Aarno Dietz

The Development of the Finnish Matrix Sentence Test

BIS-Verlag der Carl von Ossietzky Universität Oldenburg

Oldenburg, 2015

Verlag / Druck / Vertrieb

BIS-Verlag
der Carl von Ossietzky Universität Oldenburg
Postfach 2541
26015 Oldenburg
E-Mail: bisverlag@uni-oldenburg.de

ISBN 978-3-8142-2327-8

Introduction

Understanding speech is the base of our culture – and not being able to understand separates us humans from each other (as Kant already stated in a precise way). Hence, caring for methods to precisely assess the individual performance in understanding speech is of high practical and clinical importance. Moreover, it is highly desirable if the same (or at least very similar) speech recognition measures are applicable to several languages with the aim to make the clinical and research results obtained in one language directly transferable to results for another language.

Since connecting the Finnish and the German culture and language is a well-established and trained behavioral pattern by Aarno Dietz (who is a native speaker of both languages), it takes no wonder that in his dissertation he has selected to bring the most suitable German speech recognition test (the OLSA or Oldenburger Satztest) to the Finnish language. For those who think that this is an easy, straight forward task, please read the thesis by Aarno Dietz yourself! You will find that this endeavor is not just a simple translation job, but instead a masterpiece that took into account a large number of aspects: The selection of the most familiar and phonematically appropriate words to be used, the speaker to be selected, the details of the recording and subsequent verification and optimization of the speech materials and the final selection of the most appropriate speech test parameters are by far not trivial – and Aarno Dietz has mastered these tasks in an excellent way – convince yourself!

Establishing the OLSA in Finnish (denoted as the Finnish Matrix test) constitutes, though, not just a simple bilateral connection between Finland and Germany. Instead, compatible versions with the German OLSA test have meanwhile been developed for at least 14 languages with very similar and hence highly comparable test results. Hence, Aarno Dietz brings in his new test into a whole family of languages (among them such important languages as American and British English, Spanish, French, Turkish and German) and hence gets Finnish audiology connected to international speech audiology – what a great outcome of a dissertation project!

However, Aarno Dietz would hide his main profession, i. e. being an ORL Doctor and surgeon for cochlea implants, if he would not utilize this newly developed test to assess the indication for and the achieved benefit from a cochlea implant operation for a group of his clinical patients. Again, his results are highly comparable to similar results in the German language with patients from Germany (primarily from Hannover) – and this is the ultimate proof that Aarno Dietz's work is not just plain theory and the development of some irrelevant speech materials, but instead promises to be of high clinical value for the treatment of many patients!

Among the many unique features of the current dissertation work, there is one which cannot be derived by reading the thesis itself: Aarno Dietz is the first medical doctor to be promoted by the new Faculty of Medicine and Health Sciences at Oldenburg University (as part of the European Medical School Oldenburg-Groningen). Moreover, he is the first Medical Doctor (MD/PhD) that I had the honor to supervise as primary supervisor in the field of medicine (after supervising more than 50 PhD theses in the field of physics and engineering). It is therefore my great pleasure to certify that the work produced by Aarno Dietz is among the finest PhD theses for the medical doctorate to be found in Germany – please read yourself and you will get convinced!

Needless to say that – besides being a great clinician and researcher – Aarno Dietz is also a great colleague and friend and a caring family father. The amazing care and devotion to all goals and tasks that Aarno is performing (mostly simultaneously) again becomes transparent from his thesis work – please read yourself and get impressed!

Oldenburg, January 2015, Birger Kollmeier

Abstract

Until the present, no speech audiometric test utilizing sentences in interfering noise has been available for the Finnish language. In this thesis, the development of the Finnish matrix sentence test in noise is described in detail and compared with the matrix tests in other languages. Additionally, the characteristics and properties of the new Finnish test were investigated in cochlear implant recipients.

The difficulty of understanding speech in interfering noise is usually the first and most relevant symptom of hearing impairment. Speech audiometric tests are used in the diagnostics of hearing impairment, since these reflect the subject's hearing performance better than the traditionally measured pure-tone thresholds. Speech tests in interfering noise are used nowadays to achieve a better simulation of everyday hearing situations. In addition to determining the attenuation component of the hearing impairment, these speech tests also help to assess suprathreshold distortions occurring in the auditory system, which may further reduce the speech recognition performance. In addition to their use in audiology diagnostics, speech tests are also important for the verification of rehabilitation outcomes with hearing aids and implantable hearing devices (middle ear and cochlear implants) as well as for research applications.

During the development of a new Finnish speech audiometric test in noise, special attention was paid to incorporating comprehensive clinical applications as well as to achieving an internationally comparable test procedure. The minimum requirements for new audiometric tests were established by the European Union funded HearCom project and its successor HurDig. Therefore, a so-called matrix sentence test was developed for the Finnish language. The matrix sentence test was originally developed by Björn Hagerman (1982) for the Swedish language and this was further developed in the Institute of Physics at the University of Oldenburg *(Wagener et al.; 1999a-c).* In contrast to the original Hagerman test, the procedure for the Oldenburg sentence test (OLSA) accounts for co-articulation between the words in order to maintain the naturalness and prosody of the sentences. The OLSA fulfills the Hear-

Com standards and has already been adapted for many different languages. It is widely used in audiologic centers where it has clinical and research applications.

The base matrix of the Finnish test consists of 10x5 simple, frequently used words with the representative phoneme distribution. The speech material was recorded by a female news anchor. Special attention was paid to the naturalness, clarity, volume and speaking rate during the recording sessions. Similarly as for the OLSA, the recording procedure accounts for co-articulation between the words. Accordingly, 100 sentences including all possible combinations of two consecutive words, were recorded at least two times. The concept of preserving co-articulation produces more naturally sounding resynthesized sentences. All of the desired combinations of the word matrix could be realized. During the optimization procedure, the speech intelligibility was balanced across the individual words of the newly resynthesized sentences. In the following evaluation measurements, the intelligibility function of the test was determined. The equivalence between the different test lists was also checked. From the performed measurements, the reference values for normal hearing subjects on adaptive measurements were determined.

The Finnish matrix sentence test is the first speech audiometric sentence test in noise for the Finnish language. It is an accurate new speech audiometric test, which provides internationally highly comparable test results. In comparison to other existing matrix tests, the Finnish test displays a lower speech recognition threshold in normal hearing persons, which can be accounted for by language- and speaker-related factors. The slope of the intelligibility function, however, is highly comparable with the values obtained in the other tests.

In a first application study, the Finnish matrix sentence test proved to be suitable for measurements of the speech intelligibility in noise in cochlear implant recipients. The results were compared with previously published data of patients tested with the OLSA. Similar speech-reception thresholds were measured for cochlear implant users in both countries. Furthermore the correlation was evaluated between the new Finnish matrix sentence test and the established Finnish word test conducted under quiet conditions.

Zusammenfassung

In der vorliegenden Dissertation wird die Entwicklung eines sprachaudiometrischen Satztestverfahrens im Störgeräusch für die finnische Sprache beschrieben. Da es bisher keinen finnischen Satztest im Störgeräusch gab, wurde ein finnischer Matrixtest nach Vorbild anderer bereits verfügbarer Satztestverfahren entwickelt. Die Entwicklung des finnischen Matrixtests wird im Detail beschrieben und mit Matrixtests anderer Sprachen verglichen. Zusätzlich wird die klinische Anwendung des neuen finnischen Tests anhand von mit Cochlea Implantaten versorgten Patienten beschrieben.

Die meisten Menschen bemerken ihre Schwerhörigkeit erst durch Probleme bei der Kommunikation bei Nebengeräuschen und Störschall. In der Diagnostik von Hörverlusten werden neben dem Reintonaudiogramm auch sprachaudiometrische Verfahren verwendet, die die funktionelle Beeinträchtigung des Hörschadens widerspiegeln. Um die Alltagssitutationen besser zu erfassen, werden zunehmend Satztestverfahren im Störschall verwendet. Man kann damit außer der Abschwächungs-komponente des Hörverlustes auch eine mögliche Verzerrungskomponente ermitteln, die zu einer zusätzlichen Verminderung der Spracherkennung führen kann. Darüber hinaus sind Satztestverfahren im Störschall wichtig für die Überprüfung von Rehabilitationsergebnissen mit Hörgeräten und implantierbaren Hörhilfen (Mittelohrimplantate und Cochlea Implantate).

Bei der Entwicklung des finnischen Satztestverfahrens im Störschall wurde einerseits Wert auf ein vielseitiges Anwendungsspektrum gelegt und anderseits auf eine internationale Vergleichbarkeit. Die Anforderungen für neue Sprachaudiometrie wurden in den von der EU geförderten Projekten HearCom und Hur-Dig definiert. Daher wurde für den neuen finnischen Satztest der sogenannte Matrixtest gewählt. Der Matrixtest wurde ursprünglich von Björn Hagerman (1982) für die schwedische Sprache entwickelt und darauf basierend an der Abteilung medizinische Physik der Universität Oldenburg weiterentwickelt *(Wagener et al., 1999a–c)*. Bei der deutschen Version des Matrixtests, des sogenannten Oldenburger Satztests (OLSA) wurde bei der

Entwicklung das Konzept der Koartikulation eingeführt, um die Natürlichkeit und Prosodie der Sätze zu erhalten und zu gewährleisten. Der OLSA ist inzwischen im ISO Standard aufgenommen worden und wird sowohl im klinischen Alltag als auch im Forschungsbereich angewendet.

Die Basismatrix des finnischen Matrixtests besteht aus 10x5 einfachen, häufig vorkommenden Wörtern mit für die finnische Sprache repräsentativer Phonemverteilung. Das Sprachmaterial wurde von einer finnischen Nachrichtensprecherin aufgesprochen, wobei auf Natürlichkeit, Klarheit und gleichbleibende Sprechgeschwindigkeit und Lautstärke geachtet wurde. Die aufgenommenen Sätze wurden so verarbeitet, dass unter Berücksichtigung der Koartikulation alle gewünschten Kombinationen der Matrix zusammengesetzt werden konnten. Mit den neu zusammengesetzten Sätzen fanden die Optimierungsmessungen statt, die dazu genutzt wurden, die Verständlichkeiten aller Einzelwörter anzugleichen. Nach der Optimierung der Sätze wurden Evaluationsmessungen durchgeführt, um die Bezugskurve des Tests zu bestimmen und die Gleichwertigkeit der Testlisten zu überprüfen. Aus den bei der Evaluation durchgeführten Messungen konnten zudem die Bezugswerte für Normalhörende bei adaptiven Messungen ermittelt werden.

Mit dem finnischen Matrixtest steht nun ein akkurates und international vergleichbares Sprachtestverfahren in finnischer Sprache zur Verfügung. Der neue Matrixtest reiht sich in die Matrixtests verschiedener Sprachen ein, die nach gleichen Standards und Methoden entwickelt wurden. Im Vergleich mit den bestehenden Matrixtests stellt sich heraus, dass zwar die Steigung der Sprachverständlichkeitsfunktion sehr gut mit anderen Matrixtests vergleichbar ist, allerdings liegt die Sprachverständlichkeitsschwelle etwas niedriger. Dies lässt sich durch Sprach- und Sprecher-bezogene Faktoren erklären.

In einer ersten Anwendungsstudie zeigte sich der finnische Matrixtest, ebenso wie der OLSA, geeignet für die Messungen von Patienten mit Cochlea Implantaten. Vergleichbare Messresultate werden für die Patienten beider Länder ermittelt. Zusätzlich wurde die Korrelation zwischen dem neuen Matrix Test und dem etablierten finnischen Worttest ermittelt.

Contents

* Includes text passages that have been cited from the original publication

1 Introduction

1.1 General Introduction and Motivation

Understanding speech is probably the most important task of human hearing. The healthy auditory system can differentiate between many different noises and sounds and this enables us to understand speech even in an environment where the noise level is actually higher than the speech level. Accordingly, the difficulty to comprehend speech in the presence of background noise is the most common complaint of persons with hearing impairment *(Dirks et al, 1982; Smoorenburg & van Golstein-Brouwers, 1986; Kramer SE et al., 1998; Theunissen et al., 2009).* Problems of understanding speech in a noisy environment can be much more severe than would be expected on the basis of the pure-tone thresholds (*Smoorenburg GF, 1992; Nielsen & Dau, 2009).* It has also been demonstrated that subjects with similar degrees and configurations of hearing impairment show very different abilities to understand speech in the presence of interfering noise *(Plomp 1978; Crandell 1991, Smoorenburg GF 1992).* Since typical everyday situations are most often noisy, the assessment of the patient's ability to cope with these situations is therefore crucial in order to ensure adequate hearing rehabilitation. The traditional assessment of the hearing impaired patient, which is based on the evaluation of pure-tone thresholds, does not adequately measure the function and performance of the auditory system (Carhart & Tillman, 1970; Nilsson et al., 1994; Wilson et al., 2007). Therefore the recommendation to include speech-in-noise tests into the standard audiologic examination was presented as early as 1970 by Carhart and Tillman.

Today, the necessity to perform speech-in-noise tests is undisputed in audiologic diagnostics, rehabilitation and research applications.

Sensorineural hearing impairment can affect the speech intelligibility in two distinct ways. There are investigations indicating that suprathreshold distortions in auditory processing may make even greater contributions to deficits in speech recognition than the direct attenuation of the sound signal in the hearing impaired patients *(Plomp, 1978; Glasberg & Moore, 1989; Middelweerd* *et al., 1990; Smoorenburg, 1991; Kollmeier, 1998; Van Summers et al., 2013; Bernstein et al., 2013).* Recently, it was shown by van Summers et al. (2013) that high-frequency hearing impairment might be associated with distortions in sound processing even at the lower frequencies. Thus it appears that the traditional audiometric measurements (pure-tone thresholds and word-recognition in quiet) are only of marginal use when evaluating a patient's functional hearing performance and this reservation extends to the assessment of the indications for rehabilitation. Consequently, individuals susceptible to background noise may gain only limited benefit from conventional amplification with hearing aids, which only restores audibility to the frequencies where the impairment is present. They may, however, benefit from more sophisticated devices with advanced noise reduction algorithms *(Hohmann & Kollmeier, 1995).* Speech recognition tests in background noise are therefore important for the verification and quantification of hearing rehabilitation results.

