Carl von Ossietzky
Universität Oldenburg

Faculty V: Mathematics and Science
Master Study: Hearing Technology and Audiology

MASTER THESIS

Development and Clinical Validation of The Indonesian Matrix Sentence Test

written by
Felicia Primadita (3048587)

Supervisors:
Prof. Dr. Dr. Birger Kollmeier and
Dr. rer. nat. Anna Warzybok

Oldenburg, May 12^{th} 2017

Contents

List of Figures

List of Tables

Abbreviations

DTT	Digit Triplet Test	Test using sequences of three digits, usually used as hearing screening test.
EYD	Ejaan Yang Disempurnakan (*eng:Perfected Spelling System*)	Latest spelling standard for Indonesian language.
GÖSA	Göttinger Satztest	Speech audiometry to measure speech Intelligibility level in quiet and in noise using daily sentences.
HI	Hearing impaired	
HINT	Hearing in Noise Test	Speech audiometry developed by [Soli and Wong, 2008] to measure SI using daily sentences.
NH	Normal hearing	
NPT	Nilai Ambang Persepsi Tutur	see SRT
OLSA	Oldenburger Satztest	Oldenburg sentence test/matrix sentence test in German language.
PSS	Perfected Spelling System	see EYD
PTA	Pure-Tone Average	Average threshold of hearing level at 0.5, 1, 2, and 4 kHz obtained from the pure-tone audiogram.
RMS	Root-Mean-Square	
SD	Standard Deviation	
SDS	Speech Discrimination Score	Lowest threshold, where 100% of speech intelligibility is reached.
SI	Speech Intelligibility	
SNHL	Sensorineural Hearing Loss	
SRT	Speech Reception Threshold	Threshold, where 50% speech intelligibility is reached.
SNR	Signal-to-Noise Ratio	

Abstract

Object: The aim of this thesis was the development of the Indonesian matrix sentence test *INDMatrix* for speech intelligibility measurements in noise and the clinical validation with hearing-impaired (HI) patients.

Study design: The test development followed the international standard procedure, and included recordings, cutting of the sound files, optimization of the speech material and evaluation measurements. Firstly, a 50-word base matrix consisting of 10 names, 10 verbs, 10 numerals, 10 nouns and 10 adjectives was established and recorded with a female native Indonesian speaker. In the optimization measurements, the speech reception thresholds (SRT) of 50% intelligibility and the slopes (S_{50}) of the word-specific speech recognition functions were obtained for each word realization. By applying level corrections, the speech material was homogenized in speech intelligibility. The evaluation measurements aimed at the establishment of normative data for normal-hearing (NH) listeners, the assessment of the training effect and the verification of test list equivalence. In cooperation with clinical partners in Indonesia, the INDMatrix test was validated for its application in clinical population.

Study sample: Twenty three NH listeners participated in the optimization measurements conducted in Oldenburg. The evaluation and validation measurements were a multicenter study involving 3 measurement sites (Oldenburg, Bandung, and Jakarta). Total of 25 NH listeners participated in the evaluation measurement in Oldenburg and Bandung, and 23 HI listeners in the validation measurement in Bandung and Jakarta.

Results: At the optimization stage, the word realization level was adjusted to the average word-specific SRT (-8.6 dB SNR) with maximum adaptation of ±3 dB. Due to the level correction, the standard deviation (SD) of the test-specific SRT decreased from 2.2 dB to 0.6 dB. In that way, the homogeneity of speech items in intelligibility was improved. The SRT reference value for the NH-listeners measured adaptively in noise resulted in -8.2 ± 0.8 dB SNR (SD across listeners). At least 2 training lists of 20 sentences are necessary before starting the real measurement to avoid the training effect. The average of the test-specific SRT and slope across test lists was -9.1 dB (SD: 0.1 dB)and 14.5%/dB (SD: 1.0 %/dB). The INDMatrix shows a high test sensitivity of 91.3%.

Conclusion: The final version of the INDMatrix test with 20 test lists of 10 sentences has been successfully established and fulfills the international standards. The results are comparable to those of matrix tests established for other languages.

Zusammenfassung

Ziel: Das Ziel dieser Masterarbeit war die Entwicklung des Indonesischen Satztests *INDMatrix* für Messung der Sprachverständlichkeit im Störgeräusch und die klinische Validierung mit schwerhörenden (SH) Patienten.

