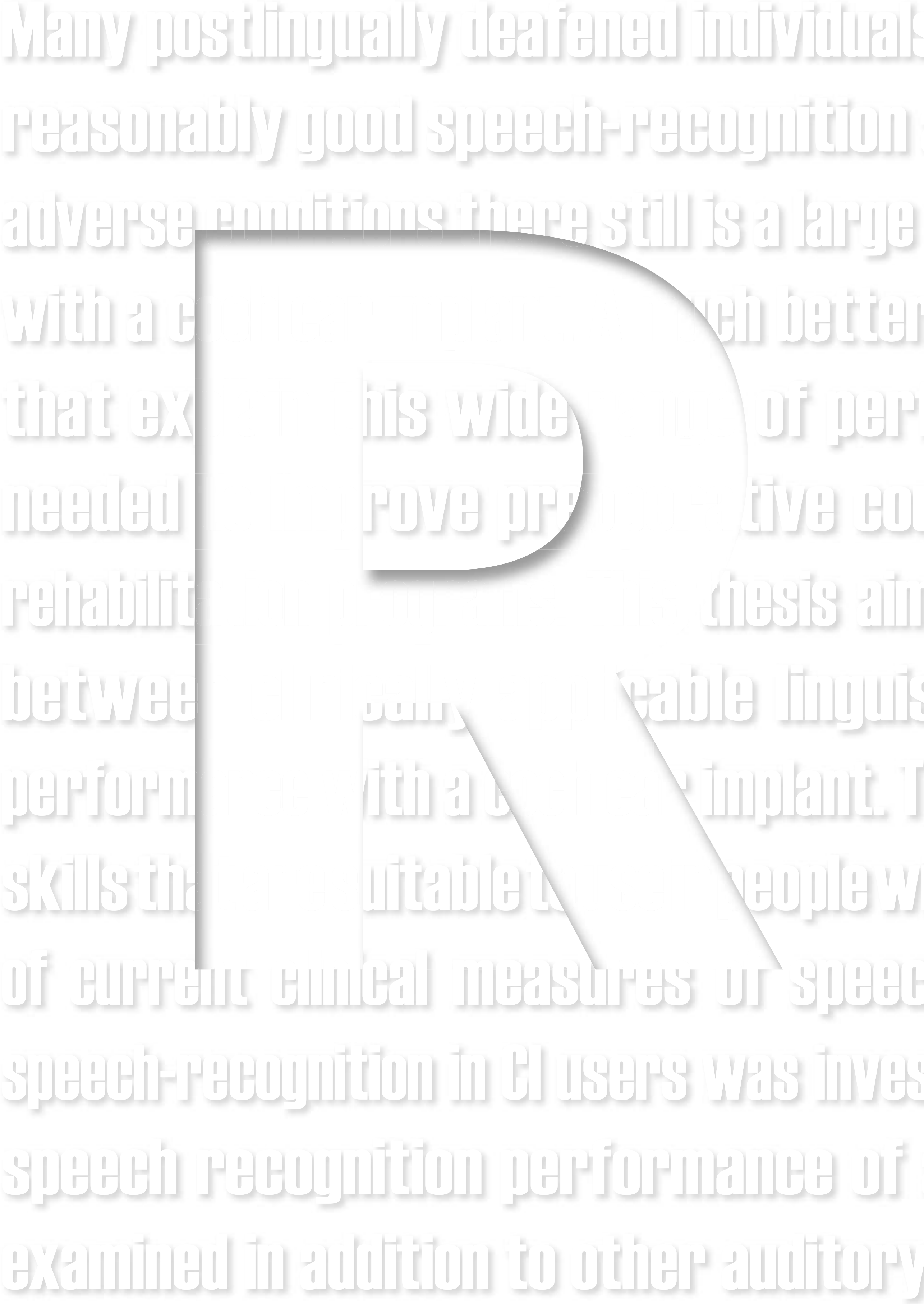
For evaluation of CI performance with current clinical measures the results suggest to start with phoneme recognition tests (CVC) in quiet. When CVC phoneme scores exceed about 40% correct phonemes, the DIN test is suitable to extend the test battery to avoid ceiling effects and to also test auditory aspects of speech-in-noise recognition. From DIN scores of around -2 dB SNR or more favorable, the SIN test seems a suitable addition and can be repeatedly and reliably measured. When SIN scores are nevertheless poor, linguistic or cognitive abilities should be evaluated to obtain a better understanding of the locus of the poor performance. Using keyword scoring in the SIN test can enable more hearing impaired listeners to achieve results of which more are in the reliable range (that is, SRTs of <15 dB SNR).

GENERAL CONCLUSIONS

The main purpose of cochlear implants for postlingually deafened adults is to restore communication abilities. Speech-recognition performance measures are, thus, very relevant to evaluate cochlear implant outcome. The results of this thesis demonstrate that measures of the recognition of monosyllables in quiet have their limits and that adaptive tests of speech recognition in noise can provide more information. The DIN test informs primarily about auditory capacity and is less influenced by cognitive and linguistic abilities. The SIN test is more suitable for evaluation of auditory functioning in everyday listening situations. For listeners that are not able to perform the standard SIN test, keyword scoring can be used as an alternative.

The visual measures of lexical-access ability were sensitive to differences in language proficiency. The auditory measures of lexical access showed that lexical-decision RT, accuracy, and an efficient balance between accuracy and RT were influenced by the quality of the auditory input signal in addition to linguistic abilities. Furthermore, response patterns were different for CI users in contrast to those of listeners with normal hearing listening to vocoded speech, indicating that they have adapted their processing strategy for listening to degraded input. CI users take more time to evaluate the uncertain input before they make a decision and seem to be used to guessing and select the most probable target word when they mishear parts of the stimulus.

Sentence-in-noise recognition of normal-hearing listeners with lower linguistic abilities, including non-natives listeners, was largely explained by lexical-access ability and to a lesser extent by vocabulary size. For CI users the influence of linguistic abilities seems more complex and interacts with the auditory capacity. For CI users with reliable SIN scores, working-memory capacity and to a lesser extent lexical-access ability predict sentence-in-noise performance. These measures, thus, may provide information for counselling pre-operatively and can help explain performance post-operatively. Performance of these patients may be resource-limited and rehabilitation should focus on strengthening the use of linguistic and cognitive resources. For CI users with very poor speech-in-noise recognition the quality of input seems to be limiting the effective use of good linguistic and cognitive skills. For these patients rehabilitation should focus on enhancing the quality of the bottom-up stimulus either by fitting or additional technology. aLDT measures can be helpful in identifying whether speech-recognition performance is data limited or resource limited and hence determining the rehabilitative approach.


Many postlingually deafened individuals
reasonably good speech-recognition
adverse
still is a large
with a c
ch better
that ex
his wide
of per
needed
rove
tive co
rehabilit
thesis aim
betwee
cable linguis
perform
ith a
implant. T
skills tha
uitable t
people w
of current clinical measures of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

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 and hearing-impaired adults', *International Journal of Audiology*, 47 Suppl 2: S53-71.

ANSI. 1997. "S3.5-1997, Methods for the calculation of the speech intelligibility index." In, 90-119. New York: American National Standards Institute.

Baayen, H., R. Piepenbrock, and H. Van Rijn. 1993. *The CELEX lexical data base CD-ROM* (University of Pennsylvania, Linguistic Data Consortium: Philadelphia, USA).

Bailey, I. L., and J. E. Lovie. 1980. 'The Design and Use of a New near-Vision Chart', *American Journal of Optometry and Physiological Optics*, 57: 378-87.

Balota, David A, and James I Chumbley. 1985. 'The locus of word-frequency effects in the pronunciation task: Lexical access and/or production?', *Journal of Memory and Language*, 24: 89-106.

Balota, David A., and James I. Chumbley. 1984. 'Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage', *Journal of Experimental Psychology: Human Perception and Performance*, 10: 340-57.

Baskent, D., J. Clarke, C. Pals, M. R. Benard, P. Bhargava, J. Saija, A. Sarampalis, A. Wagner, and E. Gaudrain. 2016. 'Cognitive Compensation of Speech Perception With Hearing Impairment, Cochlear Implants, and Aging: How and to What Degree Can It Be Achieved?', *Trends in Hearing*, 20: 1-16.