The use of background noise makes possible more complex binaural measurements. Measurements where the speech signal and the noise are presented from different directions simulate everyday listening situations better than measurements in which both signals are originating from the same source. The measured binaural parameters are the intelligibility level difference (ILD) and the binaural intelligibility level difference (BILD) (*Levitt & Rabiner, 1967; Bronkhorst & Plomp, 1989).* Only with these types of measurement can the possible binaural benefit offered by the rehabilitation of both ears be determined. The ILD quantifies the capabilities of a listener to separate a signal coming from a different direction than the background noise. The ILD is the difference of the SRT where the signal originates from the front and the noise is from the side ($S_0N_{\pm 90}{}^{o}$) and the SRT where noise and signal are both coming from the front (S_0N_0). In normal hearing subjects, the improvement is about 6-12 dB depending on the acoustics of the sound field room; this is explained by the head shadow effect and binaural processing in the brain. The BILD measurement can be used to differentiate between the head shadow effect and binaural processing. In this measurement, the binaural processing benefit is measured by excluding the head shadow effect. The BILD is quantified by the difference of the SRT at the $S_0N_{\pm 90}{}^{o}$ presentation as compared to the same presentation (binaural presentation) but with the ear that is directed to the noise source plugged (monaural presentation). When binaural processing is undisturbed the result without the plug is usually 3–6 dB better than with the plug and this is a reflection of the binaural processing occurring in the brain. The possibility to examine spatial hearing makes these tests suitable for a large variety of applications including hearing research, room acoustics and speech transmission systems. *(Levitt & Rabiner, 1967; Oldenburger Satztest: Handbuch und Hintergrundwissen, 2000)*

To date, no evaluated and validated speech test in noise has been available in the Finnish language. The aim of this thesis was to develop a matrix sentence test for the Finnish language. During the development of the test, the same principles were used as established by the EU-project HearCom and its successor HurDig, with the goal of yielding a test that would be comparable to those developed for other languages.

1.2 Review of the literature

1.2.1 Different types of speech tests

Different types of speech intelligibility tests in noise have been developed in many countries and for many different languages. The existing speech intelligibility tests differ not only because they use different languages but also since they exploit a large variety of methodical parameters and presentation modes. As a consequence, none of these tests are directly comparable since they vary extensively in their normative data, namely in the speech-recognition-threshold and the steepness of the intelligibility curve function which are considered as the most important speech intelligibility measures. An even more complex problem is the test comparability between different languages, since language specific factors may contribute crucially to the speech intelligibility in the background noise *(Wagener & Brandt, 2005; Zokoll et al., 2012*). Since the speech tests are so different, it is very difficult to make a valid comparison between international studies since it remains unclear whether the measured differences are attributable to the effect under investigation or due to the test language, test procedure or the subject group. Therefore normative data must be available in order to adequately interpret given test results for their relevancy. One parameter of particular importance is the steepness of the intelligibility function since this can help the investigator to interpret the measured SRTs correctly *(Wagener et al, 1999a–c; Theunissen, 2009).*

In the earliest tests, which were based on short words presented in noise, the intelligibility score was calculated as the percentage of correctly repeated words *(Lutman, 1997).* These tests, however, were not optimal for the assessment of a listener's ability to follow conversational speech, since isolated words lack the essential characteristics of spoken language, such as word transitions, temporal fluctuations, normal spectral weighting, intonation and prosody. The listener's ability to exploit semantic and syntactic cues as well as the redundancy is not considered when using short words *(Plomp & Mimpen, 1978; Nilsson et al., 1994; Wagener et al., 1999a–c; Nielsen & Dau, 2009)*. Additionally word tests were found unsuitable for advanced measurements such as those used for hearing aid or cochlear implant fittings, since the duration of the presented word may not have been sufficiently long to allow the sound processing algorithms to take full effect *(Nilsson et al., 1994; Nielsen & Dau, 2009, Muller-Deile, 2009, Kollmeier et al., 2014)*. In the clinical setting, when frequent re-testing is often necessary, there is also a high risk of familiarization and learning of the speech material due to the limited number of words, thus reducing the test re-test reliability *(Nilsson et al., 1994; Wagener et al., 1999a–c; Theunissen et al., 2009, Kollmeier et al., 2014)*.

Therefore sentence-length test material has been developed in order to simulate more closely everyday situations. These sentences also provide for very detailed measurements, since several stimuli are being tested during the same trial. The speech perception in noise test (SPIN) was one of the first tests using sentences as speech material *(Kalikow et al., 1977)*. This test was designed to measure the intelligibility in percentage terms at fixed signal-noise-ratios (SNR). Though providing reliable measurements, the percent intelligibility measures are inherently limited by floor and ceiling effects, especially when the test subject's performance deviates substantially from the norm *(Lutman, 1997; Nilsson et al., 1994)*.

The alternative to the percent intelligibility is to measure the speech-reception threshold (SRT). The SRT is defined as the presentation level, expressed as the signal-noise-ratio, at which the test subject can recognize 50 % of the speech material correctly. SRT measurements are not subject to the aforementioned floor or ceiling effects *(Plomp, 1978; Plomp & Mimpen, 1979; Levitt 1978)*. From the test subject's point of view, speech audiometry designed for the determination of the SRT feels equally difficult, independently of whether the test person has a hearing impairment or not. The technique for SRT determination is derived from adaptive measurements, where the presentation level of the stimuli is decreased or increased, depending upon the test subject's ability to repeat the material correctly *(Plomp & Mimpen, 1979; Levitt 1978, Brandt & Kollmeier, 2002)*. In this way, adaptive measurements effectively place the presentation levels into the region of the test subject's SRT and over a sequence of trials, it is possible to conduct an accurate SRT estimation which is done by averaging the presentations levels in the latter part of the testing sequences *(Levitt, 1978; Nilsson et al., 1994; Brandt & Kollmeier, 2002).* In order to make an accurate SRT estimation during adaptive measurements, the speech material must differ but remain of equal known difficulty so that it cannot be memorized by the listener. In settings where frequent re-testing is required (e.g. rehabilitation, research), a very large pool of test sentences with equal intelligibility has to be available.

1.2.2 The Plomp-type sentence test

The speech intelligibility test in noise developed by Plomp & Mimpen (1979) represents the ground work that many subsequent tests, such as the Hearing-in–noise-test (HINT), referred to during their development *(Nilsson et al., 1994; Wong & Soli, 2005; Hällgren et al., 2006; Wong et al., 2007; Jansen et al., 2012; Nielsen & Dau, 2009)*. These Plomp-type tests make use of meaningful everyday sentences, maintaining the pronunciation and content characteristics of conversational speech. Though being efficient for diagnostics due to natural sounding sentences and accurate SRT estimation, these tests show a high degree of redundancy, thus limiting their use in settings when frequent re-testing is required, such as in research and rehabilitation applications *(Wagener et al., 1999a-c).* Therefore a very large pool of different sentences must be available to prevent them being memorized by the test subjects. If one wishes to be sure of obtaining reliable results, then it is not possible to repeat measurements with the same test list until sufficient time has passed with the usual recommendation being at least 6 months. Different Plomp-type tests have been developed for different languages. Due to considerable variations in test parameters and presentation modes, it is not possible to make reliable comparisons across the languages. Table 1 summarizes some of the variables which influence the results of speech in noise tests.

Table 1: Variables influencing the speech-reception threshold of speech audiometry.

Stimulus	*Speech material*	Style and Content Phonemes/Words/Sentences Meaningful vs nonsense words/sentences
	Type of noise	long-term spectrum of noise stationary vs fluctuating
	Speaker	male vs female, articulation, clarity educated (professional) speaker
Presentation	*Signal level* *Noise level* *Noise presentation*	 continuous vs gated noise
Response	*Open vs closed set* *Scoring method*	 sentence-, keyword- or word scoring
Subject variables	*Hearing impairment* *Auditory processing* *Age* *Cognition* *Language*	

Figure 1 illustrates the different speech recognition curve functions of different Plomp-type sentence tests. Soli & Wong (2008) addressed this problem by setting standards for the development of the HINT-type test so that it would provide comparable measures of speech intelligibility in noise for each language. Nevertheless significant variations in the SRT measures and more importantly in the steepnesses in the discrimination curve function occurred, -2.6 to -4.7 dB (SRN) and 9.0–14.6 %/dB, respectively, meaning that comparability across the languages is far from optimal.

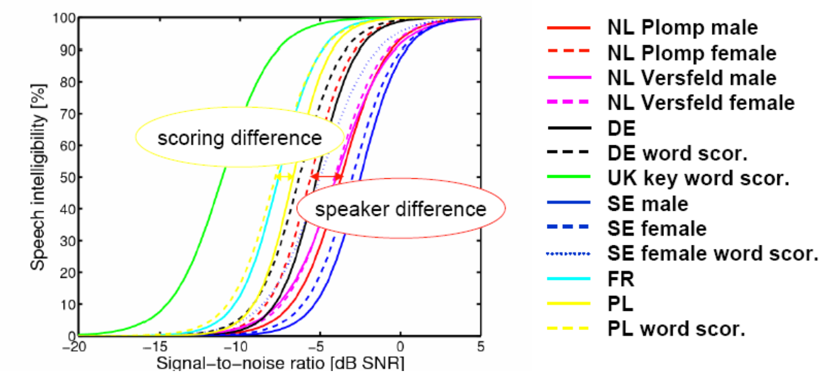


Figure 1: Plomp-type sentence tests. Cross-language comparison.
Wagener, K. Multilingual speech test in several European countries (www.HearCom.eu)

1.2.3 The matrix sentence test

Another approach which has been adopted in the development of a more standardized speech test in noise is the so-called matrix test, which is based on syntactically fixed sentences (e.g., name-verb-numeral-adjective-object for the English language) selected from a 5 x 10 word matrix. The first matrix test was originally developed by Björn Hagerman and it was done in Swedish *(Hagerman, 1982)*. The sentences are formed randomly by combining each word at one position within the sentence with any other word in the neighboring position(s), yielding 10^5 possible different sentences. Since the speech material consists of only 50 words that can be recorded and perceptually optimized in an appropriate way *(Wagener et al., 1999a and b)*, it exhibits high homogeneity, which results in a discrimination curve function of high steepness. A high steepness in the intelligibility function is desirable in order to be able to measure also subtle changes in the hearing performance. In the original Hagerman test, only the sentences in the base matrix were recorded and word transitions were avoided since this was thought to facilitate the re-synthesizing of the sentences *(Hagerman, 1982)*. Due to omission of the word transitions, the newly formed sentences were unnatural sounding *(Nilsson et al, 1994; Nielsen & Dau, 2009)*. Further refinements of this test were devised by Wagener et al. for the German language, when the concept of preserving co-articulation was introduced into the development of the test, providing natural prosody in the newly synthesized sentences *(Wagener et al., 1999a-c)*. Due to the practically unlimited amount of sentences, in addition to being valuable as a diagnostic tool, the matrix test is suitable also for applications when repeated testing is required, such as research and rehabilitation when the subjects are usually tested several times under different conditions. One reported disadvantage, however, is the training effect which is encountered even in experienced listeners *(Nilsson et al., 1994)*. In order to reduce the impact of training on the SRT, it is necessary to start each examination with two training sessions *(Wagener et al., 1999a–c)*. Figure 2 compares the intelligibility function of Matrix tests for different languages.

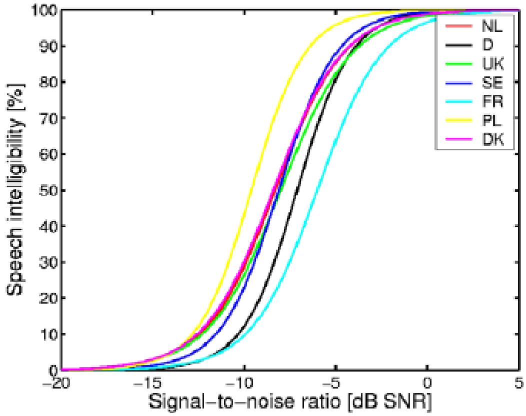


Figure 2: Matrix sentence tests. Cross-language comparison.
Wagener, K. Multilingual speech test in several European countries (www.HearCom.eu)

One of the aims of the HearCom-project and its successor HurDig, funded by the European Union, sought to achieve the harmonization of hearing diagnostic tests throughout Europe. Minimum quality requirements were established in order to reach highest comparability in testing results between the European countries. Since there will be free movement of patients within the European Union, it is most important that there should be a standardization of the therapy indications in the Member States across Europe. These aforementioned quality requirements were implemented in the Oldenburg measurement application, which also incorporated a common measurement software platform, the so-called Oldenburg measurement application. These newly developed speech audiometry tests for different languages all fulfill the HearCom and HurDig standards and are therefore highly comparable between the different languages.

1.2.4 *Finnish speech audiometry*

To date, there has been no sentence test in background noise available for the Finnish language. In Finland, the only commonly available speech audiometric test for clinical use consists of isolated bi-syllabic meaningful words presented in quiet. This word test was analyzed for word and phoneme discrimination by Jauhiainen *(1974)*. Bi-syllables were selected because there are not enough monosyllables with which to build a word audiometry test in the Finnish language. The speech material for this test was developed as early as 1952 by Palva. In 1968, these words were newly recorded and some of most old-fashioned words were omitted *(Jauhiainen, 1974)*. The final speech material consisted of six lists of 30 words. Unfortunately at the time of the selection and recording of the words in 1968, no data was available on the word frequencies or the phoneme distribution in Finnish. A few years later, Pesonen (1971) analyzed a representative sample of newspaper texts for word frequency and found that only two words which were represented in the 400 most frequent Finnish newspaper words actually had been included in the test lists. The speech material was analyzed according to the phoneme distribution and it was found that the correlation coefficient was 0.776 based on the phoneme frequencies in a selection of Finnish literary material *(Mikkonen, 1969; Pesonen, 1971, Jauhiainen, 1974)*. Later, in order to make the word discrimination thresholds more uniform, corrected word lists were compiled, where the five least discriminated words were dropped, thus further compromising the phonetical balancing of the final word lists. Table 2 shows an example of one of seven test lists. The greatest disadvantage of the Finnish words is that the test lists were not evaluated according to their perceptive equivalence. As a consequence, the test results differ depending on which list is being tested. In addition, due to the limited speech material, this test exhibits a very high degree of redundancy. Although the Finnish word test does achieve accurate measurements at the threshold level, for suprathreshold measurements, a ceiling effect occurs even in patients with a severe hearing impairment as well as in cochlear implant users.

Due to the constraints of a speech test performed in silence there have been a few attempts made to develop a speech in noise test for the Finnish language *(Palva, 1955, Jokinen, 1973, Pekkarinen, 1988)*. These tests used tape recordings with fixed signal-to-noise ratios and were very time-consuming and were never implemented into clinical work. A computerized adaptive Finnish speech in noise test was later developed using newly chosen and phonetically balanced isolated bi-syllabic words with superimposed noise *(Laitakari & Laitakari, 1997, Laitakari, 1996, 2001, Laitakari & Uimonen, 2001)*. The methods described by Plomp and Mimpen (1979) were adapted during the development of this test. Unfortunately this test was never evaluated and validated according to the requirements nowadays demanded for new test procedures and a necessary requirement were it to be implemented in the clinics across the country.