Design der Studie: Die Entwicklung des Matrixtests folgte dem internationalen Standardverfahren und umfasste die Sprachaufnahme, das Schneiden von Aufnahmedateien, die Optimisierung des Sprachmaterials und die Evaluationsmessung. Zunächst wurde eine 50-Wort Basismatrix mit 10 Namen, 10 Verben, 10 Nummern, 10 Nomen und 10 Adjektiven konstruiert und von einer Indonesischen Muttersprachlerin aufgenommen. In der Optimierungsmessung wurden die Verständlichkeitsschwellen bei 50% der Sprachverständlichkeit und die Steigung (S_{50}) der wortspezifischen Diskriminationsfunktion für jede Wortrealisierung bestimmt. Durch Anwendung der Pegelkorrektur wurde die Verständlichkeit des Sprachmaterials homogenisiert. Die Evaluationsmessungen zielten auf die Bestimmung der Referenzdaten von Normalhörenden (NH), die Bewertung vom Trainingeffekt und von der Testlistäquivalenz. Durch Zusammenarbeit mit klinischen Partnern in Indonesien wurde INDMatrix zur Anwendung in der klinischen Population validiert.

Probanden: 23 NH nahmen an den in Oldenburg durchgeführten Optimierungsmessungen teil. Die Evaluation- und Validierungsmessungen waren eine Multicenter-Studie, wo 3 Messorten involviert waren (Oldenburg, Bandung und Jakarta). Insgesamt gab es 25 NH für die Evaluationsmessung in Oldenburg und Bandung und 23 SH für die Validierungsmessung in Bandung und Jakarta.

Ergebnisse: Bei der Optimierung wurde die Pegel der Wortrealization auf den Mittelwert der wortspezifischen Verständlichkeitsschwelle (-8.6 dB SNR) mit maximaler Pegelkorrektur von ± 3 dB angepasst. Aufgrund der Pegelkorrektur sank die Standardabweichung der testspezifischen Verständlichkeitsschwelle von 2.2 dB auf 0.6 dB und dadurch verbesserte sich die Homogenität der Verständlichkeit vom Sprachmaterial. Der adaptiv gemessene Referenzwert der Verständlichkeitsschwelle im Störgeräusch betrug -8.2 $\pm$ 0.8 dB SNR (Standardabweichung über gesamte Hörer). Mindestens 2 Traininglisten mit jeweils 20 Sätzen, um den Training-Effekt zu verhindern, setzt die Sprachverständlichkeitsmessung voraus. Der Mittelwert der testspezifischen Verständlichkeitsschwelle und der Steigung über gesamte Testlisten betrug -9.1 dB und 14.5%/dB mit Standardabweichung von 0.1 dB und 1.0%/dB. INDMatrix hat eine hohe Testsensitivität von 91.3%.

Zusammenfassung: Die finale Version vom INDMatrix Test mit 20 Testlisten, jeweils 10 Sätzen wurde erfolgreich erstellt und sie erfüllt das internationale Standardverfahren. Die enstandenen Ergebnisse sind mit den in anderen Sprachen entwickelten Matrixtests vergleichbar.

Abstrak

Tujuan : Tujuan dari tesis ini adalah pengembangan tes kalimat Bahasa Indonesia berbasis matriks *INDMatrix* sebagai alat uji pendengaran untuk mengukur kemampuan percakapan dengan suara latar bising dan juga validasi klinis dengan penderita gangguan pendengaran.

Desain penelitian: Tahap pengembangan ini mengikuti prosedur berstandar internasional yang mencakup tahap rekaman, pemotongan data rekaman, optimalisasi materi bahasa, dan tahap evaluasi. Pertama-tama, 50 kata yang terdiri dari 10 nama, 10 verba, 10 angka, 10 nomina, dan 10 adjektiva dipilih ,disusun menjadi matriks dasar, dan direkam dengan suara wanita berbahasa ibu Indonesia. Melalui pengukuran dalam tahap optimalisasi rekaman didapatkan nilai ambang persepsi tutur (NPT) dan gradien S_{50} fungsi diskriminasi khusus kata dari setiap realisasi kata. Dengan mengkoreksi nilai kata, homogenitas materi bahasa pada matrix dasar meningkat dari segi kejelasannya. Tahap evaluasi bertujuan untuk mendapatkan nilai referensi / nilai baku untuk pendengar normal, menilai *training effect*, dan menguji ekuivalensi setiap deret. Dengan bekerjasama dengan partner-partner klinis di Indonesia, INDMatrix dapat diuji validitasnya dengan populasi klinis.