Baskent, D., C. L. Eiler, and B. Edwards. 2010. 'Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss', *Hearing Research*, 260: 54-62.

Benard, M. R., J. S. Mensink, and D. Baskent. 2014. 'Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities', *Journal of the Acoustical Society of America*, 135: El88-El94.

Benichov, J., L. C. Cox, P. A. Tun, and A. Wingfield. 2012. 'Word Recognition Within a Linguistic Context: Effects of Age, Hearing Acuity, Verbal Ability, and Cognitive Function', *Ear and Hearing*, 33: 250-56.

Besser, J., T. Koelewijn, A. A. Zekveld, S. E. Kramer, and J. M. Festen. 2013. 'How Linguistic Closure and Verbal Working Memory Relate to Speech Recognition in Noise-A Review', *Trends in Amplification*, 17: 75-93.

Besser, J., A. A. Zekveld, S. E. Kramer, J. Ronnberg, and J. M. Festen. 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.

Blamey, P., P. Arndt, F. Bergeron, G. Bredberg, J. Brimacombe, G. Facer, J. Larky, B. Lindstrom, J. Nedzelski, A. Peterson, D. Shipp, S. Staller, and L. Whitford. 1996. 'Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants', *Audiology and Neuro-Otology*, 1: 293-306.

Blamey, P., F. Artieres, D. Baskent, F. Bergeron, A. Beynon, E. Burke, N. Dillier, R. Dowell, B. Fraysse, S. Gallego, P. J. Govaerts, K. Green, A. M. Huber, A. Kleine-Punte, B. Maat, M. Marx, D. Mawman, I. Mosnier, A. F. O'Connor, S. O'Leary, A. Rousset, K. Schauwers, H. Skarzynski, P. H. Skarzynski, O. Sterkers, A. Terranti, E. Truy, P. Van de Heyning, F. Venail, C. Vincent, and D. S. Lazard. 2013. 'Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients', *Audiology and Neuro-Otology*, 18: 36-47.

Bosman, A.J., and G.F. Smoorenburg. 1995. 'Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment', *Audiology*, 34: 260-84.

Bradlow, Ann R., and Jennifer A. Alexander. 2007. 'Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners', *The Journal of the Acoustical Society of America*, 121: 2339-49.

Broersma, M. 2007. "Why the'president'does not excite the'press: The limits of spurious lexical activation in L2 listening." In *16th International Congress of Phonetics Sciences (ICPhS 2007)*, 1909-12. Pirrot.

Broersma, M., and A. Cutler. 2011. 'Competition dynamics of second-language listening', *Quarterly Journal of Experimental Psychology*, 64: 74-95.

Budenz, C. L., M. K. Cosetti, D. H. Coelho, B. Birenbaum, J. Babb, S. B. Waltzman, and P. C. Roehm. 2011. 'The effects of cochlear implantation on speech perception in older adults', *Journal of the American Geriatrics Society*, 59: 446-53.

Carreiras, M., M. Perea, and J. Grainger. 1997. 'Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons', *Journal of Experimental Psychology-Learning Memory and Cognition*, 23: 857-71.

Collison, E. A., B. Munson, and A. E. Carney. 2004. 'Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants', *Journal of Speech Language and Hearing Research*, 47: 496-508.

Cullington, H. E., and T. Aidi. 2017. 'Is the digit triplet test an effective and acceptable way to assess speech recognition in adults using cochlear implants in a home environment?', *Cochlear Implants International*, 18: 97-105.

Dahan, D., and M. K. Tanenhaus. 2004. 'Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints', *Journal of Experimental Psychology-Learning Memory and Cognition*, 30: 498-513.

De Graaff, F., E. Huysmans, O. U. Qazi, F. J. Vanpoucke, P. Merkus, S. T. Goverts, and C. Smits. 2016. 'The Development of Remote Speech Recognition Tests for Adult Cochlear Implant Users: The Effect of Presentation Mode of the Noise and a Reliable Method to Deliver Sound in Home Environments', *Audiology and Neuro-Otology*, 21 Suppl 1: 48-54.

De Groot, A. M. B., S. Borgwaldt, M. Bos, and E. Van Den Eijnden. 2002. 'Lexical Decision and Word Naming in Bilinguals: Language Effects and Task Effects', *Journal of Memory and Language*, 47: 91-124.

De Groot, Annette MB. 2011. *Language and cognition in bilinguals and multilinguals: An introduction* (Psychology Press).

de Vet, H. C., C. B. Terwee, D. L. Knol, and L. M. Bouter. 2006. 'When to use agreement versus reliability measures', *Journal of Clinical Epidemiology*, 59: 1033-9.

Devocht, E. M. J., A. M. L. Janssen, J. Chalupper, R. J. Stokroos, and E. L. J. George. 2016. 'Monaural Beamforming in Bimodal Cochlear Implant Users: Effect of (A)symmetric Directivity and Noise Type', *Plos One*, 11.

Farris-Trimble, A., B. McMurray, N. Cigrand, and J. B. Tomblin. 2014. 'The process of spoken word recognition in the face of signal degradation', *Journal of Experimental Psychology: Human Perception and Performance*, 40: 308-27.

Feeney, MP, RL Folmer, J Vachhani, GP McMillan, CS Watson, and GR Kidd. 2013. "Test performance of a computer version of the digits-in-noise test for adult hearing screening." In *Poster Presented at the 36th Annual Midwinter Meeting of the Association for Research in Otolaryngology*.

Festen, J. M., and R. Plomp. 1990. 'Effects of Fluctuating Noise and Interfering Speech on the Speech-Reception Threshold for Impaired and Normal Hearing', *Journal of the Acoustical Society of America*, 88: 1725-36.

Finley, C. C., T. A. Holden, L. K. Holden, B. R. Whiting, R. A. Chole, G. J. Neely, T. E. Hullar, and M. W. Skinner. 2008. 'Role of Electrode Placement as a Contributor to Variability in Cochlear Implant Outcomes', *Otology & Neurotology*, 29: 920-28.

Friesen, Lendra M., Robert V. Shannon, Deniz Baskent, and Xiaosong Wang. 2001. 'Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants', *The Journal of the Acoustical Society of America*, 110: 1150-63.

Fullgrabe, C., and S. Rosen. 2016. 'On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds', *Frontiers in Psychology*, 7: 1268.

Goldinger, Stephen D. 1996. 'Auditory Lexical Decision', *Language and Cognitive Processes*, 11: 559-68.

Goverts, S.T., E. Huysmans, S.E. Kramer, A.M.B. De Groot, and T. Houtgast. 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: 1702-08.

Goy, H., M. Pelletier, M. Coletta, and M. K. Pichora-Fuller. 2013. 'The effects of semantic context and the type and amount of acoustic distortion on lexical decision by younger and older adults', *Journal of Speech Language and Hearing Research*, 56: 1715-32.

Grainger, J., and A. M. Jacobs. 1996. 'Orthographic processing in visual word recognition: a multiple read-out model', *Psychological Review*, 103: 518-65.

Grainger, J., and J. Segui. 1990. 'Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies', *Attention, Perception, & Psychophysics*, 47: 191-98.

Groot, A.M.B.de, S. Borgwaldt, M. Bos, and E. Eijnden van den. 2002. 'Lexical Decision and Word Naming in Bilinguals: Language Effects and Task Effects', *Journal of Memory and Language,* 47: 91-124.

Grosjean, F. 1980. 'Spoken Word Recognition Processes and the Gating Paradigm', *Perception & Psychophysics*, 28: 267-83.

Hagerman, B. 1982. 'Sentences for Testing Speech-Intelligibility in Noise', *Scandinavian Audiology*, 11: 79-87.

Hallgren, M., B. Larsby, B. Lyxell, and S. Arlinger. 2001. 'Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired persons', *Journal of the American Academy of Audiology*, 12: 357-70.

Haumann, S., V. Hohmann, M. Meis, T. Herzke, T. Lenarz, and A. Buchner. 2012. 'Indication criteria for cochlear implants and hearing aids: impact of audiological and non-audiological findings', *Audiology Research*, 2: e12.