Table 2: Example of one of the word lists (List 1) used in the Finnish word test in silence

purje	sail	virta	stream
seos	mixture	vattu	raspberry
tamma	mare	menkööt	go
ydin	core	järvi	lake
viikset	moustache	kitsas	miserly
kärppä	weasel	tolppa	pole
säie	fibre	erä	batch
joulu	Christmas	kaaos	chaos
kahle	bond	silmä	eye
rinta	breast	pistos	sting
temppu	trick	uhri	victim
pyörä	wheel	purra	bite
metri	metre		

2 The development of the Finnish Matrix Sentence

2.1 Introduction

The inability to understand speech in the presence of background noise is the most common complaint of persons with a hearing impairment *(Lutman et al., 1987; Kramer et al., 1998)*. The difficulty of understanding speech in a noisy environment can be much more severe than would be expected on the basis of the audiogram findings *(Lutman et al., 1987; Smoorenburg, 1992)*. Since typical everyday situations are most often noisy, the assessment of the patient's ability to cope with these situations provides the clinician with a more concise understanding of his/her problem. It also helps to assess any suprathreshold distortions occurring in the auditory system due to hearing impairment that are mostly independent from the sensitivity loss as assessed by the tone audiogram *(Plomp, 1978; Glasberg & Moore, 1989; Middelweerd et al., 1990; Smoorenburg, 1991; Kollmeier, 1998; Van Summers et al., 2013; Bernstein et al., 2013)*. Different types of speech audiometry have been used for the assessment of speech recognition. Apart from audiology diagnostics, speech recognition tests are also of crucial importance for hearing rehabilitation, hearing research, including room acoustics and speech transmission systems. Therefore, different types of speech intelligibility tests in noise have been developed in many countries and for many languages *(Plomp & Mimpen, 1979; Hagerman, 1982; Kollmeier & Wesselkamp, 1997; Wagener et al., 1999; Hällgren et al., 2006; Wilson et al., 2007; Wong et al., 2007; Soli & Wong, 2008; Luts et al., 2008; Shiroma et al., 2008; Nielsen et Dau, 2009; Ozimek et al., 2010; Hochmuth et al., 2012; Jansen et al., 2012; Wagener et al., 2003)*.

Speech intelligibility tests may differ in the speech material used (e.g., syllables, digits, words, sentences) and the homogeneity of the test lists and the steepness of the speech discrimination function *(Theunissen et al., 2009)*. The discrimination function can be derived from the proportion of correct responses at different signal-to-noise ratios (SNR). Using sentences for speech audiometry means that the test is more representative of a realistic communication situation than simply listening to words or syllables. Since several test items can be presented within the same trial, sentence tests allow for a very efficient measurement of the speech-reception threshold (SRT), which corresponds to the SNR at which 50 % speech recognition is achieved *(Plomp & Mimpen, 1979; Nilsson et al., 1994; Soli & Wong, 2008;Theunissen et al., 2009)*.

There are two types of sentence tests commonly used. The first type of test makes use of meaningful everyday sentences. The test proposed by Plomp and Mimpen (1979) for the Dutch language is an example of this kind of test as is the Hearing in noise tests (HINT) for English *(Nilsson et al., 1994)*. These kinds of speech audiometric tests with meaningful sentences have been further developed for many other languages *(Kollmeier & Wesselkamp, 1997; Nilsson et al., 1994; Hallgren et al., 2006; Wong et al., 2007; Soli & Wong, 2008; Luts et al., 2008; Shiroma et al., 2008; Nielsen et Dau, 2009)*.

Though being efficient for diagnostics due to natural sounding sentences and accurate SRT estimation, these so called Plomp-type tests suffer from a high degree of redundancy, thus limiting their use in settings when frequent retesting is required, such as research and rehabilitation applications *(Wagener et al., 1999a–c)*. The second type of test is based on syntactically fixed sentences (e.g., name-verb-numeral-adjective-object for the English language) selected from a 5 x 10 word matrix. The first matrix test was developed by Hagerman for Swedish *(Hagerman, 1982)*. Further refinements of this test were made by Wagener et al. for the German language, when the concept of preserving co-articulation was introduced in the development of the test, providing natural prosody to the synthesized sentences *(Wagener et al., 1999a–c)*. Although the speech material of the word matrix is limited, each word at one position within the sentence may be combined with any other word in the neighboring position(s), yielding 100 000 possible different sentences. Since it can generate a virtually limitless number of sentences, the matrix test is suitable also for applications when repeated testing is required, such as research and rehabilitation when the subjects are usually tested several times under different conditions. One reported disadvantage, however, is the training effect even in experienced listeners *(Wagener et al., 1999a–c; Hochmuth et al., 2012; Jansen et al., 2012, Wagener et al., 2003),* although this can be reduced by starting each examination with two training sessions.

At the present time, there is no sentence test in background noise available for the Finnish language. The most common speech audiometric test in clinical use consists of isolated bi-syllabic words presented in quiet *(Jauhiainen, 1974)*. Although balanced, the word lists were not evaluated according to their perceptive equivalence, so that the test results may differ depending on which list is being tested. Additionally, due to the limited word material this test has been found to be unsuitable for frequent retesting. Therefore a sentence in noise test was developed for the Finnish language.

2.2 The Design of the Finnish Matrix Sentence Test

2.2.1 Development of the base matrix

The design and the selection of the speech material follow the same structure as the Swedish test of Hagerman *(Hagerman, 1982)*. The composition of the matrix consists of 10 names, 10 verbs, 10 numerals, 10 adjectives and 10 objective nouns, e.g., "Sofia pyysi kolme punaista sukkaa" ("Sofia asked for three red socks").

The words included in the matrix are chosen to be commonly recognized by all Finnish speakers and to contain all phonemes at a similar frequency as in everyday spoken language. Special attention was paid to the semantic neutrality and the familiarity of the words, also for different age groups. The words were chosen from everyday spoken language and two word-frequency dictionaries were used for reference. The Frequency dictionary of Finnish was based on material collected by Saukkonen (1979) in the 1960's, so that some of the words were already old-fashioned. The Frequency dictionary of Finnish collected in 2004 by the IT Center for Science (Helsinki, Finland), consisting of the 9996 most frequent words appearing in newspapers was used as the main reference. Phonetic balancing was made by comparing the words with the phoneme frequencies in Finnish texts and speech published by Vainio (1996).

During the selection of the speech material, an over-representation of the long phonemes /ɑ:/ and /ä:/ was detected, especially in the third person singular in verbs' present tense and in objective nouns' partitive case. To overcome this problem, half of the verbs had to be changed into the past tense and half of the two-syllable adjectives and nouns had to be substituted with three-syllable words. The distribution of the two- and three-syllable words was even in names, verbs and adjectives. To reduce the complexity of the matrix, one numeral was rejected and substituted with "a pair of". In the end, no more than 4 suitable three-syllable numerals were found. Therefore in the numerals, there were 4 three-syllable and 6 two-syllable numerals and in the objective nouns, there were 6 three-syllable and 4 two-syllable nouns used for balancing the overall distribution of the syllables.

The base matrix chosen according to the aforementioned principles is shown in Table 3. The phoneme distribution of the speech material in comparison to the reference distribution is shown in Figure 3 a,b.

Table 3: The Finnish speech material with the 50-word matrix consists of ten sentences with the same syntactical structure. The words in bold are randomly selected forming one of the test sentences.

Name	Verb		Numeral		Adjective		Noun	
Elina	etsii	(searches)	pari	(a pair of)	halpaa	(cheap)	autoa	(cars)
Harri	huomasi	(notices)	kaksi	(two)	kallista	(expensive)	bussia	(busses)
Johanna	järjesti	(arranged)	**kolme**	(three)	keltaista	(yellow)	kelloa	(watches)
Kerttu	lainasi	(borrowed)	neljä	(four)	pientä	(small)	kenkää	(shoes)
Mikko	näkee	(sees)	viisi	(five)	**punaista**	(red)	kirjaa	(books)
Juhani	ostaa	(buys)	kuusi	(six)	sinistä	(blue)	kuppia	(cups)
Olga	**pyysi**	(asked)	seitsemän	(seven)	suurta	(big)	mattoa	(carpets)
Petteri	tahtoo	(wants)	kahdeksan	(eight)	tuttua	(familiar)	pöytää	(tables)
Sofia	tarvitsi	(needed)	yhdeksän	(nine)	uutta	(new)	rengasta	(wheels)
Ville	valitsee	(chooses)	kymmenen	(ten)	vanhaa	(old)	**sukkaa**	(socks)

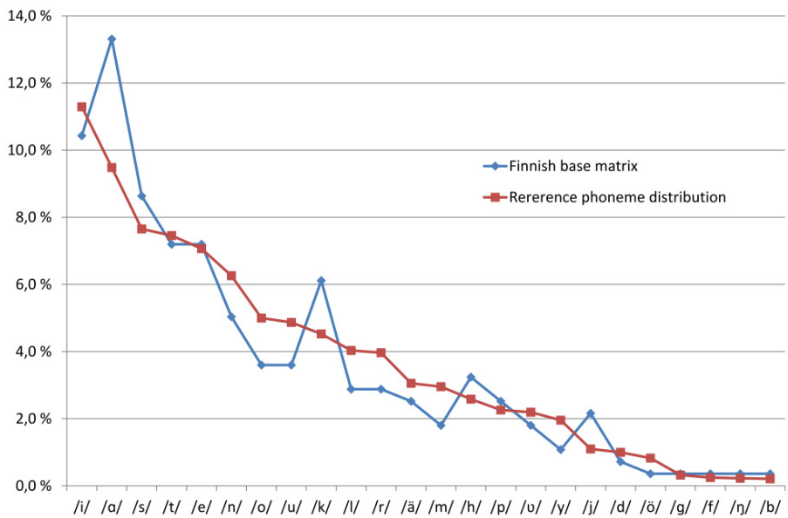


Figure 3: a The distribution of long phonemes in the base matrix in comparison to the reference phoneme distribution.

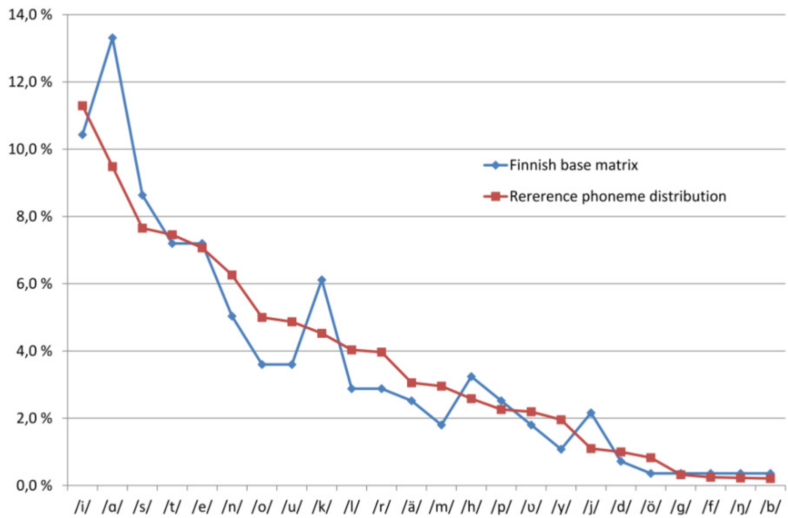


Figure 3 b: The distribution of short phonemes in the base matrix in comparison to the reference phoneme distribution.

2.2.2 Recordings

The speech material was recorded in a sound insulated room in the "House of Hearing" in Oldenburg, Germany, using a Neumann KM184 cardiod microphone (Georg Neumann GmbH, Berlin, Germany) and a standard Windows PC with a USB Soundcard RME Fireface UC (Distribution Audio AG, Haimhausen, Germany) and Adobe Audition software Version 2.0 (Adobe Systems Software Ireland Limited, Dublin, Ireland) with a sampling rate of 44.1 kHz and 32 bit resolution. The recording room had reverberation times (T_{30}) of less than 0.5 s for all frequencies between 125 Hz and 8 kHz. The signal-to-noise ratio of the final recordings was better than 40 dB. The speaker was instructed to keep the same speech level and distance from the microphone during the recordings. The speaker was asked to use natural speech effort, speech rate and intonation for all sentences and to enunciate all words clearly. The compliance with these instructions was continuously verified by four listeners outside the recording booth. Mrs. Marjukka Havumäki, a trained female speaker with standard Finnish pronunciation, who is a news anchor for Finland's national public service broadcasting company, YLE, spoke the sentences. In contrast to the original Hagerman test, in which only the base matrix was recorded, the procedure for the Oldenburg

Sentence Test accounts for co-articulation between the words *(Wagener et al., 1999a–c)*. Accordingly 100 sentences, including all possible combinations of two consecutive words, were recorded at least two times. The concept of preserving co-articulation produces more natural sounding sentences. All recorded sentences were reviewed with respect to speech rate, loudness, intonation, artifacts and clarity in order to select the 100 best sentences for further processing. The post processing of the recorded materials also included high pass filtering at 50 Hz in order to remove any potential low frequency humming sounds. All sentences were equalized in terms of their RMS (root mean square) level in order to adjust for potential loudness differences of the speaker during the recording session.

2.2.3 Test lists used for the Finnish matrix sentence test

The test lists used for the Finnish matrix sentence test derived from 30 generic ten-item test lists (i.e. 30 base lists containing 10 sentences each), which fulfilled the following requirements: a) they seem purely random b) each word transition appears equally often in all sentences of the 30 base lists, c) each word occurs exactly once per 10-item-list, d) all 10 sentence lists can be combined freely with each other. Instead of the theoretically possible 10^5 sentences, the 30 base test lists contain 300 sentences. Due to the seemingly random nature of the sentences, the test subjects cannot memorize them.

2.2.4 Cutting the speech material and resynthesizing the sentences

In order to generate the sentences from the base test lists, the 100 recorded and selected sentences were cut into individual words. For each sentence, four cutting points for each word transition were established. The cutting points were chosen such that the co-articulation between two words was included in the first word sound file. The individual words were then edited to remove recording artifacts if necessary. Finally the words were concatenated into new sentences according to the 30 base test lists. In the concatenation, two consecutive sound files were placed together with individually optimized cross-fading parameters in order to obtain transitions as natural as possible. These cross-fading parameters were stored for each transition in order to be able to re-generate the sentences in the later processing steps. The newly synthesized sentences were all reviewed by five native speakers.

2.2.5 *Development of the masking noise*

The spectral and temporal properties of the speech signal and the background noise affect the results of a speech intelligibility test in noise. In order to achieve optimal spectral masking and thus a steep discrimination function, the background noise should exhibit the same long-term average spectrum as the speech signal *(Wagener et al., 1999a–c; Theunissen et al., 2009; Wagener & Brandt, 2005)*. This eliminates accidental differences between the speech signal and the noise and makes it possible to have a steep intelligibility function in the test. The noise should not include any intelligible words in order to prevent unwanted informational masking. The use of spectrally matched noise is also well documented with sentence material *(Wagener et al., 1999; Hochmuth et al., 2012; Jansen et al., 2012; Wagener et al., 2003; Wagener & Brandt, 2009; Dreschler et al., 2001)*. The masking noise was therefore generated from the recorded speech material by 30-fold superposition of all individual sentences, creating a stationary noise without fluctuations. The superposition of sentences was done as described by Wagener et al. (1999, 2003).