Subjek penelitian: Dua puluh tiga pendengar normal berpartisipasi dalam tahap optimalisasi yang dilaksanakan di Oldenburg. Tahap evaluasi dan validasi merupakan penelitian multisenter yang melibatkan 3 tempat (Oldenburg, Bandung, dan Jakarta). Sebanyak 25 pendengar normal berpartisipasi dalam tahap evaluasi di Oldenburg dan Bandung. Sebanyak 23 orang dengan gangguan pendengaran ikut serta dalam tahap validasi di Bandung dan Jakarta.

Hasil penelitian: Dalam tahap optimalisasi, nilai realisasi kata disesuaikan dengan rata-rata NPT khusus kata (-8.6 dB SNR) dengan adaptasi maksimal ±3 dB. Dengan adanya koreksi nilai pada kata, standar deviasi (SD) dari NPT khusus test menurun dari 2.2 dB menjadi 0.6 dB. Dengan begitu, homogenitas kejelasan materi bahasa meningkat. Nilai normal NPT yang diukur secara adaptif dengan latar suara bising adalah -8.2 ± 0.8 dB SNR (SD seluruh pendengar). Setidaknya 2 deret, masing-masing 20 kalimat, diperlukan sebelum pengukuran yang sesungguhnya untuk menghindari *training effect*. NPT khusus test dan gradien rata-rata dari seluruh deret adalah -9.1 dB (SD: 0.1 dB) dan 14.5%/dB (SD: 1.0%/dB). INDMatrix menunjukkan sensitivitas yang tinggi sebesar 91.3%.

Kesimpulan: Hasil akhir INDMatrix dengan 20 deret masing-masing 10 kalimat berhasil diciptakan dan memenuhi standar internasional. Hasil penelitian INDMatrix sebanding dengan tes kalimat berbasis matriks yang telah dikembangkan dalam bahasa-bahasa lain.

1. Introduction

In most common situations, target speech is not the only sound perceived by the ears and transmitted to the brain, as ears do not have the ability to separate important and unimportant sounds. Whenever noise exists, noise will be also perceived by the ears and transmitted to the brain. Hearing and understanding speech in background noise occurs in everyday communication and is considered as one of the most complex activities in daily life. Not only hearing-impaired people (with or without hearing aid), also normal-hearing often complain about communication difficulties in noisy environments [Bharadwaj et al., 2015]. It is known that a pure-tone audiogram can not reflect and predict these difficulties [Plomp, 1978]. For a better characterization of a patient's problem, a test assessing her or his communication ability in noise is necessary [Zokoll et al., 2015]. Assessment of speech perception in noise plays an important role in many fields, such as telecommunication, room acoustics, audiology, language and reading disorders, or evaluation of hearing aids and cochlear implants [Ozimek et al., 2010].

According to the World Health Organization (WHO) data [WHO et al., 2007], Indonesia with prevalence of 4.2% in 2002 has the third biggest population suffering from hearing loss in the South-East Asia (SEA) region. According to Statistics Indonesia (the Indonesian official statistics) [Statistik, 2011], about 3 million people (ca.1.5% of 236 million Indonesian population) in Indonesia suffer from hearing difficulties. About 15% of them reported severe hearing difficulties. The difference to the WHO data may occur due to the definition of hearing-impaired population. Statistics Indonesia defines the hearing-impaired population as citizens older than 10 years of age with hearing difficulties despite of hearing aids. Citizens with hearing aids who are able to hearing normally are not counted in the hearing-impaired population. The number of hearing aid- and cochlea implant-users increases. According to Abiratno [Abiratno, 2016b], there are more than 600 cochlea implant-users in Indonesia. These users should be supported by appropriate facilities of hearing diagnostics, fitting, and rehabilitation. Voice test, Rinne and Weber test, and Schwabach test had been used mostly before the independence of Indonesia and are still used in the small clinics for hearing diagnostics [Radhi, 2012a] [Soewito, 1985]. However, these tests do not give any quantitative information of the hearing loss. Pure-tone audiometry was introduced in the big cities in Indonesia in 1960s and is used as a standard in hearing diagnostics. However, pure tone audiometry has shown many deficits, for example: (1) instability of measurement results of hearing thresholds by different audiometricians conducted on the same subject; (2) variations of anatomical structure of the bone and subcutaneous tissues at the mastoid area lead to uncertainty of the bone-conducted measurement results; (3) lack of information about suprathreshold and sensitivity above hearing threshold intensity; (4) inability of pure-tone threshold to predict speech perception ability in noise [Soewito, 1985].