Heinrich, A., H. Henshaw, and M. A. Ferguson. 2015. 'The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests', *Frontiers in Psychology*, 6: 782.

Heydebrand, G., S. Hale, L. Potts, B. Gotter, and M. Skinner. 2007. 'Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation', *Audiology Neuro-otology*, 12: 254-64.

Holden, L.K., C.C. Finley, J.B. Firszt, T.A. Holden, C. Brenner, L.G. Potts, B.D. Gotter, S.S. Vanderhoof, K. Mispagel, G. Heydebrand, and M.W. Skinner. 2013. 'Factors affecting open-set word recognition in adults with cochlear implants', *Ear and Hearing*, 34: 342-60.

Horn, J. L., and R. B. Cattell. 1967. 'Age Differences in Fluid and Crystallized Intelligence', *Acta Psychologica*, 26: 107-&.

Houtgast, T., and J. M. Festen. 2008. 'On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise', *International Journal of Audiology*, 47: 287-95.

Huysmans, E., J. de Jong, J.H. van Lanschot-Wery, J.M. Festen, and S.T. Goverts. 2014. 'Long-term effects of congenital hearing impairment on language performance in adults', *Lingua*, 139: 102-21.

Jansen, S., H. Luts, K. C. Wagener, B. Kollmeier, M. Del Rio, R. Dauman, C. James, B. Fraysse, E. Vormes, B. Frachet, J. Wouters, and A. van Wieringen. 2012. 'Comparison of three types of French speech-in-noise tests: a multi-center study', *International Journal of Audiology*, 51: 164-73.

Kaandorp, M. W., A. M. B. De Groot, J. M. Festen, C. Smits, and S. T. Goverts. 2016. 'The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise', *International Journal of Audiology*, 55: 157-67.

Kaandorp, M. W., C. Smits, P. Merkus, J. M. Festen, and S. T. Goverts. 2017. 'Lexical-Access Ability and Cognitive Predictors of Speech Recognition in Noise in Adult Cochlear Implant Users', *Trends in Hearing*, 21: 1-15.

Kaandorp, M. W., C. Smits, P. Merkus, S. T. Goverts, and J. M. Festen. 2015. 'Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users', *International Journal of Audiology*, 54: 48-57.

Kollmeier, B., and M. Wesselkamp. 1997. 'Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment', *Journal of the Acoustical Society of America*, 102: 2412-21.

Larsby, Birgitta, Mathias Hällgren, and Björn Lyxell. 2012. "Working memory capacity and lexical access in speech recognition in noise." In *3rd International Symposium on Auditory and Audiological Research (ISAAR 2011), 24-26 August 2011, Nyborg, Denmark*, 95-102. The Danavox Jubilee Foundation.

Lazard, D. S., C. Vincent, F. Venail, P. Van de Heyning, E. Truy, O. Sterkers, P. H. Skarzynski, H. Skarzynski, K. Schauwers, S. O'Leary, D. Mawman, B. Maat, A. Kleine-Punte, A. M. Huber, K. Green, P. J. Govaerts, B. Fraysse, R. Dowell, N. Dillier, E. Burke, A. Beynon, F. Bergeron, D. Baskent, F. Artieres, and P. J. Blamey. 2012. 'Pre-, Per- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time', *Plos One*, 7: 1-11.

Leung, J., N. Y. Wang, J. D. Yeagle, J. Chinnici, S. Bowditch, H. W. Francis, and J. K. Niparko. 2005. 'Predictive models for cochlear implantation in elderly candidates', *Archives of Otolaryngology-Head & Neck Surgery*, 131: 1049-54.

Luce, P. A., and D. B. Pisoni. 1998. 'Recognizing spoken words: the neighborhood activation model', *Ear and Hearing*, 19: 1-36.

Luteijn, F, and DPH Barelds. 2004. "Handleiding van de Groninger Intelligentie Test 2 (GIT-2)." In.: Lisse: Harcourt Test Publishers.

Lyxell, B., J. Andersson, U. Andersson, S. Arlinger, G. Bredberg, and H. Harder. 1998. 'Phonological representation and speech understanding with cochlear implants in deafened adults', *Scandinavian Journal of Psychology*, 39: 175-9.

Lyxell, B., J. Andersson, S. Arlinger, G. Bredberg, H. Harder, and J. Ronnberg. 1996. 'Verbal information-processing capabilities and cochlear implants: implications for preoperative predictors of speech understanding', *Journal of Deaf Studies and Deaf Education*, 1: 190-201.

Lyxell, B., U. Andersson, E. Borg, and I.S. Ohlsson. 2003. 'Working-memory capacity and phonological processing in deafened adults and individuals with a severe hearing impairment', *International Journal of Audiology*, 42 Suppl 1: S86-S89.

Marslen-Wilson, W. D. 1987. 'Functional parallelism in spoken word-recognition', *Cognition*, 25: 71-102.

Marslen-Wilson, W. D. . 2013. "Issues of Process and Representation in Lexical Access." In *Cognitive models of speech processing: The second Sperlonga meeting*, 187. Psychology Press.

Marslen-Wilson, W. D., and A. Welsh. 1978. 'Processing interactions and lexical access during word recognition in continuous speech', *Cognitive Psychology*, 10: 29-63.

Mattys, S.L., M.H. Davis, A.R. Bradlow, and S.K. Scott. 2012. 'Speech recognition in adverse conditions: A review', *Language and Cognitive Processes*, 27: 953-78.

McClelland, J.L., and J.L. Elman. 1986. 'The TRACE Model of Speech Perception', *Cognitive Psychology*, 18: 1-86.

McGraw, Kenneth O, and Seok P Wong. 1996. 'Forming inferences about some intraclass correlation coefficients', *Psychological methods*, 1: 30.

McMurray, B., A. Farris-Trimble, M. Seedorff, and H. Rigler. 2016. 'The Effect of Residual Acoustic Hearing and Adaptation to Uncertainty on Speech Perception in Cochlear Implant Users: Evidence From Eye-Tracking', *Ear and Hearing*, 37: e37-51.

McQueen, J. M., and F. Huettig. 2012. 'Changing only the probability that spoken words will be distorted changes how they are recognized', *The Journal of the Acoustical Society of America*, 131: 509-17.

McRae, K., D. Jared, and M. S. Seidenberg. 1990. 'On the Roles of Frequency and Lexical Access in Word Naming', *Journal of Memory and Language*, 29: 43-65.

Meyer, T. A., S. A. Frisch, D. B. Pisoni, R. T. Miyamoto, and M. A. Svirsky. 2003. 'Modeling open-set spoken word recognition in postlingually deafened adults after cochlear implantation: some preliminary results with the neighborhood activation model', *Otology and Neurotology*, 24: 612-20.

Moore, D. R., M. Edmondson-Jones, P. Dawes, H. Fortnum, A. McCormack, R. H. Pierzycki, and K. J. Munro. 2014. 'Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age', *Plos One*, 9: e107720.

Mosnier, I., J. P. Bebear, M. Marx, B. Fraysse, E. Truy, G. Lina-Granade, M. Mondain, F. Sterkers-Artieres, P. Bordure, A. Robier, B. Godey, B. Meyer, B. Frachet, C. Poncet, D. Bouccara, and O. Sterkers. 2014. 'Predictive factors of cochlear implant outcomes in the elderly', *Audiology and Neuro-Otology*, 19 Suppl 1: 15-20.

Munson, Benjamin. 2001. 'Relationships between vocabulary size and spoken word recognition in children aged 3 to 7', *Contemporary Issues in Communication Science and Disorders*, 28: 20-29.

Nilsson, M., S.D. Soli, and J.A. Sullivan. 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: 1085-99.

Norman, D. A., and D. G. Bobrow. 1975. 'Data-Limited and Resource-Limited Processes', *Cognitive Psychology*, 7: 44-64.

Norris, D., and J. M. McQueen. 2008. 'Shortlist B: a Bayesian model of continuous speech recognition', *Psychological Review*, 115: 357-95.

Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric theory* (McGraw-Hill Inc.: New York).