2.2.6 *Optimization measurements*

The purpose of the optimization procedure was to balance the speech intelligibility across the individual words of all sentences. This is necessary in order to achieve a steep discrimination function for the test. It was shown that the speech recognition curve of a sentence test can be predicted by the convolution of the mean word-specific recognition curve and the distribution of the word-specific SRT values of all single words. Accordingly, the word-specific recognition function was initially determined. Based on this data, level corrections of ±3 dB were subsequently made in order to equalize the intelligibilities of all words as far as possible. The theoretical background for this procedure has been described in earlier publications by Wagener et al. (1999, 2003) and Hochmuth et al. (2012).

The optimization measurements were performed in the Kuopio University Hospital with 21 test subjects, aged from 22 – 44 years (mean 30 years). The test subjects had normal hearing confirmed by pure tone audiometry at the beginning of the session (pure tone threshold < 15 dB HL for octave frequencies 125 Hz-8 kHz). All measurements were performed monaurally on the better ear. All test subjects had been born in Finland and were Finnish native speakers. Measurements were done in a sound-attenuated booth using sound-field equalized Sennheiser HDA200 headphones (Sennheiser Electronics

GmbH & Co KG, Wedemark-Wennebostel, Germany) and a Fujitsu Lifebook (Fujitsu-Siemens Computers GmbH, München, Germany) with the onboard sound card Realtek HD Audio (Realtek Semiconductor Corp., Taiwan, China) and the Oldenburg Measurement Applications software (OMA, Version 1.3, HörTech gGmbH, Oldenburg, Germany). A calibration of the equipment was performed before the commencement of the optimization measurements and confirmed after their completion. For the optimization measurements, the 30 ten-item test lists were rearranged into 10 thirty-item test lists. Each subject was tested with all 10 test lists of 30 sentences at ten different SNR values (from -14 dB to -5 dB) with a constant noise level of 65 dB SPL in order to determine the intelligibility function of all the recorded words with the order of test lists and SNRs being randomized. Two training lists were measured prior to the actual measurements at 0 dB SNR and -4 dB SNR to familiarize the test subject with the test concept and the speech material. The masking noise started 500 ms before the sentences and ended 500 ms after the sentences. The test subject repeated the understood words (i.e. open set presentation) and the audiometrist recorded the correctly repeated words (word scoring). Analysis of the optimization measurements resulted in the necessity to perform further measurements at additional 5 SNR levels (-2 dB, -4 dB, -16 dB, -18 dB and -20 dB). Overall, 18 test subjects were tested at 15 SNR levels and 3 test subjects were tested at 10 SNR levels.

2.2.7 Evaluation measurements

The aim of the evaluation measurements was to make sure that the test lists obtained after the optimization would be equivalent with respect to speech intelligibility. The evaluation measurements also provided reference values for further clinical applications. The optimization resulted in 14 ten-item lists (see below). These lists were combined to seven 20-item lists. A total of 21 newly recruited native Finnish speaking test subjects, aged 21-38 years (mean 23 years) were measured at the Kuopio University Hospital with the same equipment and set-up as used for the optimization measurements. They were all normal hearing (pure tone threshold < 15 dB HL for octave frequencies 125 Hz-8 kHz). To investigate the training effect, seven 20 sentence lists were measured adaptively prior to the actual measurements. The data from the training lists was also used for assessing the training effect for the Finnish Matrix Test. The subsequent evaluation measurements were performed at constant SNRs (-12.5 dB, -10.5 dB, -8.5 dB). The SNRs were selected to reach approximately 20 %, 50 % and 80 % word recognition rates. The order of the test lists and the SNRs were randomized and the measurements were performed at a constant noise level of 65 dB SPL.

2.3 Results

2.3.1 Results of the optimization measurements

From the data of the optimization measurements, the psychometric function for each individual word realization was determined. The psychometric function was derived by fitting a sigmoidal function to the combined raw data of all test subjects for each individual word realization, yielding the SRT and slope of the function for each word. The function used was the logistic function:

$$SR(l) = \frac{100}{1 + e^{-4m(l - SRT)}},$$

where SR is the speech recognition rate in percentages, l is the level in dB SNR, SRT is the speech-reception threshold in dB SNR, and m is the slope of the psychometric function at the SRT. By using this approach it was possible to obtain a mean word-specific SRT of -10.4 ± 2.3 dB (SNR) and a mean slope of the intelligibility function of 18.9 ± 7.1 %/dB across 500 word realizations. Figure 4 shows an example of the measured SRTs of the numerals before the optimization procedure. The slopes of the psychometric functions did not differ extensively between the word positions within a sentence. The SRTs of the adjectives were somewhat higher than the SRTs of the other word positions (see Figure 5).

In the next step, each word had to be adjusted in its level in order to bring its SRT as closely as possible to the average SRT of the whole speech material. The amount of this level correction per word needed to be limited in order to prevent unnatural sounding sentences. The maximum level correction was set to ±3 dB in order to preserve natural sounding sentences (see also Hochmuth et al., 2012). The level adjustment to each word was determined based on the difference between the word's SRT and the average SRT of the whole speech material. Reasonable optimizations could not be obtained for all 500 word realizations. This could be due to one of the following three reasons: (1) Even after the maximum level correction of ±3 dB, the word's SRT was still more than 2 dB away from the target SRT. (2) The slope of the word's psychometric function was less than 5%/dB. (3) Based on the available data from the optimization measurements, a reliable estimate of SRT and slope could not be obtained for the word realization. There were 15 word realizations that fulfilled the first criterion. The two other criteria were not met by any of the words.

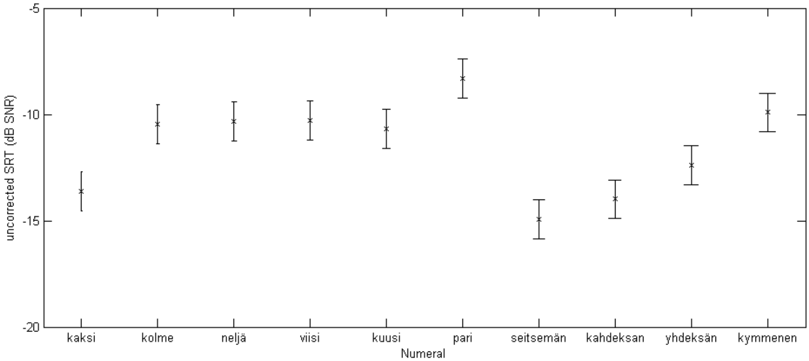


Figure 4: The mean SRTs +/- SD of the numerals in the base matrix before optimization.

Any ten-item test list that contained two or more words that could not be optimized was discarded from the total pool of test lists. After this deletion, 14 test lists from the original 30 ten-item test lists remained for further evaluation. These 14 ten-item test lists were combined to 7 test lists each with 20 sentences. After the level adjustments, 15 sample sentences with especially conspicuous differences in the loudness (theoretically max. 6 dB between two consecutive words) of the individual words, were presented to three native listeners and reviewed for natural sound. However, no subjective loudness differences were noticed on three presentations, until they were pointed out.

Based on the level corrections, an average SRT of -10.4 ± 0.6 dB SNR could be expected for the whole speech material. The expected SRTs and standard deviations for the different word positions are indicated in Figure 5.

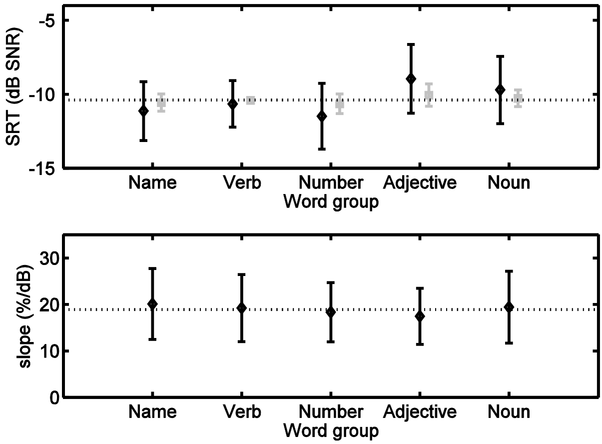


Figure 5: SRT (top panel) and slope (bottom panel) of the speech material before (black diamonds) and after (gray squares) applying the level adjustments. Error bars indicate standard deviations. The dotted lines mark the mean across all word realizations. The SRT values for the speech material after the level adjustments are expected from the optimization measurements but have not been measured.

2.3.2 Training effect of the Finnish Matrix Sentence Test

The training effect was studied on the basis of the seven adaptive measurements prior to the evaluation measurements. On average, the test subjects improved by 2.3 dB SNR from the first to the seventh measurement. The largest improvement was seen between the first and second measurement (mean 1.1 dB SNR). The difference between the first and the third measurement was 1.8 dB. The difference between the third and seventh measurement was only 0.5 dB. This means that the major contribution to the training effect had occurred between the first and the second measurements. From the third measurement onwards, the performance of the test subjects improved only marginally.

Pairwise non-parametric Friedman tests on the SRTs from the consecutive adaptive measurements were performed in order to detect differences between consecutive measurements. Assuming a significance level of 5%, significant differences between the first and the second and between the second and the third measurement were found. No significant differences were found between the following consecutive measurements. An additional one-way repeated measures analysis of variance (ANOVA) followed by a multiple comparison procedure according to Holm-Sidak yielded the same results.

If one pools the adaptively measured SRTs from the third measurement onwards, one obtains an expected range of -9.7 ± 0.7 dB SNR for normal hearing test subjects in the adaptive measurements (mean and standard deviation). The data from the training sessions is displayed in Figure 6.

2.3.3 Results of the evaluation measurements

The objective of the evaluation measurements at fixed SNRs was twofold. First, the list specific speech recognition functions were determined. This was done by fitting the logistic model function (Equation 1) to the results of all subjects for each test list *(Brandt & Kollmeier, 2002)*. Secondly, the inter-individual differences between test subjects in terms of SRT and slope were determined by pooling the data from all test lists and fitting the logistic model function to the data for each test subject. The mean SRT and the slope of the 14 ten-item test lists was -10.1 ± 0.1 dB SNR and 16.7 ± 1.2 %/dB, respectively. The values of the mean SRT and the slope of the test subjects were -10.1 ± 0.7 dB SNR and 17.5 ± 2.2 %/dB, respectively. The final intelligibility function of the Finnish matrix sentence test is shown in Figure 7.

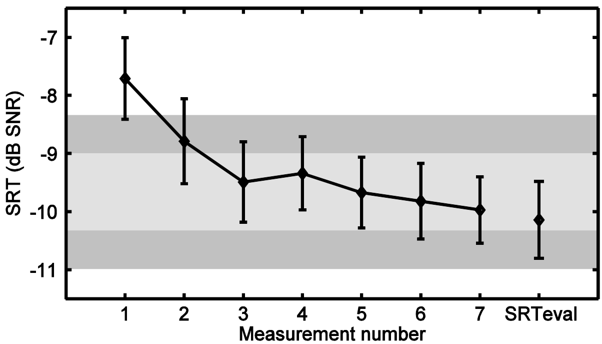


Figure 6: Improvement of test subjects during the training sessions for the evaluation measurements. SRTs were determined adaptively. Error bars indicate standard deviations across test subjects. The last error bar indicates the mean SRT and standard deviation across test subjects as determined from the evaluation measurements at fixed SNRs. The light and dark gray areas show the regions ±1 and ±2 standard deviations around the mean SRT when pooling all adaptive training measurements from the third measurement onwards.

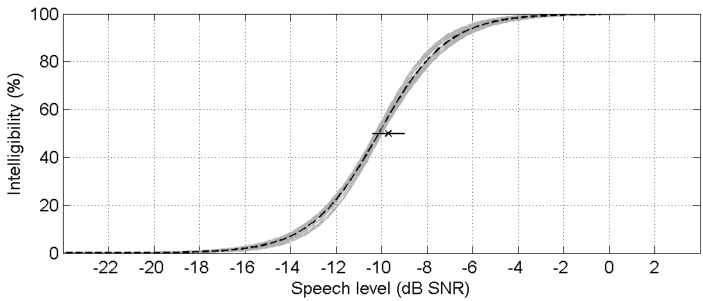


Figure 7: The final speech intelligibility function of the Finnish matrix sentence test. The dashed black line represents the mean intelligibility function of the test. The grey lines represent the intelligibility functions for each of the 14 final test lists. The diagonal cross with the short line represent the estimated SRT (± 1 standard deviation) on adaptive measurements.

The small standard deviation of SRTs between test lists compared to the standard deviation between test subjects indicates that the test lists are interchangeable. For an additional assessment of possible differences between test lists, three one-way repeated measures ANOVAs were performed on the intelligibilities at the three fixed SNRs. No significant differences were found between the different test lists for -12.5 dB SNR and for -10.5 dB SNR ($F_{13,277}$=1.641, p=0.075; $F_{13,269}$=1.324, p=0.199, respectively). For -8.5 dB SNR, the one-way repeated measures ANOVA reported a significant difference between test lists ($F_{13,271}$=1.893, p=0.032) which could not be verified with a post-hoc multiple comparison procedure according to Holm-Sidak.

2.4 Discussion

The Finnish Matrix Test is the first sentence test in noise for the Finnish language. It was developed according to the same principles as the already existing tests of same structure for Swedish, German, Danish, French and Spanish. The Finnish Matrix Sentence Test matches very well with these tests with respect to list-specific recognition function and intelligibility across the subjects. This can be explained by the predefined methodological standards used for the development of these new speech tests. For the first time, comparable studies across the languages are now possible. In the clinical setting, it also facilitates the standardization of therapy indications (e.g. hearing rehabilitation) across Europe, which will be of considerable importance when the free movement of patients will become a reality across the European Union.

With the exception of the Swedish Matrix test devised by Hagerman, in all subsequent tests co-articulation has been accounted for during the recording sessions. This has meant that the re-synthesized sentences sound much more natural, thus eliminating one of the main criticisms raised against the original Hagerman test *(Hagerman, 1982; Wagener et al, 1999a-c).*

During the optimization of the speech material for matrix tests, adjustments up to ± 4 dB SPL were normally used. The amount of adjustment was determined up to the limit without impairing the naturalness of the spoken sentences. In the present study, a maximum level adjustment of ± 3 dB SPL was used. Despite the level adjustments some representations of words could not be adequately optimized. Therefore 140 sentences of the originally re-synthesized 300 sentences are represented in the final test version. Nevertheless, the remaining 14 10-item test lists could be demonstrated to be highly equivalent, which contributes to the validity of the test. There was a standard deviation between test lists of 0.1 dB SNR, which is even less than that reported for the German Sentence Test OLSA (0.16 dB SNR) or the original Hagerman test (0.3 dB SNR). In the final Finnish Matrix Test, the 20-item test lists were created by combining any of the 10-item lists. The seemingly random nature of the sentences makes them impossible to be memorized by the test subjects.