The first speech audiometry test, well-known in Indonesia as *Gadjah Mada (GaMa) Phonetically Balanced Lists* (mono- and bisyllabics), was first introduced at the ENT-national congress in Jogjakarta in 1973 by Soewito. Due to the test's deficits, such as: inhomogeneity of the test lists, poor speaker articulation and incompatibility to the new Indonesian orthography, these test lists were updated, renewed, and expanded to mono- and bisyllabic word lists and lists of phrases and numbers in 1985 [Soewito, 1985]. This caused a huge progress in Indonesian hearing diagnostics, but since the measurements are conducted in quiet, this test does not reflect the patients' complaints of experiencing poor speech recognition in noisy conditions.

The awareness of assessing speech intelligibility in noise in Indonesia has been risen by the establishment of the Indonesian version of the Hearing in Noise Test (IndoHINT) developed by Abiratno [Abiratno, 2016a]. The HINT is one of tests that emphasizes the importance of assessing speech recognition in noise. The HINT [Abiratno, 2016a] [Nilsson et al., 1994] was first developed for the English language in the House Ear Research Institute, LA, USA, using revised Bamford-Kowal Bench(BKB) [Nilsson et al., 1994] sentences and the Speech Reception Threshold (SRT)-method introduced by Plomp and Mimpen [Plomp and Mimpen, 1979]. The HINT in English contains 24 test lists of 10 sentences. The length of the sentences, phoneme distribution, and the intelligibility of the sentences within a list are equivalent, so that HINT fulfilled the standards of realibility and validity.

Abiratno [Abiratno, 2016a] developed 27 test lists of 10 sentences for the IndoHINT. The speech material represents daily conversation situations [Abiratno, 2016a]. HINT has been developed for other languages as listed by Soli and Wong [Soli and Wong, 2008] and can be applied to measure speech intelligibility in quiet as well in noise. However, the available test lists can not be applied twice to the same listener, at least within a short period of time, since the sentences are easily memorized.

The problem of limited number of test lists in a lot of speech audiometry tests using meaningful words or sentences does not occur in the so called "matrix sentence test". The matrix test was invented by Björn Hagerman for clinical purposes to measure speech intelligibility in noise for the Swedish language. The idea of this speech test was to generate sentences from a fixed format, the so called "base matrix" which contains of 5 word groups with 10 words in each group (10 names, 10 verbs, 10 numbers, 10 adjectives and 10 nouns) [Hagerman, 1982]. Sentences are generated by randomly picking up one word of each word groups and combining them into a sentence. The syntactical structure of the sentences is always the same. Each word occurs once in a test list of 10 sentences. The sentences are all grammatically correct, but semantically unpredictable, so it is impossible for the subjects to memorize the sentences or a sentence list [Kollmeier et al., 2015]. In this way, the same subject can be tested repeatedly with the same speech test (even the same test list) [Kollmeier et al., 2015]. Until now matrix sentence test has been developed for more than 14 languages, as listed in [Kollmeier et al., 2015].

Advantages of the matrix test are the high comparibility across language (better than HINT) and the fact that, the limited speech material may be better suited for patients with a substantial degree of hearing loss or higher age than speech tests with more complex speech material [Kollmeier et al., 2015].

Another advantage of the matrix test is the possibility to be performed in a closed-set response format, i.e., the test instructor or audiometrician does not need to comprehend the language of the measured listener. The limited vocabulary of base matrix enables displaying it to the listener, so that the listener can select the understood words by himself/herself on a desktop or touchscreen. In that way, the matrix sentence test can be used all around the world. Developing matrix sentence test for Indonesian language should contribute to an accurate and valid speech audiometry independent on the country the listener lives.

The aim of this master thesis focused on the development of the Indonesian matrix sentence test, which followed the international standard procedure. Furthermore, the Indonesian matrix sentence test (*INDMatrix*) was clinically validated to proof that the INDMatrix is comparable to other audiometry tests applied in the clinics and can complement the result of those tests.

2. Theoretical Background

2.1. Speech intelligibility measurements

Speech audiometry assesses the capability of an individual listener to understand speech [Kollmeier et al., 2014]. It is one of the most important and fundamental components of modern audiology and hearing research. Speech audiometry tests are used in diagnostics, hearing device fitting and evaluation of the benefit of hearing devices, evaluation and assessment of the 'effective' hearing impairments of individual listeners [Kollmeier et al., 2015].