Pals, C., A. Sarampalis, and D. Baskent. 2013. 'Listening Effort With Cochlear Implant Simulations', *Journal of Speech Language and Hearing Research*, 56: 1075-84.

Pichora-Fuller, M. K. 2008. 'Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing', *International Journal of Audiology*, 47: S72-S82.

Pichora-Fuller, M. K., and H. Levitt. 2012. 'Speech Comprehension Training and Auditory and Cognitive Processing in Older Adults', *American Journal of Audiology*, 21: 351-57.

Pichora-Fuller, M.K. 2003. 'Cognitive aging and auditory information processing', *International Journal of Audiology*, 42 Suppl 2: 2S26-2S32.

Pisoni, D. B. 2000. 'Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception', *Ear and Hearing*, 21: 70-78.

Pittman, A. L., and M. A. Rash. 2016. 'Auditory Lexical Decision and Repetition in Children: Effects of Acoustic and Lexical Constraints', *Ear and Hearing*, 37: e119-28.

Pittman, A. L., and B. C. Schuett. 2013. 'Effects of Semantic and Acoustic Context on Nonword Detection in Children With Hearing Loss', *Ear and Hearing*, 34: 213-20.

Plant, K., R. van Hoesel, H. McDermott, P. Dawson, and R. Cowan. 2016. 'Influence of contralateral acoustic hearing on adult bimodal outcomes after cochlear implantation', *International Journal of Audiology*, 55: 472-82.

Plomp, R., and A.M. Mimpen. 1979. 'Improving the reliability of testing the speech reception threshold for sentences', *Audiology*, 18: 43-52.

Ramscar, Michael, Peter Hendrix, Bradley Love, and R Harald Baayen. 2013. 'Learning is not decline: The mental lexicon as a window into cognition across the lifespan', *The Mental Lexicon*, 8: 450-81.

Ratcliff, R., P. Gomez, and G. McKoon. 2004. 'A diffusion model account of the lexical decision task', *Psychological Review*, 111: 159-82.

Roditi, R. E., S. F. Poissant, E. M. Bero, and D. J. Lee. 2009. 'A Predictive Model of Cochlear Implant Performance in Postlingually Deafened Adults', *Otology & Neurotology*, 30: 449-54.

Ronnberg, J., M. Rudner, C. Foo, and T. Lunner. 2008. 'Cognition counts: a working memory system for ease of language understanding (ELU)', *International Journal of Audiology*, 47 Suppl 2: S99-105.

Ronnberg, J., E. Samuelsson, and E. Borg. 2000. 'Visual cognitive tests, central auditory function and auditory communication', *Scandinavian Audiology*, 29: 196-206.

Rosemann, S., C. Giessing, J. Ozyurt, R. Carroll, S. Puschmann, and C. M. Thiel. 2017. 'The Contribution of Cognitive Factors to Individual Differences in Understanding Noise-Vocoded Speech in Young and Older Adults', *Frontiers in Human Neuroscience*, 11.

Rosen, S., A. Faulkner, and L. Wilkinson. 1999. 'Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants', *Journal of the Acoustical Society of America*, 106: 3629-36.

Rubenstein, H., S. S. Lewis, and M. A. Rubenstein. 1971. 'Evidence for Phonemic Recording in Visual Word Recognition', *Journal of Verbal Learning and Verbal Behavior*, 10: 645-57.

Schlichting, Liesbeth. 2005. *Peabody picture vocabulary test-III-NL* (Hartcourt Assessment BV: Amsterdam, the Netherlands).

Shannon, R. V., Q. J. Fu, and J. Galvin, 3rd. 2004. 'The number of spectral channels required for speech recognition depends on the difficulty of the listening situation', *Acta Otolaryngology Suppl*: 50-4.

Smits, C., and J. M. Festen. 2011. 'The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: steady-state noise', *The Journal of the Acoustical Society of America*, 130: 2987-98.

Smits, C., S. T. Goverts, and J. M. Festen. 2013. 'The digits-in-noise test: assessing auditory speech recognition abilities in noise', *Journal of the Acoustical Society of America*, 133: 1693-706.

Smits, C., T.S. Kapteyn, and T. Houtgast. 2004. 'Development and validation of an automatic speech-in-noise screening test by telephone', *International Journal of Audiology*, 43: 15-28.

Studebaker, G. A. 1985. 'A Rationalized Arcsine Transform', *Journal of Speech and Hearing Research*, 28: 455-62.

Tabossi, P., and L. Laghi. 1992. 'Semantic Priming in the Pronunciation of Words in 2 Writing Systems - Italian and English', *Memory & Cognition*, 20: 303-13.

Theunissen, M., De W. Swanepoel, and J. Hanekom. 2009. 'Sentence recognition in noise: Variables in compilation and interpretation of tests', *International Journal of Audiology*, 48: 743-57.

Turner, J. E., T. Valentine, and A. W. Ellis. 1998. 'Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision', *Memory & Cognition*, 26: 1282-91.

Tyler, L.K., R. Waksler, and W. D. Marslen-Wilson. 2013. "Representation and access of derived words in English." In *Cognitive models of speech processing: The second Sperlonga meeting*, 126. Psychology Press.

Vaillancourt, H., C. Laroche, C. Mayer, C. Basque, M. Nali, A. Eriks-Brophy, S. D. Soli, and C. Giguere. 2005. 'Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations', *International Journal of Audiology*, 44: 358-69.

Van der Maas, H. L., C. V. Dolan, R. P. Grasman, J. M. Wicherts, H. M. Huizenga, and M. E. Raijmakers. 2006. 'A dynamical model of general intelligence: the positive manifold of intelligence by mutualism', *Psychological Review*, 113: 842-61.

Van Rooij, J. C., and R. Plomp. 1990. 'Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses', *The Journal of the Acoustical Society of America*, 88: 2611-24.

Van Wieringen, A., and J. Wouters. 2008. 'LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands', *International Journal of Audiology*, 47: 348-55.

Van Wijngaarden, S.J., H.J.M. Steeneken, and T. Houtgast. 2002. 'Quantifying the intelligibility of speech in noise for non-native listeners', *The Journal of the Acoustical Society of America*, 111: 1906-16.

Versfeld, N.J., L. Daalder, J.M. Festen, and T. Houtgast. 2000. 'Method for the selection of sentence materials for efficient measurement of the speech reception threshold', *The Journal of the Acoustical Society of America*, 107: 1671-84.

Wagenaar, Willem A. 1969. 'Note on the construction of digram-balanced Latin squares', *Psychological Bulletin*, 72: 384.

Wagener, K, T Brand, and B Kollmeier. 1999. 'Entwicklung und Evaluation eines Satztests fr die deutsche Sprache III: Evaluation des Oldenburger Satztests', *Zeitschrift fr Audiologie/Audiological Acoustics*, 38: 8695.

Wagner, A. E., P. Toffanin, and D. Baskent. 2016. 'The Timing and Effort of Lexical Access in Natural and Degraded Speech', *Frontiers in Psychology*, 7: 398.

Wang, X., and L. E. Humes. 2010. 'Factors influencing recognition of interrupted speech', *The Journal of the Acoustical Society of America*, 128: 2100-11.

Weiss, Deborah, and James J. Dempsey. 2008. 'Performance of Bilingual Speakers on the English and Spanish Versions of the Hearing in Noise Test (HINT)', *Journal of the American Academy of Audiology*, 19: 5-17.

Wong, L. L., and S. K. Keung. 2013. 'Adaptation of scoring methods for testing cochlear implant users using the Cantonese Hearing In Noise Test (CHINT)', *Ear and Hearing*, 34: 630-6.

Zekveld, A.A., E.L.J. George, S.E. Kramer, S.T. Goverts, and T. Houtgast. 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: 576-84.

Zokoll, M. A., S. Hochmuth, A. Warzybok, K. C. Wagener, M. Buschermohle, and B. Kollmeier. 2013. 'Speech-in-Noise Tests for Multilingual Hearing Screening and Diagnostics', *American Journal of Audiology*, 22: 175-78.