The rate of speaking of the Finnish Matrix Test is very homogenous for the complete speech material. With 226 ± 19 syllables per minute, the syllable frequency of the final Finnish sentences is very comparable to the syllable frequency of the OLSA which is 233 ± 27 syllables per minute *(Wagener et al., 1999a-c)*. This speech rate should be sufficiently slow to make the test suitable for measurements with severely hearing impaired patients and cochlear implant users. However, it is still fast enough so that the sentences sound natural.

As is the case with the other international tests, the most important training effect takes place between the first and the second test list. In the present study, the standard deviation of adaptively measured SRTs after the second 20-item test list was 0.7 dB SNR. Assuming normally distributed test results (SRTs) in adaptive measurements for normal hearing test subjects, the 95% confidence interval for the test result was in a ± 1.4 dB range around the average SRT. It is therefore advisable to conduct at least two training lists prior to the actual measurement (Figure 6).

The optimization of the speech material is conducted in order to achieve a steep slope of the test-specific recognition curve. In the current study, the recognition scores of all word groups (i.e. at each position in the sentence) were found to be very close to each other. In the Spanish Matrix Test, the names were easier to understand than the other word groups even after optimization *(Hochmuth et al., 2012)*. Figure 5 shows the SRTs for each word group before and after optimization, as well as the slope of the intelligibility function for each word group.

With the final slope of 16.7 %/dB, a steep intelligibility function was achieved which is very close to that of the OLSA (17.1 dB/%) and higher than the slope for the Danish test (12.6 %/dB), the Spanish test (13.2 %/dB) or the French test (14.0 %/dB) *(Wagener et al., 1999a-c; Wagener et al., 2003; Hochmuth et al., 2012; Jansen et al., 2012)*. For further comparison with existing international Matrix Tests see Table 4.

Table 4: The comparison of key specifications of the existing international Matrix Tests with the Finnish Matrix Test. If available, values are given as mean ± standard deviation. The variability of SNRs is given across subjects (i.e. test lists are pooled)

	SRT (adaptive)	**SRT (constant SNR)**	**Slope**	**Reference**
German (OLSA)	-6.2 ± 1.3 dB SNR	-7.1 ± n/a dB SNR	17.1 ± 1.6 %/dB	Wagener 1999, Wagener & Brand 2005
Danish	n/a	-8.4 ± 1.0 dB SNR	12.6 ± 0.8 %/dB	Wagener 2003
French	n/a	-6.0 ± 0.6 dB SNR	14.0 ± 1.6 %/dB	Jansen 2012
Spanish	-6.2 ± 0.8 dB SNR	-6.8 ± 0.8 dB SNR	13.2 ± n/a %/dB	Hochmuth 2012
Finnish	-9.7 ± 0.7 dB SNR	-10.1 ± 0.7 dB SNR	16.7 ± 1.2 %/dB	current publication

The most relevant information for the Finnish Matrix Sentence Test is represented by the mean data across the subjects. The expected SRT for normal hearing young adults for adaptive measurements is -9.7 ± 0.7 dB SNR after two 20-item training lists. The evaluation measurements (performed at fixed SNRs) corresponded very well to the adaptively measured SRTs, with an SRT of -10.1 dB ± 0.7 dB SNR. The Finnish Matrix Test shows higher intelligibilities compared to the other tests at comparable SNRs (Table 4). Due to the fixed methodical standards during the development of these tests, differences in measured SRTs and recognition functions can most likely be explained by language specific reasons and by speaker dependent factors.

The Finnish language is phonologically rather straightforward with distinct articulatory and acoustical characteristics. For example, its vowel structure may be more intelligible than is the case for other languages since there are only minor phonological differences between most of the long and short vowels *(Ylinen et al., 2005)*. In the Finnish language, front and back vowels never occur together in a two-syllable word (vowel harmony). The vowel in the first syllable determines the type of vowel in subsequent syllables. This means that the occurrence of vowels in a word is predictable, which is especially beneficial when there is background noise, when only a fraction of the word may be understood. It was shown that this phonetic rule is learned by Finnish children very early, before the age of 3 years and any violation to this vowel harmony has been shown to elicit so-called mismatch negativity responses in the brain *(Aaltonen et al, 2008)*. Furthermore there is some evidence that the linguistic entropy of speech stimuli (i.e. information content of linguistic stimuli) has an effect on the SRT, which was up to 4 dB (SRT) in young adults *(van Rooij & Plomp, 1991; Versfeld et al., 1999)*. The same principle can be applied to phonetic rules, such as the rather strict vowel harmony present in Finnish.

Differences in SRT across languages can also be attributed to speaker-dependent factors *(Theunissen et al., 2009; Versfeld et al., 1999).* In contrast to the matrix tests in other languages, here a highly trained professional female speaker was used; in fact a news anchor on the Finland's national public service broadcasting company, YLE. Especially in unfavorable hearing conditions, there is a substantial benefit of clear and articulate pronunciation to the speech intelligibility *(Mullennix et al.,1989; Smiljanić & Bradlow, 2005)*. Therefore it is very likely that her very clear and articulate speech has contributed to the lower SRTs of the Finnish test version. Speaker differences were previously studied. Versfeld et al. (1999) found that the SRTs of the same speech material for four different speakers (2 male and 2 female speakers) ranged from -2.6 to -1.1 dB, which was statistically significant. It remains unclear the extent to which either the speaker or the language dependent factors have contributed to the better SRTs found with the Finnish test.

2.5 Conclusions

The Finnish Matrix Test in noise is the first evaluated sentence test in noise for the Finnish language. The development of the test was conducted according to the principles implemented in the EU projects HearCom and HurDig in order to establish comparable speech audiometric tests across Europe *(Zokoll et al., 2013).* The new Finnish Matrix Test provides reliable test results with the same characteristics as the other equivalent European tests.

3 Application study of the new Finnish matrix test in cochlear implant recipients

3.1 Introduction

During the last decades, cochlear implants (CIs) have become standard treatment for both children and adults with severe to profound bilateral hearing loss. The quality of the rehabilitation with CIs depends on both adequate surgical care as well as meticulous postoperative follow-up. Regular check-ups need to be carried out in order to evaluate the progress of the patient's rehabilitation. These allow for therapy adjustments to the needs of each individual patient (i.e. cochlear implant fitting and speech therapy). In order to adequately follow-up and to quantify the rehabilitation results, different types of speech audiometry have been applied. In the literature, short words (i.e. monosyllables, CVC, etc.) presented in silence are often still used for the presentation of rehabilitation results with CIs (Gifford et al., 2008). These tests, however, are not sensitive enough to permit an adequate evaluation of the rehabilitation results with CIs (Gifford et al., 2008; Müller-Deile, 2009, Hey et al., 2014). The short duration of the stimulus is often not sufficient to allow the adaptation of the sound coding algorithms to the sound and the test results do not correlate well with the functional hearing of the patients (Nilsson et al., 1994; Wagener et al, 1999; Müller-Deile 2010). Since the results are reported as the percentage of the correctly identified words, these tests are often susceptible to a saturation effect (Gifford et al., 2008). In CI rehabilitation applications, frequent testing is usually required in order to ensure the optimal fitting of the devices. Therefore the test variability has to be as low as possible in order to be able to choose proper and accurate programming parameters.

Speech audiometry in noise reflects everyday listening situations much better and helps to assess the functional auditory communication ability of the patients. In speech tests in noise, the speech reception threshold (SRT) is usually measured adaptively. The SRT represents the signal to noise ratio at which the test subject repeats 50 % of the test items correctly. This type of measurement is not susceptible to the saturation effect often observed in speech audiometry in quiet performed at constant suprathreshold levels. Speech in noise tests, especially with sentences as speech material, normally also exhibit much steeper intelligibility functions, which facilitates the accurate measurement of the SRT with lower test-retest variability (Brandt et Kollmeier, 2002; Müller-Deile, 2009; Bosman et Smoorenburg, 1995).

Several sentence tests in noise have been developed for the German language. For example, the Göttingen Sentence Test (GÖSA, Kollmeier & Wesselkamp, 1997) consists of meaningful everyday sentences. This test, however, may be sometimes too difficult for CI patients due to the rather fast speech rate and the use of colloquial speech and pronunciation (mean speech rate: 279 +/- 38 syllables/min); (Kollmeier et Wesselkamp, 1997; Müller-Deile, 2009). The so-called HSM-test was recorded with a slower speech rate (mean 222 +/- 40 syllables/min) and utilizes clearly enunciated sentences and this test has been found to be more suitable for conducting measurements in CI users (Hochmair-Desoyer et al., 1997, Müller-Deile, 2010). There are however differences in the intelligibility across the different test lists. Furthermore the familiarization with the speech material in tests with meaningful sentences limits their application in settings where repetitive testing is required such as in CI users (Hey et al., 2003; Müller-Deile, 2010). A minimum of one year is commonly recommended before repeating the same test lists, thus reducing the applicability of this test in repeated measurements.

The Oldenburg Sentence Test (OLSA), developed by Wagener et al. (1999) has proved to be suitable for the measurements with CI recipients (Hey et al., 2003, 2010; Rader, 2008; Müller et al., Müller-Deile, 2009; Fredelake et Hohmann, 2012). The moderate speech rate of 233 ± 27 syllables/min is appropriate for patients with hearing impairments as well as for CI users. The OLSA test lists were optimized to achieve equal intelligibility by using normal hearing subjects. The evaluation with 20 normal hearing test subjects resulted in a mean SRT of -7.1 +/-0.16 dB (SNR) and a slope of the intelligibility function of 17.1 +/- 1.65 %/dB (see also Table 4). The steep slope of the intelligibility function allows for very accurate SRT determination with only marginal test-retest variability.

Hearing impaired patients and especially CI users show a very wide performance range in their speech perception. It has been shown in several studies that there is quite a large variety in the SRTs in CI users both in quiet and noise (Fredelake & Hohmann, 2012; Hey et al., 2010, Müller-Deile 2009). For a subset of CI patients, with lower than average performance, the OLSA may sometimes be too difficult to allow accurate SRT measurements (Hey et al., 2010). The characteristics of the OLSA were examined with respect to the SRTs, slope of the intelligibility function and test-retest variability in CI users by Müller-Deile (2009) and Hey et al. (2010). In these studies, the OLSA was found to be well suited for advanced and repetitive testing in CI patients. For many studies in which repetitive measurements are required (i.e. speech processor or speech coding strategies research, binaural hearing research etc.), the OLSA is chosen for the determination of the SRT. Nowadays in clinical as well as in research applications, the OLSA is widely accepted and is state-of-the-art for (adaptive) speech-in-noise testing for assessing the performance capacity in everyday hearing situations as well as for monitoring changes in maximum performance for documentary reasons of CI patients.

In Finland, the only available speech test is a word test of meaningful two-syllabled words presented in quiet (Jauhiainen, 1974). Seven test lists of 25 words are available which were not tested for their perceptive equivalence. Therefore the seven different test lists exhibit unequal difficulties (see also Table 2). The speech material was devised in 1952 and therefore includes uncommon and old-fashioned words that are no longer used anymore in the normal Finnish colloquial language. The phonemic balancing was disturbed when there was a revision of the lists done in the late 1960s, where five of the least intelligible words were omitted (Jauhiainen, 1974). This test is suitable for measurements at the threshold level (i.e. for the functional check-up of the sensitivity adjustment of the speech processor and microphone characteristics), however, for suprathreshold testing, it appears to evaluate the vigilance of the patients rather than their functional hearing. On the other hand, a substantial portion of the Finnish CI users already achieve 100% intelligibility after a rather short time of device usage. The extent to which extent the 100% scoring is due to the high redundancy of the test or due to the real rehabilitation success cannot be determined. From the clinical perspective, the Finnish two-syllable word test in quiet is not applicable for the testing of CI patients.

The aim of the study was to investigate

1. whether the new Finnish matrix sentence test in noise would be suitable for measurements of CI patients.
2. if the Finnish matrix sentence test would correlate with the established Finnish word test in quiet
   a) in patients, presenting for CI candidacy.
   b) in patients, who have received a CI in the Kuopio University Hospital.

3. the intelligibility function of the Finnish matrix sentence test in a cohort of 13 experienced CI patients.
4. the extent to which there would be comparability between the rehabilitation results of CI patients measured with the OLSA and the Finnish matrix sentence test.

3.2 Material and Methods

The new Finnish matrix sentence test was implemented at the Hearing Center of the Kuopio University Hospital in March 2014. After the implementation, all CI patients with at least one year of implant use, who came for a regular follow-up until the beginning of June 2014 were measured with the matrix test and the Finnish word test in quiet. A total of 33 unselected CI patients, with severe pre- or post-lingual hearing impairment, were enrolled into the study. Eighteen patients were unilateral CI users, 7 patients had bilateral CIs and 8 patients used bimodal stimulation. All patients were measured in the best-aided condition. In 4 patients with bilateral CIs additional measurements were performed unilaterally on each ear. Three bilateral CI users with a very long interval between the first and the second CI had not developed any speech understanding in the second ear and were measured only with bilateral stimulation. Patients using bimodal stimulation were measured also in the CI only and hearing aid only condition.

In order to facilitate the necessary training for this new test, patients were introduced to the principles of the matrix-type test and the speech material was shortly reviewed to them by showing the matrix in its written form. All measurements were done with randomized 20-item test lists with fixed noise level of 65 dB (SPL). The first training list was presented with a fixed SNR of +10dB, so that most of the words were understood. This was done in order to potentially speed up the training for the test. Subsequently, two adaptive test lists were conducted, the first for additional training and the second for the actual SRT measurement.

In a subset of 13 patients, with at least two years of CI experience and word scoring over 88 %, three additional test lists were measured with the algorithm described by Brand and Kollmeier (2002) for concurrent slope (of the intelligibility function) and SRT estimation. The SRT and slope for each patient was calculated from the three measurements. The results represent the mean SRT and the mean slope of the inter-individual measurements.

Sixteen patients referred to the Hearing Center for the evaluation of CI rehabilitation were measured in the best aided condition.

All measurements were made in a sound room in the S_0N_0 condition, in which the signal and noise are presented from 0^0 azimuth from the same loudspeaker. The Oldenburg Measurement Applications software (OMA, Version 1.3, HörTech gGmbH, Oldenburg, Germany) was used in conjunction with a HP Probook (HP, Palo Alto, California, United States) and an Auritec mobile audiometry device ear 3.0 (Auritec GmbH, Hamburg, Germany) which directly supplied a Genelec 8140A active loudspeaker (Genelec Oy, Iisalmi, Finland). Calibration of the equipment was performed before the commencement of the measurements. The statistical analysis was done using IBM SPSS Statistics Version 19 (IBM Corporation, New York, USA).

3.3 Results

The mean word score and median for the Finnish word test in 33 cochlear implant patients in the best aided condition were 89 % and 96 %, respectively. Mean and median scoring for the first training list of the matrix test presented at +10dB (SNR) were 93% and 95 %, respectively.