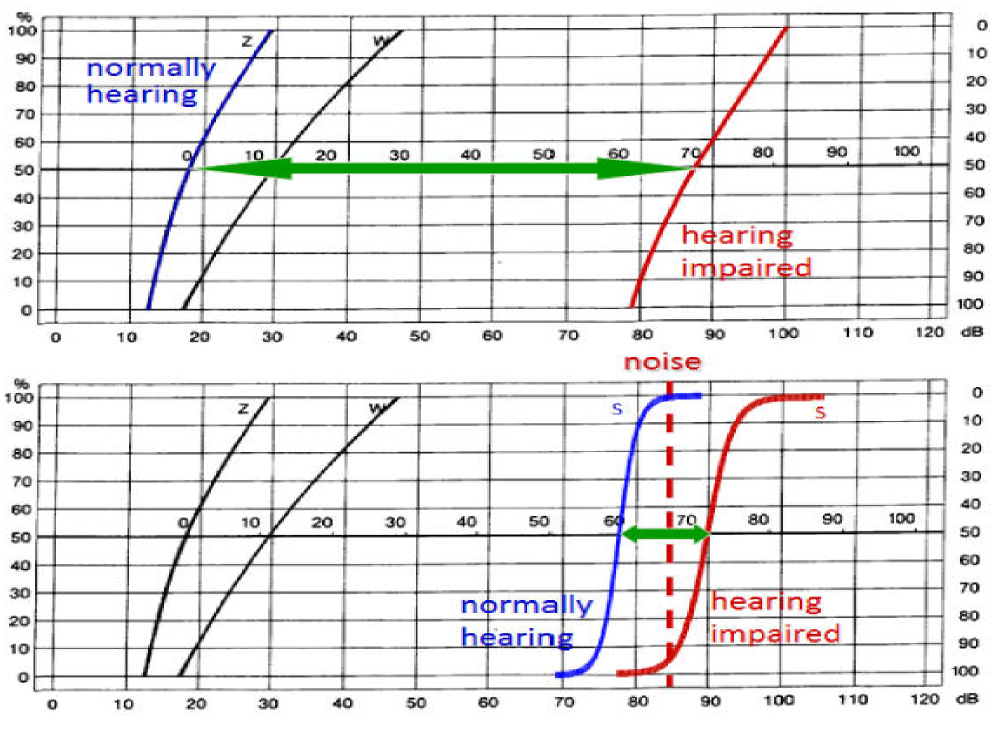


Figure 2.1.: Speech intelligibility measurements in quiet (upper panel) and in noise (lower panel) [Kollmeier et al., 2014].

According to [Kollmeier et al., 2014], speech intelligibility (SI) is defined as the proportion of correctly repeated speech items of a speech test. Speech material may consist of syllables, words or sentences. By definition, SI is quantitatively measureable by means of a standardized speech material (with controlled playback conditions). SI is usually represented in the so called "intelligibility function" that shows percentage of correctly understood speech items as a function of speech level when measurement is conducted in quiet or as a function of signal-to-noise ratio (SNR) when speech intelligibility is measured in noise [Kollmeier et al., 2014].

A speech intelligibility measurement in quiet tests mainly the audibility of speech items. Furthermore, a tone audiogram can be verified. The remaining auditory capacity or discrimination loss can be observed by presenting short speech items at an optimized (individual) presentation level. Speech intelligibility in noise aims at the simulation of a natural speech communication situation and quantifies not only the consequence of loss in audibility but also characterizes the suprathreshold processing deficits.

Figure 2.1 shows the inter-individual differences between normal-hearing (NH) and hearing-impaired (HI) listeners between two common performances of speech intelligibility: The upper panel presents a speech intelligibility measurement in quiet, the lower panel presents a speech intelligibility measurement with noisy background. Measuring speech intelligibility in quiet yields a relatively large inter-individual difference between NH and HI listeners and the accuracy of this measurement is moderate (about 5 dB). Measuring SI in noise, on the other hand, yields a small the inter-individual difference and the accuracy of this measurement to distinguish a NH from a HI-listener has to be higher (about 1 dB) [Kollmeier et al., 2014].