S

Many postlingually deafened individuals

reasonably good speech-recognition

adverse conditions there still is a large

with a coc ... uch better

that exp ... ns wide range of perf

needed ... ove preoperative co

rehabilitatio ... This thesis aim

between clinically ... le linguis

performance with a coc ... nplant. T

skills th ... people w

of current cli ... asures of speec

speech-recognition in CI users was inves

speech recognition performance of

examined in addition to other auditory

Summary

Samenvatting

Many postlingually deafened persons who receive a cochlear implant (CI) achieve reasonably good speech-recognition abilities in quiet. However, there still is a large range of speech-recognition performance in quiet and even more in noise. A better understanding of the underlying factors that explain this wide range of performance is needed to improve pre-operative CI counselling and to optimize user-specific rehabilitation programs. In this thesis auditory, linguistic/cognitive, and biomedical factors that potentially affect speech recognition were distinguished. An increasing interest has emerged in non-auditory factors, like working-memory capacity and linguistic skills, which influence speech recognition in challenging conditions. However, the number of studies on the effect of these factors on CI performance, although increasing, is still limited. This thesis focused on the relations between clinically applicable linguistic measures and speech-recognition performance with CI. Because words are the primary meaning-bearing elements of language, lexical-access ability and vocabulary size were used as measures of linguistic abilities. As these studies are focused on people with limited auditory abilities, there is a need for a speech-recognition test which is applicable over a wide range of auditory capabilities. Additionally, linguistic abilities need to be tested such that the results are not limited by poor auditory performance. The first chapters of this thesis are dedicated to the selection of tests that are applicable to CI users.

The goal of the study described in Chapter 2 was to investigate the feasibility, reliability, and validity of the recently developed Digits in Noise test (DIN) for measuring speech-recognition performance in CI users and hearing-aid users. The DIN test was designed to measure speech-in-noise recognition ability comparable to the Sentences in Noise test (SIN), but with less demand on linguistic and cognitive skills. The results showed that the DIN test was feasible for more participants than the SIN test. The DIN test also showed high reliability and a small measurement error as compared to the SIN test. Overall the DIN and SIN tests were highly correlated. The results of this study informed that current clinical measures are not always suitable for measuring CI outcome. In all following studies the DIN test was added to the standard test battery with phoneme recognition in monosyllable word lists and sentence-in-noise recognition by means of the SIN test.

The study described in Chapter 3 was conducted to investigate the effect of linguistic abilities on different measures of speech-in-noise recognition. To measure linguistic abilities independent of auditory performance, as would be needed in pre-operative testing of CI-candidates, it was decided to start with using visual tests of lexical access and vocabulary size. Young adult listeners with normal hearing were tested to avoid interaction

with auditory capacity or biomedical factors. Additionally, the effect of keyword scoring compared to whole-sentence scoring on SIN thresholds was examined. Variation was introduced in linguistic abilities by including three groups of 24 young normal-hearing listeners: high-educated native Dutch-, lower-educated native Dutch-, and high-educated non-native listeners. Lexical-access was measured with well-known tests that were built specifically for this study: a lexical-decision test (in which listeners have to discriminate quickly between words and pseudowords) and a wordnaming test (in which listeners have to read a list of words as quickly as possible). Evaluation of the lexical-access variables showed that lexical-access ability was best measured with the combined results of both tests. Non-native listeners performed much poorer than native listeners on the SIN test in stationary noise and even more so in fluctuating noise. Lexical-access ability explained a large part of the variance in SIN thresholds in the whole study group and a smaller part in the two groups of native listeners. The results described in this chapter showed that lexical decision and word naming are suitable measures of lexical-access ability and they confirmed that the DIN test is hardly influenced by cognitive or linguistic skills. In Chapters 4 and 5 the DIN test was, therefore, used as a measure of speech recognition in noise that represents mainly auditory factors. It was also shown that keyword scoring resulted in a 1.0 dB more favorable speech reception threshold (SRT) than sentence scoring, making the test applicable to a wider range of listeners.

In the exploratory study described in Chapter 4 the effects of linguistic and cognitive measures on different measures of speech recognition were examined in 24 postlingually deafened CI users. In addition to visual measures of lexical-access ability and vocabulary size, two other cognitive factors, namely working-memory capacity (the reading-span test) and a test of the ability to recognize visually masked sentences (with a visual analogue of the sentence-in-noise test: the Text Reception Threshold test, TRT), were included as predictor variables. Furthermore, biomedical and audiological factors were examined. No significant relations were found between the predictor variables and phoneme recognition with monosyllables in quiet or DIN thresholds. In the correlation analyses a derived variable was included that reflects the cognitive and linguistic factors in SIN thresholds by taking the difference between SIN and DIN: SRT_{diff}. Correlation analyses revealed that better SIN scores were related to better working-memory capacity and a shorter duration of hearing loss. Better SRT_{diff} scores were related to better working-memory capacity and shorter lexical-decision RTs. The results of the regression analyses with pre-operative and post-operative factors suggest that poor verbal working-memory capacity and, to a lesser extent, poor lexical-access ability limit speech-recognition ability in listeners with a CI.

The final study described in Chapter 5 aimed to gain more insight in the interaction between measures of linguistic skills in the auditory domain and speech understanding as it occurs in daily life. The studies in Chapters 3 and 4 showed that the relation between visual lexical-access ability and speech recognition in noise was weaker in CI users than in normal hearing listeners. It was found that auditory factors were the main predictor of speech-recognition outcome in CI users and that there might be an interaction between auditory acuity and linguistic skills. In the auditory modality the input signal is degraded for CI users, which affects their performance and might affect lexical access. The objectives of the study described in this chapter were to examine the effect of degraded input, by either CI-simulated stimuli (with a noise-band vocoder) or by CI use, on auditory lexical-decision reaction times (RT) and accuracy. The results showed lower accuracy and longer RTs for degraded input by the vocoder manipulations and CI use, but the response patterns were different between groups. CI users were relatively slow in pseudoword responses and had a large bias toward real-word responses, which was not present in the normal-hearing vocoder group. CI users, thus, seem to have adopted a different lexical-access strategy than normal hearing listeners listening to vocoded speech, possibly because CI users are used to listening to degraded stimuli. Correlation analyses with measures of speech recognition in CI users revealed relations between monosyllable phoneme scores in quiet and real-word response times in the auditory lexical-decision test. The sentence recognition in noise performance was more related to an efficient balance between response accuracy and response time, that is, response efficiency both in pseudoword responses as in overall responses. These results show that measures of auditory lexical decision together with visual lexical decision might help in understanding speech-recognition outcome with CI use.

The final chapter presents a general discussion of the findings and clinical implications. The studies inform on how speech recognition in CI users can be measured with current clinical tests. The results demonstrate that measures of recognition of monosyllables in quiet have their limits and that adaptive tests of speech recognition in noise can provide more and relevant information. The DIN test informs primarily about auditory aspects of speech-in-noise recognition and is less influenced by cognitive and linguistic abilities. This test may also be used for follow-up measurements to monitor progress in speech-recognition performance after cochlear implantation. The SIN test is more suitable for evaluation of auditory functioning in everyday listening situations, because it involves more linguistic and cognitive processing. For listeners that are not able to perform the standard sentence test, keyword scoring can be used as an alternative. The relation between linguistic and cognitive skills and speech-recognition performance with CI seems complex and interacts with auditory capacity. For CI users with very poor speech-in-noise recognition, the quality of input seems to be limiting the effective use of good linguistic and cognitive skills. Overall, visual measures of linguistic abilities and working memory should be used to improve pre-operative CI counselling and can be used, next to auditory measures of lexical access, in the evaluation of speech-recognition outcome with CI.

SAMENVATTING

Als mensen ernstig slechthorend worden na het ontwikkelen van hun taal (postlinguaal) en een cochleair implantaat (CI) ontvangen, kunnen ze spraak vaak vrij goed verstaan in een rustige omgeving. De onderlinge verschillen in de vaardigheid om spraak te verstaan in stilte, en nog meer in lawaai zijn echter groot. Het is daarom belangrijk om de onderliggende factoren, die deze grote variatie in het verstaan van spraak verklaren, beter te begrijpen. Door meer inzicht in deze factoren kan de voorlichting van CI-kandidaten voorafgaand aan de operatie worden verbeterd en kan het revalidatietraject met het CI nog beter worden afgestemd op de individuele CI-gebruiker.