The mean SRT on adaptive measurement after two 20 item training lists was -3.60 ± 2.09 dB (SNR). The distribution of the SRTs in the different CI users is shown in Figure 8.

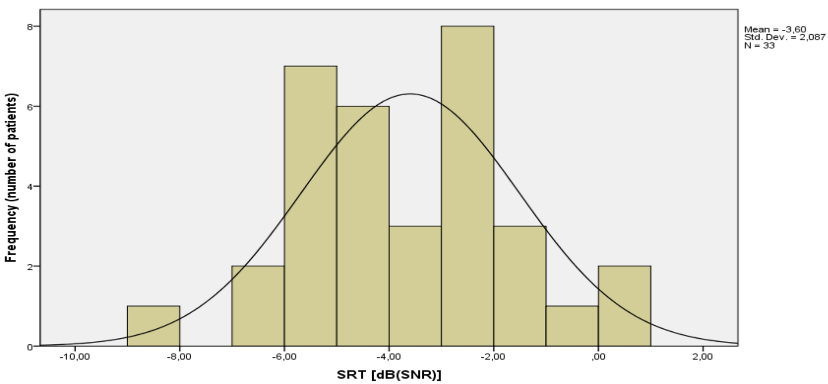


Figure 8: The distribution of the SRTs measured with the Finnish matrix sentence test across the CI users.

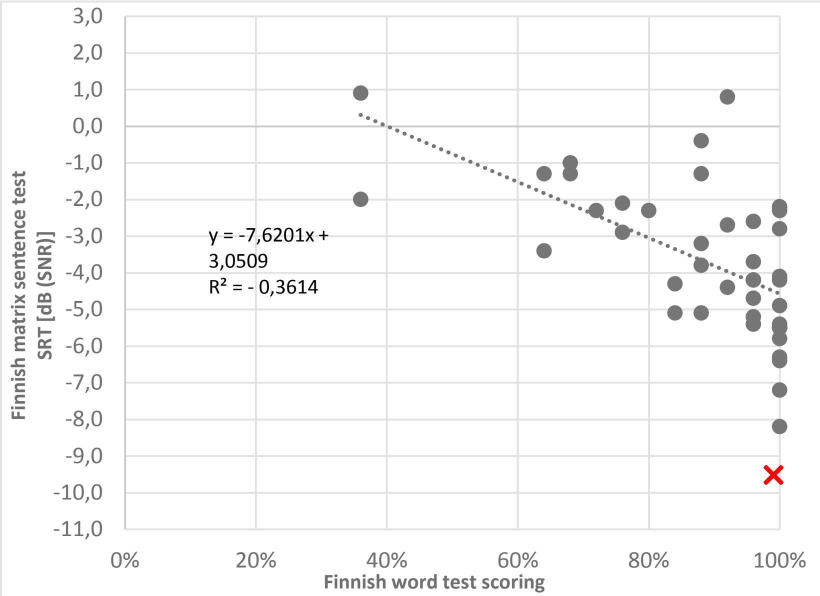


Figure 9: Scatter plot of the SRTs of the Finnish matrix sentence test in noise with the word scoring of the Finnish word test in quiet. The red cross represents the expected SRT for normal hearing subjects.

To evaluate the possible training effect, the mean SRT of the second training list (first adaptive SRT measurement) was measured and was -2.76 ± 2.36 dB (SNR), i.e. approximately 0.8 dB higher than the final measurement result obtained after the training phase had been completed. This learning effect found after the second training list was still statistically significant (paired sample T-test $p<0.001$).

In the calculation of the correlation between the Finnish matrix sentence test in noise and the Finnish word test in quiet, a total of 41 measurements were available (measurements of the 33 patients in the best aided condition and the 8 additional measurements of the 4 bilateral CI users measured on each side alone). There was a weak correlation between the SRTs of the matrix test and the Finnish word test scoring (R^2= -0.3614) as shown in Figure 9. This illustrates well the saturation or ceiling effect of the Finnish word test in a substantial portion of CI patients. Even though the first training of the matrix test administered at a presentation level of +10 dB (SNR) should not be regarded as a valid measurement, the correlation with the SRTs is nevertheless better than the correlation of the Finnish word test and the SRTs (Figure 10).

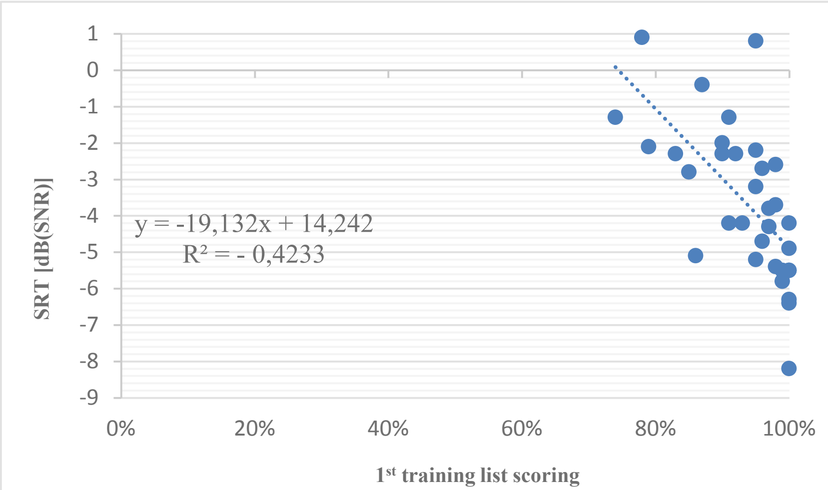
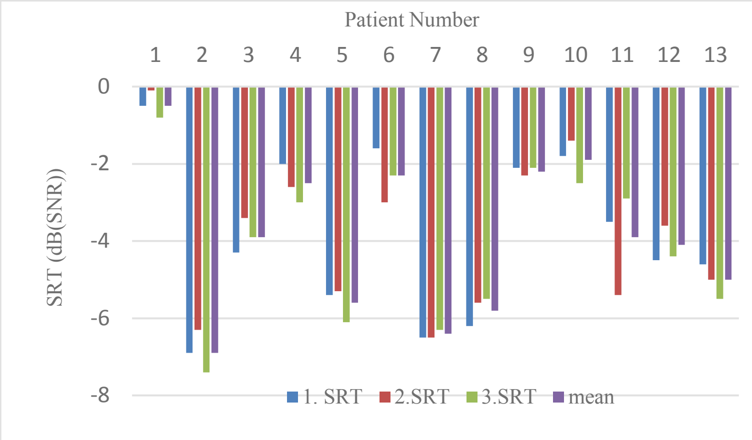


Figure 10: Scatter plot between the SRTs in noise and the first training list presented at a constant SNR of +10 dB in CI patients.

In the subgroup of 13 CI patients with at least two years of usage and a word score of over 90 % the mean SRT was -3.77 ±1.49 dB (SNR) and the mean slope of the intelligibility function was 15.0 ±4.17 %/dB. The results of the 3 consecutive measurements at 20 % and 80 % intelligibility for SRT and slope estimation are illustrated in Figure 11.



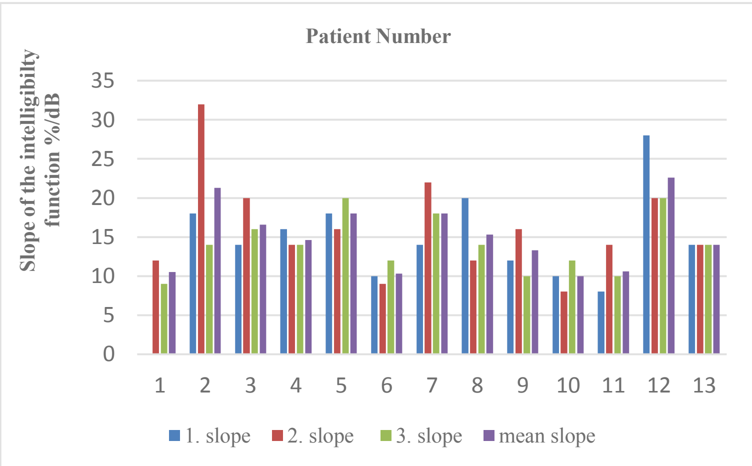


Figure 11: Finnish matrix sentence test SRT and slope estimation for 13 CI users.

The SRT estimation during the slope measurements was -3.92 ±1.88 dB (SNR), which is only slightly lower than for the SRT measurement and the difference was also not statistically significant (paired sample T-Test p=0.42).

In the CI candidates the mean SRT was +1.4 ±4.07 dB(SNR). The mean word score in the Finnish word test was 73% ± 21%. The mean score for the first training list with +10 dB SNR was 82% ± 15%. There was no correlation between the SRT measurements and the Finnish word test scoring (R^2= -0.1009) as shown in Figure 12.

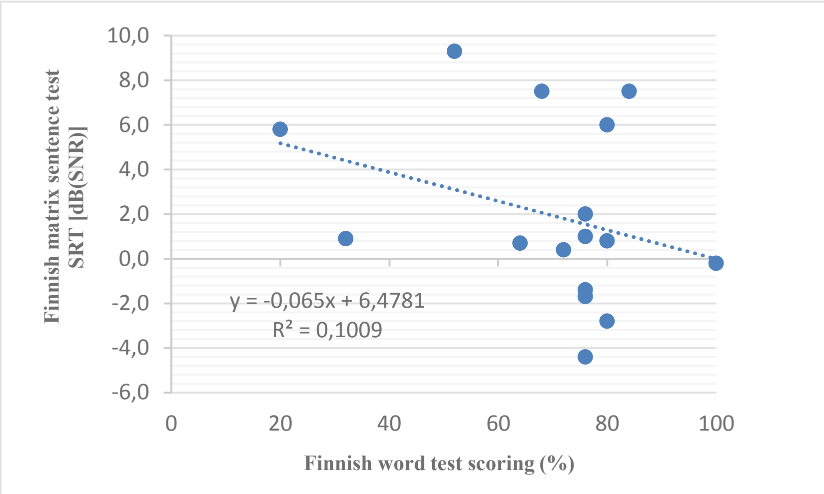


Figure 12: Scatter plot between the SRTs of the Finnish matrix sentence test in noise and the word scoring of the Finnish word test in quiet in 16 CI candidates. The solid line denotes the regression curve denoted by y

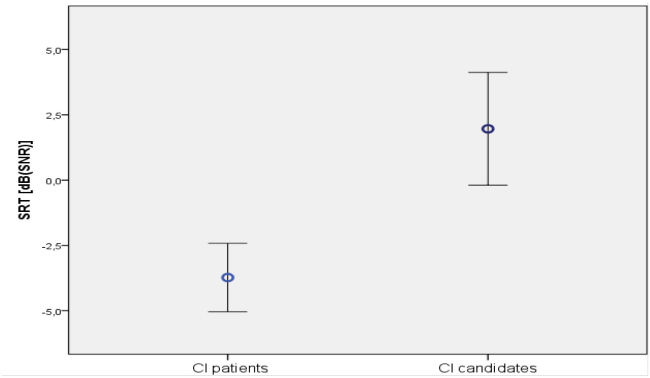


Figure 13: The mean SRT measured with the Finnish matrix sentence test and standard deviation of the CI users and the CI candidates.

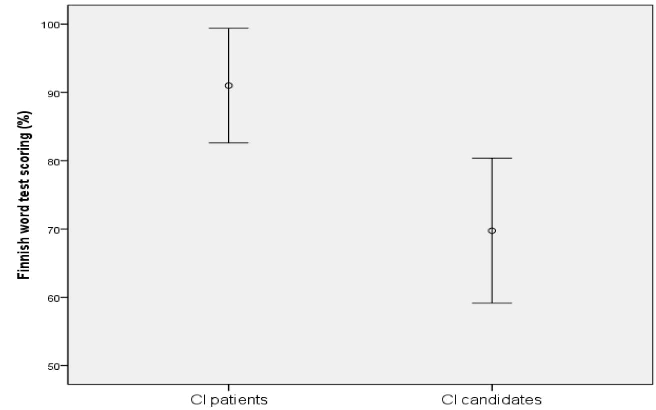


Figure 14: The mean word scoring at the Finnish word test in quiet in CI users and CI candidates.

The mean performance of the CI users was significantly better than the CI candidates in both the speech recognition in noise as well as in the word intelligibility in quiet (Figure 13 and 14). The difference was statistically significant especially for the SRT measurements (independent sample T-test; $p<0.001$).

A test-retest measurement with the matrix test during a separate follow-up appointment (measurement interval > 4 weeks) was possible in only 3 patients and in these individuals the difference ranged from +0.2 to -0.5 dB (SNR) (mean difference 0.4 dB [SNR]).

Bilateral CI treatment improved the performance in 4 patients in the S_0N_0 condition, ranging from -1.4 to -0.9 dB (SNR) (mean SRT improvement -1.08 dB [SNR]).

In three patients with bimodal stimulation, the hearing aid actually impaired their speech reception performance but in two patients a slight improvement was detected. In the remaining patients, no beneficial or detrimental effect of the additional hearing aid could be observed. It is noteworthy that in all bimodal stimulated patients, the SRT on the hearing aid side alone was in the positive range.

3.4 Discussion

Hearing rehabilitation with implantable hearing devices has been very successful in the recent years. Constant improvements in CI system technology and surgical techniques have led to the continuing expansion of indications and this has stimulated the development of reliable and sensitive speech intelligibility tests. This is the background reason why the Finnish matrix sentence test was developed since there was a clear need for an accurate speech intelligibility test procedure in noise for diagnostic, rehabilitation and research applications. During the implementation of new test procedures, it is important to validate the test in the different groups of patients in which it will be used in order to gather knowledge about the test's limitations and possible errors. In this first application study, normative data for the Finnish CI users were gathered and compared with the German CI recipients.

On scheduled follow-up visits to the Hearing Center of the Kuopio University Hospital, all CI patients were measured successfully with the new Finnish matrix sentence test and valid data was collected, i.e., all patients were able to perform the test in an open test format (indicating that the subjects responses were given verbally), and both the non-adaptive training phase and the adaptive measurement phase yielded stable and reproducible measurement results. The same applied also to all of the patients referred to Kuopio University Hospital who were being assessed as candidates for CI treatment. There has been some discussion about the possibility that OLSA may sometimes too difficult for CI users (Hey et al., 2014). This was also a concern during the development of the Finnish matrix sentence test. However, the first experience and measurement results in the application of the new test on CI patients were very promising and this concern about its difficulty could be disregarded. The lower normative SRT of the Finnish matrix sentence test (-10.1 dB [SNR]) in comparison to the OLSA (-7.1 dB [SNR]) seems to facilitate the application of the test in patients with more severe hearing impairment and CI users. Even in those patients who were attending the clinic to be assessed for CI treatment, the matrix test provided valid measurements in all of the cases.