In dit proefschrift worden auditieve, talige/cognitieve en biomedische factoren onderscheiden die mogelijk van invloed zijn op het verstaan van spraak. In toenemende mate is er aandacht voor de rol van niet-auditieve factoren, zoals werkgeheugencapaciteit en talige vaardigheden, bij het verstaan van spraak onder uitdagende omstandigheden. Het aantal onderzoeken naar het effect van deze factoren op het verstaan van spraak met een CI neemt weliswaar toe, maar is nog steeds beperkt. Dit proefschrift richt zich op de relaties tussen klinisch toepasbare talige maten en de vaardigheden in het verstaan van spraak met een CI. Omdat woorden de primaire dragers van betekenis zijn in een taal, werden woordherkenning (i.e., lexicale toegang, de toegang tot woorden in het mentale lexicon) en woordenschatgrootte gebruikt als maten voor talige vaardigheden.

De onderzoeken in dit proefschrift richtten zich op het spraakverstaan bij volwassenen met beperkte auditieve vermogens. Om die reden was er behoefte aan een test die het verstaan van spraak betrouwbaar kan meten bij mensen met aanzienlijke verschillen in auditieve vermogens. Bovendien moesten talige vaardigheden zodanig worden getest dat de uitkomsten niet werden beïnvloed door de auditieve beperkingen. De eerste hoofdstukken van dit proefschrift zijn daarom gewijd aan de selectie van tests die gebruikt kunnen worden bij volwassenen met een CI.

Het doel van de studie in hoofdstuk 2 was om de haalbaarheid, de betrouwbaarheid en de validiteit te onderzoeken van de onlangs ontwikkelde Digits in Noise test (DIN test, i.e. cijfers-in-ruis test) voor het meten van spraakverstaanvaardigheden bij CI-gebruikers en hoortoestelgebruikers. De DIN-test werd ontwikkeld om de vaardigheden in het verstaan van spraak in ruis te meten op een manier die vergelijkbaar is met de Sentences in Noise test (SIN test, i.e. zinnen in ruis test), maar minder beroep doet op de talige en cognitieve vaardigheden. De resultaten toonden aan dat de DIN-test voor meer deelnemers afneembaar was dan de SIN-test. De DIN-test had een hoge betrouwbaarheid en een kleine meetfout in vergelijking met de SIN-test. De uitkomsten op de DIN- en SIN-test waren sterk gecorreleerd. Uit deze onderzoeksresultaten bleek dat de tests die momenteel klinisch worden gebruikt voor het meten van spraakverstaan niet voor alle CI-gebruikers goed bruikbaar zijn. De DIN-test werd daarom in alle overige studies in dit proefschrift toegevoegd aan de standaard testbatterij, die tests bevat voor het verstaan van fonemen in eenlettergrepige woorden en voor het verstaan van zinnen in ruis (SIN-test).

In de studie beschreven in Hoofdstuk 3 werd het effect onderzocht van talige vaardigheden op verschillende maten voor het verstaan van spraak in ruis. De deelnemers aan dit onderzoek waren jongvolwassenen met een normaal gehoor, waardoor interactie met auditieve vermogens of biomedische factoren werd beperkt. Talige vaardigheden werden in kaart gebracht met visuele tests voor het meten van lexicale toegang en woordenschatgrootte. Deze tests werden gekozen omdat ze talige vaardigheden meten op een manier die niet beïnvloed wordt door auditieve vermogens, wat noodzakelijk was voor het beoogde gebruik van deze tests bij preoperatief onderzoek bij CI-kandidaten. In deze studie werd ook onderzocht wat het effect was op de uitkomsten van de SIN-test als de responsen van deelnemers werden gescoord op basis van correct herhaalde sleutelwoorden in plaats van het correct herhalen van de volledige zin.

Het onderzoek werd uitgevoerd bij drie groepen van 24 jongvolwassenen met een normaal gehoor die van elkaar verschilden in talige vaardigheden: hoogopgeleid met Nederlands als moedertaal (native), lager opgeleid met Nederlands als moedertaal (native) en hoogopgeleid met een andere moedertaal en Nederlands lerend (non-native). Voor het meten van de vaardigheid in lexicale toegang werden twee bekende testmethodes gebruikt, gemeten met tests die specifiek voor dit onderzoek waren gemaakt. Namelijk een test waarbij deelnemers dienden aan te geven of een visueel aangeboden woord wel of geen bestaand woord is in het Nederlands (zgn. lexicale decisietest) en een test waarbij woorden zo snel mogelijk dienden te worden voorgelezen (zgn. woordleestaak). Uit een evaluatie van de uitkomstmaten bleek dat de vaardigheid van lexicale toegang het best kon worden gemeten met de gecombineerde resultaten van de twee tests. Non-native luisteraars presteerden beduidend minder goed dan native luisteraars op de SIN-test in stationaire ruis en nog minder goed in fluctuerende ruis. De vaardigheid in lexicale toegang verklaarde een groot deel van de variantie in SIN-drempels in de hele studiegroep en een kleiner deel in de twee groepen met native luisteraars. De resultaten die in dit hoofdstuk worden beschreven, toonden aan dat de tests voor lexicale decisie en de woordleestaak geschikt zijn voor het meten van de vaardigheid in lexicale toegang. Daarnaast bevestigden de resultaten van deze studie dat de uitkomsten van de DIN-test nauwelijks worden beïnvloed door cognitieve of talige vaardigheden. In hoofdstuk 4 en 5 werd de DIN-test daarom gebruikt als een maat voor het verstaan van spraak in ruis waarmee voornamelijk de invloed van auditieve factoren in kaart wordt gebracht. In de studie werd ook aangetoond dat het scoren van sleutelwoorden bij de SIN-test resulteerde in een 1.0 dB gunstigere spraak in ruis drempel (SRT) dan het scoren van de gehele zin. De SIN-test wordt met deze scoringswijze voor een groter aantal luisteraars bruikbaar.

Hoofdstuk 4 beschrijft de resultaten van een verkennende studie waarin de relatie tussen talige en cognitieve maten en verschillende maten voor spraakverstaan werd onderzocht bij 24 CI-gebruikers die postlinguaal doof werden. Talige vaardigheden werden gemeten met visuele tests voor lexicale toegang en woordenschatgrootte. Daarnaast werden de uitkomsten van een test voor werkgeheugencapaciteit (zgn. reading-span test) en een test voor het herkennen van visueel aangeboden gemaskeerde tekst (de Text Reception Threshold-test, TRT, analoog aan de SIN test) opgenomen als voorspellende variabelen. Verder werden biomedische en audiologische factoren onderzocht. Er werden geen significante relaties gevonden tussen de voorspellende variabelen en het verstaan van fonemen in eenlettergrepige woorden in stilte of het verstaan van cijfers in ruis (DIN-test). Voor verdere correlatieanalyse werd een aanvullende maat berekend die de invloed van cognitieve en talige vaardigheden op het verstaan reflecteert, nl. het verschil tussen de SIN- en DIN-score (SRT_{diff}). Uit correlatieanalyses bleek dat betere SIN-scores verband hielden met een beter werkgeheugen en een kortere duur van gehoorverlies. Betere SRT_{diff}-scores waren gerelateerd aan een beter werkgeheugen en kortere reactietijden bij de lexicale decisietest. De resultaten van de regressieanalyses met preoperatieve en postoperatieve factoren suggereren dat een kleinere verbale werkgeheugencapaciteit en, in mindere mate, een minder snelle lexicale toegang het verstaan van spraak beperken bij luisteraars met een CI.

De laatste studie, beschreven in Hoofdstuk 5, beoogde meer inzicht te krijgen in de interactie tussen maten van talige vaardigheden in het auditieve domein en het verstaan van spraak in het dagelijks leven. In de studies in hoofdstuk 3 en 4 kwam naar voren dat de relatie tussen *visuele* lexicale toegang en het verstaan van spraak in ruis minder duidelijk was bij CI-gebruikers dan bij luisteraars met een normaal gehoor. Hieruit bleek dat auditieve factoren het verstaan van spraak het best voorspellen bij CI-gebruikers, wat suggereert dat er mogelijk een interactie bestaat tussen de auditieve vermogens en talige vaardigheden. Voor CI-gebruikers is de kwaliteit van auditieve input verminderd, wat het verstaan van spraak nadelig beïnvloedt en wat van invloed kan zijn op de lexicale toegang. De studie in Hoofdstuk 5 had daarom als doel om het effect van verstoorde input te onderzoeken op de uitkomsten van een *auditieve* lexicale decisietest, nl. de reactietijd en de correctheid. Bij deelnemers met een goed gehoor werd de lexicale decisietest afgenomen met stimuli die het geluid van een CI simuleerden, geproduceerd via reconstructie met 6 of 10 ruisbanden. Voor CI-gebruikers werden de standaard auditieve stimuli gebruikt die door hen werden waargenomen via het CI. Uit het onderzoek bleek dat verstoorde auditieve input, veroorzaakt door gereconstrueerde spraak of CI-gebruik, leidde tot meer fouten en langere reactietijden bij de lexicale decisietest. De reactiepatronen bij deze taak bleken echter verschillend te zijn tussen de groepen. CI-gebruikers reageerden relatief traag op aangeboden pseudowoorden en neigden ernaar om een stimulus als 'bestaand' te classificeren. Dit patroon kwam niet voor in de groep deelnemers met een normaal gehoor die gereconstrueerde spraak kregen aangeboden. CI-gebruikers lijken dus gebruik te maken van een andere strategie voor lexicale toegang dan normaal horende luisteraars die naar gereconstrueerde spraak luisteren. Een mogelijke verklaring hiervoor is dat CI-gebruikers gewend zijn te luisteren naar verstoorde spraakstimuli. Correlatieanalyses met maten voor het spraakverstaan bij CI-gebruikers toonden een relatie tussen het verstaan van fonemen in eenlettergrepige woorden in stilte en de reactietijden op bestaande woorden bij de auditieve lexicale decisietest. Het verstaan van zinnen in ruis was meer gerelateerd aan een efficiënte balans tussen de correctheid van de responsen en de gemiddelde reactietijd, zowel bij aangeboden pseudowoorden als gemiddeld over alle stimuli. Deze resultaten tonen aan dat het meten van vaardigheden in lexicale decisie, zowel in de auditieve als in de visuele modaliteit, kan bijdragen aan het verklaren van de vaardigheid in het verstaan van spraak bij CI-gebruikers.

Het laatste hoofdstuk bespreekt de bevindingen van de beschreven onderzoeken en hun klinische implicaties. De onderzoeksuitkomsten bieden inzicht in de mogelijkheden en beperkingen van verschillende klinische tests voor het meten van het spraakverstaan bij CI-gebruikers. De resultaten tonen aan dat het meten van het verstaan van eenlettergrepige woorden in stilte bij CI-gebruikers zijn beperkingen heeft en dat adaptieve tests voor het verstaan van spraak in ruis meer en relevante informatie kunnen verschaffen. De DIN-test blijkt vooral goed geschikt om inzicht te geven in de auditieve aspecten van het verstaan van spraak in ruis en wordt minder beïnvloed door cognitieve en talige vaardigheden. De DIN-test kan daarnaast goed worden gebruikt om de ontwikkeling van de vaardigheden in het spraakverstaan na cochleaire implantatie te volgen. Omdat het verstaan van zinnen in ruis om meer talige en cognitieve verwerking vraagt, is de SIN-test beter bruikbaar om het auditief functioneren in alledaagse luistersituaties te evalueren. Als luisteraars onvoldoende verstaan om de standaard SIN-test uit te voeren, is het scoren van de test op basis van sleutelwoorden een alternatief. Uit de studies blijkt verder dat de relatie tussen talige en cognitieve vaardigheden en het verstaan van spraak met een CI complex lijkt en beïnvloed wordt door de auditieve vermogens. Voor CI-gebruikers die zwak zijn in het verstaan van spraak in ruis, lijkt de kwaliteit van de input het effectieve gebruik van goede talige en cognitieve vaardigheden te beperken. De studieresultaten geven aan dat het in kaart brengen van talige vaardigheden en werkgeheugen in de visuele modaliteit kan bijdragen aan het verbeteren van de voorlichting van CI-kandidaten voorafgaand aan de operatie. Daarnaast kunnen de resultaten van deze tests, in combinatie met auditieve maten van lexicale toegang, gebruikt worden bij het evalueren van de vaardigheden van CI-gebruikers in het verstaan van spraak.

D

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a c uch better
that ex his wid of perf
needed prove pre tive co
rehabilit programs. esis aim
betwee cally ap e linguis
perform with a c mplant. T
skills tha ase in people w
of current clinical measures of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory

Dankwoord

Mijn dank gaat uit naar alle mensen die mij, ieder met hun eigen inbreng, bij mijn onderzoek hebben gesteund en in staat hebben gesteld dit proefschrift te voltooien:

Mijn promotoren en copromotoren; Theo Goverts voor je onmisbare, fijne begeleiding, je helikoptervisie en dat je me steeds maar weer bleef uitdagen en inspireren; Joost Festen voor het bedenken van het onderwerp en je hulp bij analyses en correcties op de concept artikelen; Annette de Groot voor je waardevolle input vanuit de psycholinguïstiek en je uitgebreide opbouwende feedback; Cas Smits voor het vele meedenken over onder andere SRTs en je kritische blik;

Paul Merkus voor je co-auteurschap bij de artikelen uit hoofdstuk 2 en 4.

De leescommissie; prof. dr. Deniz Başkent, dr. Emmanuel Mylanus, prof. dr. Astrid van Wieringen, prof. dr. Sophia Kramer, prof. dr. ir. Wouter Dreschler en dr. Marte Otten voor jullie geïnvesteerde tijd, moeite en constructieve feedback.

Studenten Femke Langerhorst, Merel Latenstein, Tine Goossens, Astrid de Vos, Griet Mertens, Katrien Devroye, Jaap Peter Kronemeijer, Esther Timmer, Sietske Temminghoff, Sophie Radoux, Sofie Reynders-Frederix, Dick Biesheuvel, Silke Scheepmans en Lauren Blockmans voor het testen van een deel van de proefpersonen en jullie verdere bijdragen aan de verschillende studies tijdens jullie stages.

Alle ca. 180 proefpersonen voor jullie tijd en geleverde inspanning, want vermoeiend was het soms zeker.