3.4.1 Test-retest reproducibility

The most widely reported disadvantage of matrix-type speech tests is the training effect. For normal hearing subjects, the most significant learning effect happens between the first and the second training list (see Figure 6). Even in patients with prior experience with the matrix-type speech test, it is necessary to perform at least one adaptive training list before the actual measurement (Müller-Deile, 2009). In this study, all the patients performed this test for the very first time. Therefore adequate training was extremely important in order to achieve valid and reliable data. The first list was presented at SNR +10 dB, which resulted in a mean speech level of 75 dB (SPL). The aim of the first test condition was to introduce the patient to this new test and its speech material in order to attain the best possible training. Due to the rather advantageous high presentation level, which is near the upper limit of the functional input level of most of the speech processors, this test condition represents in essence a sentence test without noise. The aim of the second training list was to introduce the patient to the adaptive test procedure. The mean learning effect was -0.84 dB (SNR) between the second training list and the actual measurement, which was statistically significant. In 13 patients, further measurements were performed after the adaptive SRT determination. The SRT and slope estimates measurements showed a very minor and statistically insignificant improvement in the SRTs, which can be attributed to the further training of the patients or to the different measurement algorithm. These results indicate that two training lists of 20 sentences are sufficient but necessary to eliminate the training effect in order to obtain valid measurements with the Finnish matrix sentence test.

3.4.2 SRTs in Finnish CI users and the correlation with the word test

The mean SRT for an unselected sample of 33 CI patients was -3.60 ±2.09 dB (SNR). There is no reference data available for the Finnish CI recipients other than the patient's own subjective experience and testimonials to the therapists that they are hearing better with the CI than before. A substantial portion of the CI users will score the maximum when tested with the Finnish word test, but may still complain about significant difficulties in everyday listening situations. On the other hand, there are patients with lower scoring at the word test who seem to perform rather well in background noise. This clinical observation is now backed up by the data obtained in this study. Only a weak correlation could be demonstrated between the intelligibility of the Finnish word test in quiet and the SRT with the new Finnish matrix sentence test in noise (Figure 10). In most of the CI users, the word test scoring was between 60 and 100 % (mean 89%) irrespective of the measured SRT.

There is also a clear saturation or ceiling effect encountered in the scoring during the word test. Maximal word test scoring was seen in patients with a final SRT between -2.1 and -8.4 dB (SNR). Since the CI patients were measured with the word test on a regular basis, their high scoring may at least be partly due to the redundancy of the test.

A slightly better correlation was found between the first training list of the matrix test and the final SRT value than for the word test and the SRTs (Figure 9, 10). This can be explained since it is the same structure of the speech material and presentation as well as the fact that a 4-fold amount of items are tested in one matrix test training list as in comparison to the word test (20x5=100 words vs. 25 words). Therefore, the traditional Finnish word test can be replaced by the first training run of the matrix test.

3.4.3 The slope of the intelligibility function in Finnish CI patients

There is a tendency to underestimate the hearing benefit on the basis of the change in the SRT value. In fact the improvement of the SRT value is relatively meaningless when the slope of the intelligibility function is unknown. The slope of the Finnish matrix sentence test is 16.7 ±1.2 %/dB (SNR) for the normal hearing subjects. In order to interpret the clinical improvement of the patients, it can be helpful to use the slope of the intelligibility function to estimate the improvement in terms of intelligibility. In this study, the slope of the intelligibility function was 15.0 ±4.17 %/dB in 13 selected patients with at least 2 years of CI usage and word scoring ≥ 88 %.

Hey et al. (2003) studied the intelligibility function of 9 CI patients measured with the OLSA and found that the slope is only slightly lower than for the normal hearing subjects (15.0 %/dB compared to 17. 2 %/dB). However in a study with 56 experienced CI users, Müller-Deile calculated the mean slope of 10.3 ±3.5%/dB on adaptive measurements (Müller-Deile, 2009), which is considerably lower than the normative slope. In 26 CI users, the median slope was 14 %/dB tested with the OLSA (Hey et al., 2010). In the present study, the mean slope across the subjects was 15.0 ±4.17%/dB. There was no relation between the mean SRT and the steepness of the intelligibility function on an individual basis (see Figure 11 a,b). Müller-Deile (2009) and Hey et al. (2014) observed a tendency towards shallower intelligibility functions in patients with higher SRTs, but this could not be observed in this present series. The relatively small number of patients in this study may explain the difference. Nonetheless, the mean slope function of 15.0 ±4.17 %/dB shows that experienced CI users may be very close to the slope functions of normal hearing subjects, which is consistent with the data obtained by Hey et al. (2014). Therefore the Finnish matrix sentence test allows for accurate SRT measurements also in this group of patients.

3.4.4 *Advanced measurements with the Finnish matrix sentence test*

The Finnish matrix sentence test has been implemented in the clinical work very recently, which means that more advanced measurements have not yet been conducted. Test re-test measurements on separate follow-up visits (> 4 weeks apart) have been conducted only in three patients. The results revealed that test re-test variability was present also in this setting (range -0.2 – 0.5 dB [SNR]). Further measurements will be conducted in the future to validate the reproducibility of the SRT measurements with the Finnish matrix sentence test.

Although the benefit of bilateral CIs cannot be fully appreciated in the S_0N_0 condition, a slight benefit was detected in all of the experienced bilaterally implanted patients. Rader (2008) found a mean benefit of about -2 dB (SNR) in patients with bilateral CI in comparison to patients implanted unilaterally in the S_0N_0 condition measured with the OLSA. More advanced testing in different conditions will be conducted in the future in all of the bilaterally implanted patients in order evaluate if they benefit from the second implant and if so, by how much.

Interestingly in patients using bimodal stimulation, the benefit of the hearing aid with respect to speech understanding in noise was surprisingly poor. In only two patients could a measurable benefit of the hearing aid be observed. Furthermore, worse SRTs were detected in two patients in the bimodal condition. Generally the combination of acoustic and electric stimuli achieves the highest level of speech perception and sound quality (Tyler et al., 2002, Ching et al., 2004). Nonetheless, there is extensive variability in the benefit provided by the additional hearing aid and this also depends on how much residual hearing is available. Since at the time of writing, bilateral implantation in adults is not the common practice in Finland, patients are advised to use their hearing aids, despite their negligible residual hearing which may explain the insignificant benefit.

3.4.5 *The comparison of the rehabilitation results with cochlear implants in Finland and Germany*

The most relevant information of this study is the mean SRT of -3.60 ±2.09 dB (SNR) for 33 unselected CI recipients measured with the new Finnish matrix sentence test. When comparing the normative data of the OLSA and the Finnish matrix sentence test for normal hearing test subjects, the slope of the intelligibility function was very similar, but the SRT was 3 dB lower for the Finnish test. Due to the very comparable slope, the assumption can be made that for the hearing impaired or the CI patient, the test characteristics remain similar in both tests but at a 3 dB different threshold level. Müller et al. (2008), Rader (2008) and Fredelake & Hohmann (2012) and Zirn (2014) reported mean SRTs in the range of -0.7 – 1.7 dB (SNR) for German CI patients tested with the OLSA in the S_0N_0 condition. Slightly better SRTs were measured in 38 selected CI users by Hey et al. (2014) with a median SRT of -2.0 dB (SNR) (range -5 – 6 dB [SNR]). When applying a correction of 3 dB (SRT difference of the OLSA and the Finnish matrix sentence test) to the German results, the rehabilitation result was rather similar. With the presumption of the SRT correction, the rehabilitation results with respect to speech perception in noise between both countries appear therefore to be comparable.

In the future, the comparability of test methods will become more and more important. In anticipation of the free movement of patients within the EU, efforts are being made to ensure that there will be harmonization of the inclusion criteria for treatment. In addition, the quality control and cost effectiveness of therapeutic interventions are also becoming more important and there is an urgent need for a standardized report procedure of treatment results at both the national and international levels.

3.4.6 Indications for the rehabilitation with cochlear implants in Finland

On the basis of these first results, it is apparent that the indication criteria for CI will become much more precise as new data on the rehabilitation results of unilateral and bilateral CI treatment become available in Finland. Before the implementation of the matrix test, CI candidates presented most often with considerable subjective difficulties in their communication abilities. It was very difficult to interpret the results of the Finnish word test to make decisions with respect to CI. The same caution had to be applied to the prognosis of the possible benefit of treatment. The only guideline was the clinical experience that patients treated with a CI should state that they have improved their communication abilities. According to the first results with the Finnish matrix test, more than 90 % of the CI patients reach a SRT in the negative range. This data will help in the decision making process and in the counseling of the patients.

3.5 Conclusion

The new Finnish matrix sentence test was found to be suitable for the testing of CI patients. Valid measurements were possible in the CI patients as well as in patients referred for the evaluation of CI treatment. This is the first time that valid speech reception thresholds in noise have been available for Finnish CI users. Comparable rehabilitation results were found between the Finnish and German CI patients. It is predicted that the Finnish matrix sentence test will become a valuable tool in the evaluation of patients presenting for CI candidacy as well as in the follow-up of patients already supplied with a CI. Nonetheless, more studies, also longitudinal trials, will have to be conducted in the future in order to obtain more knowledge about the characteristics of the Finnish matrix sentence test in CI patients.

4 Summary and conclusions

Spoken language is one of the most important elements of human communication. The ability to understand speech is one of the most crucial functions of human hearing. Considerable advances have occurred in hearing rehabilitation technologies over the past decade including the widespread use of CIs for children and adults with severe-to-profound hearing loss. In addition to restoring hearing, CI treatment also promotes the acquisition of speech in prelingually deaf children. The ongoing success of the rehabilitation results with CIs as well as the continuing expansion of the indications (i.e. single-sided deafness, electric-acoustic stimulation of partly deaf ears) call for improved and more sensitive test procedures for accurately monitoring the patient's speech communication ability. It is essential to be able to assess the patient's hearing status as accurately as possible in order to provide successful and cost-effective treatment. Therefore it is also important to investigate the characteristics and especially the limitations and errors of the administered test procedures.

The only commonly used speech audiometric test for the Finnish language is a two-syllable word test in quiet. The word material was developed over fifty years ago and today not only the test itself is outdated but even many of its words are old fashioned and the test does not fulfill the international ISO standards (ISO 8253-3) for speech recognition tests for audiological applications. These demand minimum requirements of precision and comparability between different test procedures including speech recognition tests in different languages. The test lists are not balanced across their perceptive equivalence and the phonetic balancing has been impaired by the exclusion of some of the original words. Additionally, the test re-test reliability has not been evaluated and no normative noise is available. In a comparable time, a much lower accuracy can be achieved with this word test than with a sentence based speech intelligibility test. Because of the old-fashioned speech material, some words may today sound strange and some are even unrecognizable to young test subjects. Due to the limited amount of speech material (seven lists with 25 words), this test exhibits a high degree of redundancy, thus limiting its application in settings when frequent testing is required, such as the case for CI users. The established word test is appropriate for measurements at the threshold level, but for suprathreshold testing there is the danger of encountering a ceiling effect. Indeed, for hearing rehabilitation applications, the established word test was found not to be sufficiently challenging since most of the patients, even CI users, could reach the highest scores attainable with this test. For example, progress in the hearing rehabilitation actions cannot be monitored and quantified. The subject may continue to improve in terms of speech intelligibility, but the established word test is unable to capture that improvement. On the other hand, due to the absence of an introductory sentence, the presented stimulus begins abruptly so that in some instances, this test may reflect the patient's vigilance rather than his/her functional hearing. Furthermore, the stimulus is too short to allow the sound processing algorithms installed in the hearing aids or CIs to take full effect, resulting in suboptimal scoring in some patients.

Against this background, the Finnish matrix sentence test was developed to fulfil the HearCom and ISO standards for speech audiometry. The Finnish matrix sentence test was evaluated with normal hearing subjects. The speech-reception threshold was -10.1 ±0.7 dB (SNR) and the slope of the intelligibility function was 16.7 ±1.2 %/dB. The characteristics of this new test were compared with similar matrix tests developed for other languages. International comparability was achieved also in the Finnish matrix sentence test.

In the first application study, the Finnish matrix sentence test proved to be suitable for the testing of CI patients. Valid and reliable data could be obtained from this group of patients and speech-reception thresholds could be measured for all of the patients. The mean SRT of 33 unselected CI users in the S_0N_0 condition was -3.60 ±2.09 dB (SNR). The slope of the intelligibility function was 15 ±4.17 %/dB, which was only slightly less steep than for the normal hearing subjects. In the established Finnish word test, a very obvious ceiling effect (word scoring ≥ 96 %) was observed in over 50 % of the CI users. In these patients, the measured SRT, however, ranged from -1.5 to -8.2 dB (SNR), which implies extensive differences between CI recipients in their respective speech perception in noise and, in addition, in their achieved benefit of the CI, despite the fact that they may all score maximal points in the traditional word test. Therefore it is not surprising that there was only a weak correlation between the established Finnish word test and Finnish matrix sentence test. A higher correlation was found between the first test list presented at a fixed speech level (at +10 dB [SNR]) and the adaptively obtained SRT using the matrix test since both ways of conducting the Matrix test should yield a comparable performance measure in the same patient. The fact that more items are tested by using sentences than single words per unit of time also yields a higher efficiency with the sentence-type test. With a time consumption of about 5 min per list, the application of the Finnish matrix sentence test is very time efficient and makes it possible to conduct also more advanced measurements under different conditions (i.e. noise from a different direction) during the same appointment. The test was administered also to patients who were candidates for CI treatment. All measurements could be conducted reliably and in 12 of 16 patients (75 %) the SRT was within the positive range.

In the course of the harmonization of the indications to treat as well as the likelihood that there will be more movement of patients in the different Member States of the European Union, international comparability of test methods as well as treatment results are matters which are increasingly important today. The Finnish matrix sentence test showed the same characteristics in normal hearing subjects as well as in CI users as the German OLSA, but at a 3 dB (SNR) lower threshold level. After accounting for that difference, then highly comparable range of test results could be achieved in normal hearing subjects as well as in CI users in both countries.

The Finnish matrix sentence test is the first evaluated sentence test in noise for the Finnish language, which provides reliable test results with the same characteristics as other equivalent European tests. In the first application study, the test was found to be suitable for the testing CI users as well as CI candidates. Valid measurements could be obtained and the SRTs were determined in 33 unselected CI users. Comparable results for CI treatment could be obtained with Finnish CI users in comparison to German users. It is anticipated that the Finnish matrix sentence test will become a valuable tool in the evaluation of patients presenting for CI candidacy as well as in the follow-up of patients already supplied with CIs. There is, however, a need for more studies to clarify in greater detail the characteristics and limitations of the new test in different groups of patients.

5 References

Aaltonen O, Hellstrom A, Peltola MS, Savela J, Tamminen H, Lehtola H. Brain responses reveal hardwired detection of native-language rule violations. Neurosci Lett. 2008;444(1):56–59.