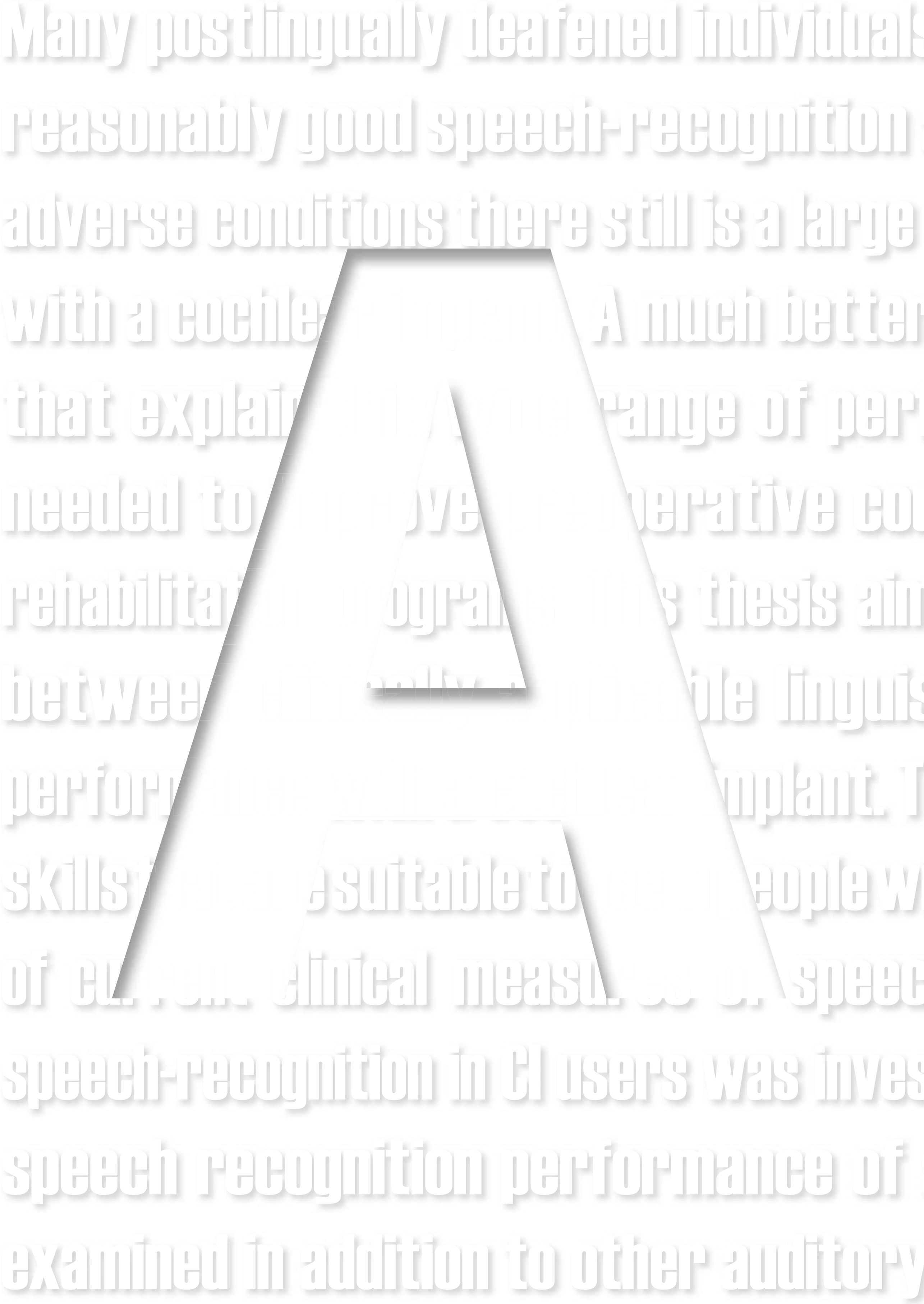
Birgit Witte voor je statistische ondersteuning; Hans van Beek voor het maken van de software voor diverse tests die gebruikt zijn in de studies; Mark Delreux voor het coverontwerp.

Mijn collega's van het audiologisch centrum en de KNO afdeling voor het werven van proefpersonen, jullie interesse en de gezelligheid tijdens het (klinische) werk; in het bijzonder Elly Amperse voor het inspreken van alle stimuli voor de 'Elly'-testen van de vierde studie uit dit proefschrift; Adriana Zekveld voor je hulp bij het definiëren van de keywords; Elke Huysmans voor alle tips, tricks en hulp bij het afronden van mijn proefschrift; Marieke Delreux en Elske Bolk voor jullie vriendschap, de fijne samenwerking, de lekkere koffie, het meedenken over onder andere de vormgeving, en jullie optreden als paranimf.

Pappa, voor het stimuleren tijdens school en opleiding en je vertrouwen. Je zou enorm trots zijn geweest.

Mijn vrienden, familie en vooral mijn gezin; Maarten, Tibbe, June, Joepe en mamma, dank voor jullie begrip, support en de welkome afleiding op de momenten dat ik het hard nodig had.

Ouderkerk aan de Amstel, december 2017


Many postlingually deafened individual
reasonably good speech-recognition
adverse conditions there still is a large
with a cochle
A much better
that explai
range of per
needed to
perative co
rehabilitat
s thesis ai
betwee
ble linguis
perfor
mplant. T
skills
suitable to
eople w
clinical measu
speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory
A

About the author

Over de auteur

ABOUT THE AUTHOR

Marre Kaandorp was born in Amsterdam on December 1st 1978 as the first of two daughters of Dick Kaandorp and Mieke Noordegraaf. When Marre was one year old the family moved to Utrecht and further eastward seven years later, where she grew up near the woods of Renkum.

She attended high school at SG Pantarijn (the former 'Wagenings Lyceum') in Wageningen from 1991 to 1997. In that year she started her studies Mechanical Engineering at the Technical University of Twente. In Enschede she enjoyed student life and was especially active as a board member and in several committees of the student swimming club 'Piranha'. The last two years of her studies she subspecialized in Biomedical Engineering, including a 3-month research-elective at RehabTech, the Monash Rehabilitation Technology Research unit in Melbourne, Australia. After graduating in 2003 she moved back to her birth town Amsterdam where she continued the research of her master thesis ('Development of a 3D visualization application for movement analysis') as a research fellow at the rehabilitation clinic of VU Medical Center (Research project: '3D visualization methods and exchange of movement data between movement analysis applications'). At the VU Medical Center her attention was drawn by the audiology center and she started her training to be a clinical physicist audiologist in 2005. After finishing this training in 2009 she continued working at the department and joined the ENT-audiology staff of VUmc in 2011, where she still works with great enthusiasm. During her training she initiated the current research under supervision of prof. dr. ir. J.M. Festen and dr. S.T. Goverts which resulted in this thesis.

Marre is happily married to Maarten Quist with whom she has two children; Tibbe (2008) and June (2010). They live in Ouderkerk aan de Amstel.

Joepe Kaandorp

OVER DE AUTEUR

Marre Kaandorp werd geboren op 1 december 1978, als eerste van twee dochters van Dick Kaandorp en Mieke Noordegraaf. Toen Marre één jaar oud was verhuisde het gezin naar Utrecht om zeven jaar later verder richting het oosten te verhuizen, alwaar ze opgroeide dichtbij de bossen van Renkum.

Ze ging naar het VWO op SG Pantarijn (voormalig Wagenings Lyceum) in Wageningen van 1991 tot 1997 en in datzelfde jaar begon ze aan haar studie werktuigbouwkunde aan de Universiteit Twente. In Enschede genoot ze van het studentenleven en was actief in het bestuur en verschillende commissies van studenten zwemvereniging 'Piranha'. De laatste twee jaar van haar studie specialiseerde ze zich in de biomedische werktuigbouwkunde. In het kader van deze specialistatie deed ze een wetenschapsstage van 3 maanden bij RehabTech aan de Monash Rehabilitation Technology Research unit in Melbourne, Australië. Na haar afstuderen in 2003 verhuisde ze terug naar haar geboortestad Amsterdam om daar het onderzoeksproject ('Development of a 3D visualization application for movement analysis') van haar afstuderen voort te zetten als junior onderzoeker bij de revalidatiekliniek van VU medisch centrum (onderzoeksproject: '3D visualization methods and exchange of movement data between movement analysis applications'). In het VUmc werd haar aandacht getrokken door het Audiologisch Centrum en begon ze haar opleiding tot klinisch fysisch-audioloog in 2005. Na het afronden van deze opleiding in 2009 bleef ze werken in het Audiologisch Centrum van het VUmc en werd ze lid van de KNO-audiologie staf in 2011, waar ze nu nog steeds met veel enthousiasme werkt. Tijdens haar opleiding begon ze aan haar promotieonderzoek onder supervisie van prof. dr. ir. J.M. Festen en dr. S.T. Goverts, hetgeen resulteerde in dit proefschrift.

Marre is gelukkig getrouwd met Maarten Quist, met wie zij twee kinderen heeft; Tibbe (2008) en June (2010). Ze wonen in Ouderkerk aan de Amstel.

Joepe Kaandorp

Many postlingually deafened individuals
reasonably good speech-recognition
adverse conditions there still is a large
with a cochlear implant. A much better
that explain this wide range of perf
needed to improve preoperative cou
rehabilitation programs. This thesis aim
between clinically applicable linguis
performance with a cochlear implant. T
skills that are suitable to use in people w
of current clinical measures of speec
speech-recognition in CI users was inves
speech recognition performance of
examined in addition to other auditory