Bernstein JG, Summers V, Grassi E, Grant KW. Auditory models of suprathreshold distortion and speech intelligibility in persons with impaired hearing. J Am Acad Audiol. 2013;24(4):307–328.

Bosman AJ, Smoorenburg GF. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. Audiology. 1995;34(5):260–284.

Brand T, Kollmeier B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. J Acoust Soc Am. 2002;111(6):2801–2810.

Bronkhorst AW, Plomp R. Binaural speech intelligibility in noise for hearing-impaired listeners. J Acoust Soc Am. 1989;86(4):1374–1383.

Carhart R, Tillman TW. Interaction of competing speech signals with hearing losses. Arch Otolaryngol. 1970;91(3):273–279.

Ching TY, Incerti P, Hill M. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. Ear Hear. 2004;25(1):9–21.

Crandell CC. Individual differences in speech recognition ability: Implications for hearing aid selection. Ear Hear. 1991;12(6 Suppl):100–108.

Cullington HE, Zeng FG. Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. Ear Hear. 2011;32(1):16–30.

Dirks DD, Morgan DE, Dubno JR. A procedure for quantifying the effects of noise on speech recognition. J Speech Hear Disord. 1982;47(2):114–123.

Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. international collegium for rehabilitative audiology. Audiology. 2001;40(3):148–157.

Fredelake S, Hohmann V. Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. Hear Res. 2012;287(1–2):76–90.

Glasberg BR, Moore BC. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand Audiol Suppl. 1989;32:1–25.

Hagerman B. Sentences for testing speech intelligibility in noise. Scand Audiol. 1982;11(2):79–87.

Hallgren M, Larsby B, Arlinger S. A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. Int J Audiol. 2006; 45(4):227–237.

HearCom deliverable D-1-3. Protocol for implementation of communication tests in different languages.Web site. http://hearcom.eu. Updated 20062012.

Hey M, Hocke T, Braun A, Scholz G, Brademann G, Muller-Deile J. Erhebung von normativen Daten fuer den Oldenburger Satztest bei CI-Patienten. 13 Jahrestagung der Deutschen Gesellschaft für Audiologie. 2010:1–3.

Hey M, Hocke T, Hedderich J, Muller-Deile J. Investigation of a matrix sentence test in noise: Reproducibility and discrimination function in cochlear implant patients. Int J Audiol. 2014;In press.

Hey M, Vorwerk W, Langer J, Vorwerk K, Begall K. Vergleich von Satztest im Störschall bei Cochlea Implantat Patienten. 6 Jahrestagung der Deutschen Gesellschaft für Audiologie. 2003.

Hochmair-Desoyer I, Schulz E, Moser L, Schmidt M. The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. Am J Otol. 1997;18(6 Suppl):S83.

Hochmuth S, Brand T, Zokoll MA, Castro FZ, Wardenga N, Kollmeier B. A Spanish matrix sentence test for assessing speech reception thresholds in noise. Int J Audiol. 2012;51(7):536–544.

Hoth S, Muller-Deile J. Audiologic rehabilitation of patients with cochlear implants. HNO. 2009;57(7):635–648.

ISO 8253 - 3:1996 Acoustics – Audiometric test methods – Part 3: Speech audiometry. International Organization for Standardization.

Jansen S, Luts H, Wagener KC, et al. Comparison of three types of French speech-in-noise tests: A multi-center study. Int J Audiol. 2012;51(3): 164–173.

Jauhiainen T. An experimental study of the auditory perception of isolated bi-syllabic Finnish words. Thesis/Dissertation, University of Helsinki. 1974;182(1).

Jokinen K. Presbyacusis. VI. Masking of speech. Acta Otolaryngol. 1973; 76(6):426–430.

Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am. 1977;61(5):1337–1351.

Kollmeier B, ed. On the four factors involved in sensorineural hearing loss. Singapore: World Scientific; 1998. Dau T., Hohmann V. and Kollmeier B., eds. Psychophysics, physiology and models of hearing.

Kollmeier B, Lenarz T, Kiessling J, et al. Contribution to the discussion surrounding the Freiburg speech test. HNO. 2014;62(1):49–53.

Kollmeier B, Lenarz T, Winkler A, et al. Indication for and verification of hearing aid benefit using modern methods of speech audiometry. In German. HNO. 2011;59(10):1012–1021.

Kollmeier B, Wesselkamp M. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. J Acoust Soc Am. 1997;102(4):2412–2421.

Kramer SE, Kapteyn TS, Festen JM. The self-reported handicapping effect of hearing disabilities. Audiology. 1998;37(5):302–312.

Laitakari K. Finnish speech in noise test: Long-term reproducibility. Scand Audiol Suppl. 2001;(52)(52):163–164.

Laitakari K. Speech recognition in noise. Development of a computerized test and preparation of test material. Scand Audiol. 1996;25(1):29–34.

Laitakari K, Laitakari J. New computerized Finnish speech in noise test and binaural hearing. Acta Otolaryngol Suppl. 1997;529:74–76.

Laitakari K, Uimonen S. The Finnish speech in noise test for assessing sensorineural hearing loss. Scand Audiol Suppl. 2001;(52)(52):165–166.

Lenarz M, Sonmez H, Joseph G, Buchner A, Lenarz T. Long-term performance of cochlear implants in postlingually deafened adults. Otolaryngol Head Neck Surg. 2012;147(1):112–118.

Lenarz T, James C, Cuda D, et al. European multi-centre study of the Nucleus Hybrid L24 cochlear implant. Int J Audiol. 2013;52(12):838–848.

Lenarz T, Pau HW, Paasche G. Cochlear implants. Curr Pharm Biotechnol. 2012.

Levitt H. Adaptive testing in audiology. Scand Audiol Suppl. 1978;(6)(6): 241–291.

Levitt H, Rabiner LR. Predicting binaural gain in intelligibility and release from masking for speech. J Acoust Soc Am. 1967;42(4):820–829.

Lutman ME. Speech tests in quiet and noise as a measure of auditory processing. In: Martin M, ed. Speech audiometry. 2nd ed. ed. London: Whurr Publishers Ltd; 1997:63–73.

Lutman ME, Brown EJ, Coles RR. Self-reported disability and handicap in the population in relation to pure-tone threshold, age, sex and type of hearing loss. Br J Audiol. 1987;21(1):45–58.

Luts H, Boon E, Wable J, Wouters J. FIST: A French sentence test for speech intelligibility in noise. Int J Audiol. 2008;47(6):373–374.

Middelweerd MJ, Festen JM, Plomp R. Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. Audiology. 1990; 29(1):1–7.

Mikkonen V. Kirjainten frekvenssi- ja informaatio-ominaisuuksia suomenkielessä. University of Turku; 1969.

Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. J Acoust Soc Am. 1989;85(1):365–378.

Müller J, Brill S, Hagen R, et al. Clinical trial results with the MED-EL fine structure processing coding strategy in experienced cochlear implant users. ORL J Otorhinolaryngol Relat Spec. 2012;74(4):185–198.

Müller-Deile J. Speech intelligibility tests in cochlear implant patients. HNO. 2009;57(6):580–592.

Müller-Deile J, Kortmann T, Hoppe U, Hessel H, Morsnowski A. Improving speech comprehension using a new cochlear implant speech processor. HNO. 2009;57(6):567–574.

Nielsen JB, Dau T. Development of a Danish speech intelligibility test. Int J Audiol. 2009;48(10):729–741.

Nilsson M, Soli SD, Sullivan JA. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am. 1994;95(2):1085–1099.

Oldenburger Satztest: Handbuch und Hintergrundwissen. HörTech gGmbH. 2000.

Ozimek E, Warzybok A, Kutzner D. Polish sentence matrix test for speech intelligibility measurement in noise. Int J Audiol. 2010;49(6):444–454.

Palva T. Studies of hearing for pure tones and speech in noise. Acta Otolaryngol. 1955;45(3):231–243.

Palva T. Finnish speech audiometry. Thesis. University of Turku, Finland; 1952.

Pekkarinen E. Effects of noise and reverberation on speech discrimination. Thesis [Ser D]. Annales Universitatis Turkuensis; 1988.

Pesonen J. Sanamuodot ja niiden kirjainrakenne suomenkielisessä sanomalehtitekstissä. [Res Rep N:o 6/1971]. Dept Spec Educ University of Jyväskylä; 1971.

Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J Acoust Soc Am. 1978;63(2):533–549.

Plomp R, Mimpen AM. Improving the reliability of testing the speech reception threshold for sentences. Audiology. 1979;18(1):43–52.

Plomp R, Mimpen AM. Speech-reception threshold for sentences as a function of age and noise level. J Acoust Soc Am. 1979;66(5):1333–1342.

Rader T, Fastl H, Baumann U. Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. Ear Hear. 2013;34(3):324–332.

Saukkonen P. Frequency dictionary of Finnish. 1.th ed. Porvoo: Söderström; 1979:536.

Shiroma M, Iwaki T, Kubo T, Soli S. The Japanese hearing in noise test. Int J Audiol. 2008;47(6):381–382.

Smiljanic R, Bradlow AR. Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. J Acoust Soc Am. 2011;130(6):4020–4031.

Smoorenburg GF. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. J Acoust Soc Am. 1992;91(1):421–437.

Soli SD, Wong LL. Assessment of speech intelligibility in noise with the hearing in noise test. Int J Audiol. 2008;47(6):356–361.

Summers V, Makashay MJ, Theodoroff SM, Leek MR. Suprathreshold auditory processing and speech perception in noise: Hearing-impaired and normal-hearing listeners. J Am Acad Audiol. 2013;24(4):274–292.

Suomen sanomalehtikielen taajussanasto. CSC – Tieteen tietotekniikan keskus Oy Web site. http://www.csc.fi/tutkimus/alat/kielitiede/ taajuussanasto-B9996/view. Updated 2004. Accessed 2/14, 2012.

Theunissen M, Swanepoel de W, Hanekom J. Sentence recognition in noise: Variables in compilation and interpretation of tests. Int J Audiol. 2009; 48(11):743–757.

Vainio M. Phoneme frequencies in Finnish text and speech. In: Iivonen A, Klippi A, eds. Studies in logopedics and phonetics. Vol 6. Series B ed. Helsinki: Department of Phonetics, University of Helsinki; 1996:181–194.

van Rooij JC, Plomp R. The effect of linguistic entropy on speech perception in noise in young and elderly listeners. J Acoust Soc Am. 1991;90(6): 2985–2991.

Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache: Design, Optimierung, Evaluation des Oldenburger Satztests (Teil 1–3). Z Audiol. 1999;38(1–3):4-15-44-56, 86–95.

Wagener K, Josvassen JL, Ardenkjaer R. Design, optimization and evaluation of a Danish sentence test in noise. Int J Audiol. 2003;42(1):10–17.

Wagener KC, Brand T. Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. Int J Audiol. 2005;44(3):144–156.

Wilson RH, Carnell CS, Cleghorn AL. The words-in-noise (WIN) test with multitalker babble and speech-spectrum noise maskers. J Am Acad Audiol. 2007;18(6):522–529.

Wilson RH, McArdle R. Intra- and inter-session test, retest reliability of the words-in-noise (WIN) test. J Am Acad Audiol. 2007;18(10):813–825.

Ylinen S, Huotilainen M, Naatanen R. Phoneme quality and quantity are processed independently in the human brain. Neuroreport. 2005;16 (16):1857–1860.

Zirn S, Hempel JM, Hemmert W, Keller S, Polterauer D. Psychometric functions of cochlear implant users in fluctuating and steady-state noise. 13th International Conference on Cochlear Implants and Other Implantable Auditory Technologie. Munich, 2014.

Zokoll MA, Hochmuth S, Warzybok A, Wagener KC, Buschermohle M, Kollmeier B. Speech-in-noise tests for multilingual hearing screening and diagnostics1. Am J Audiol. 2013;22(1):175–178.

6 Acknowledgements

A rather critical part of any dissertation is the chapter of acknowledgements. It is an almost impossible task to give credit to everyone who has in some ways contributed to this work. The list of names would go on and on, resembling the closing titles of a movie. Even then, someone would have been omitted, accidentally or incidentally. I honestly would like to thank everyone, who has in any way supported or contributed to this work.

The lack of accurate speech tests for the Finnish language bothered me since the very beginning of my clinical work with hearing impaired patients, in particular with cochlear implant recipients. For example, fitting sessions (of cochlear implants), felt sometimes rather like a waste of time, because there was no test which would allow verification of the possible improvements. Therapy decisions were based on inaccurate clinical data and relied mostly on "*clinical intuition*". Even the benefit of the hearing rehabilitation of both ears could not be tested. The availability of adequate speech tests is also a prerequisite for the implementation of evidence-based medicine in hearing rehabilitation. Therefore the development of a new Finnish speech-audiometric test owes it origins to this rather urgent clinical necessity.

The project around this thesis was initiated in early 2010 with my E-mail addressed to Prof Birger Kollmeier, with the enquiry for possible co-operation in the development of a Finnish version of the *Oldenburger Satztest*. It did not take long and his answer dropped into my mailbox after a few hours: "When do we start?" This was the beginning. Fast forward to 2014: At the time of writing, the new Finnish matrix sentence test has been already implemented in the Helsinki University Central Hospital and the Kuopio University Hospital. In co-operation with the Helsinki Central Hospital, we have also submitted a second manuscript based on the application study found in this thesis in which the new test proved its utility and accuracy in the clinical routine. I am also very pleased that the test will soon be implemented in some commercially available audiometers, which will help in the deployment of the test in the clinics across the country.

And as always in life, this project wouldn't have been possible without *a little help from my friends*. So, here we go:

My special thanks go to Prof. Birger Kollmeier for supervising this thesis and especially for starting our fruitful co-operation in the development and application of the Finnish matrix test. Thanks also to everybody in his team for all the effort, professionalism and helpfulness during all stages of this project. I send a special thank you to Dr. Michael Buschermöhle for all the excellent work we have made together. Many thanks also to Prof. Thomas Lenarz for being my co-referent for this thesis. To the North, across the Baltic Sea on the Finnish side, I would like to thank Dos. Antti Aarnisalo, for his interest and support in this project. Many thanks also to Dr. Ville Sivonen for his support and technical expertise. Thanks also to all the co-authors. Thanks to Med-El Deutschland GmbH for providing the funding to start this project. Thanks also to the Oticon Foundation, Cochlear Nordic Ab, Korvatautien tutkimussäätiö and EVO-funding of the Kuopio University Hospital for providing additional support.

Last but not least I would like to express my deepest gratitude to my dearest love Anna-Kaisa, who made all this possible in the first place and to my boys, Tobias and Benjamin, the greatest chaps of them all.

Kuopio, 13.10.2014
Aarno

7 Appendix

7.1 Original publication

Dietz A, Buschermöhle M, Aarnisalo AA, Vanhanen A, Hyyrynen T, Aaltonen O, Löppönen H, Zokoll MA, Kollmeier B. The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. Acta Otolaryngol. 2014 Jul;134(7):728-37