Another important aspect to be considered while inventing a new speech test is the number of speech items. To understand the importance of the number of speech items, output of each single speech item can be considered as a Bernoulli trial. There are only 2 possible outcomes in repeating pronounced words: true or false, i.e. the accuracy of SI measurement is given by binomial distribution. The standard error SE describes how big is the possibility that a measurement result was obtained by chance. The smaller the error, the smaller is the chance that SI is incorrectly measured. The following equation shows the influence of N number of speech items of a speech test on the SE-value:

$$SE(SI) = \sqrt{\frac{SI(1-SI)}{N}} \tag{2.1}$$

When observing SE at 50% of speech intelligibility, as shown by the dashed line in Figure 2.2a, it is obvious that increasing the number of speech items decreases SE. A high number of speech items in an SI-measurement can be achieved by using lists of sentences instead of lists of words. For example: a test list of Freiburg speech test with 20 words (N=20) has an SE of 11.2%, a test list of a matrix test with 20 five-word sentences (N=100) has an SE of 5%. Figure

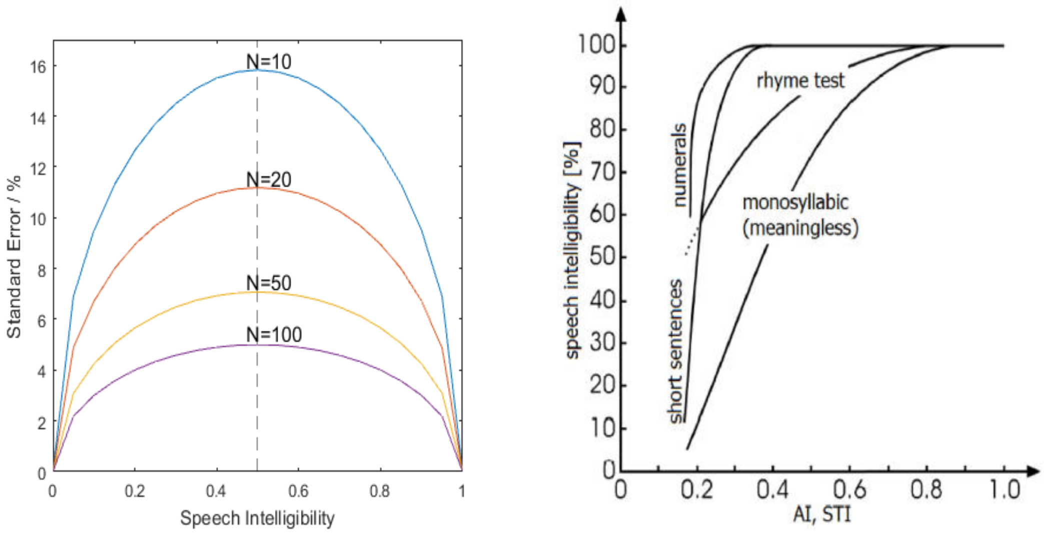


Figure 2.2.: Panel (a): Influence of the number of speech items (N) on accuracy of speech intelligibility measurements described by standard error. Panel (b): Speech intelligibility functions for several speech audiometry tests. [Kollmeier et al., 2014]

2.2b shows speech intelligibility functions for speech tests using different type of speech material including single words, numbers, or sentences. Generally, speech tests with a high number of speech items are characterized by a steep speech intelligibility function. The idea is, as soon as one element is intelligible, other elements become also intelligible, so that a steep transition from "not heard" to "heard" can be reached [Kollmeier et al., 2014]. A speech intelligibility test desires a steep slope, so that small changes in speech level can demonstrate a huge difference in the speech intelligibility value.

2.2. Speech audiometry in Indonesia

In Indonesia, there are three tests that can be used for speech audiometry: the voice test, the updated phonetically balanced list, and the Indonesian version of the Hearing in Noise Test (IndoHINT). The voice test and the phonetically balanced list measure speech intelligibility in quiet. IndoHINT measures speech intelligibility in noise.

The voice test (speech distance test) assesses speech intelligibility for whispered speech pronounced from a certain distance with expiratory reserve volume (air volume after a normal exhalation) [Boenninghaus and Lenarz, 2006]. Voice test probably does not count as a real speech audiometric test, but this test has been known before the independence of Indonesia. It is unclear, who developed or introduced this test in Indonesia. Similar to the voice test explained by [Boenninghaus and Lenarz, 2006], voice test in Indonesia used firstly numbers. Nowadays, common words are used, mostly nouns that exist around the instructor or patient [Radhi, 2012b] [Kartika, 2007].

The voice test is still commonly used in small clinics in Indonesia without a special ENT-department and audiometry. The advantage of the voice test is a very quick screening and high feasibility due to the simple technical equipment requirement (only room with minimal length of 6 m, which has no echo or which wall is damped with an absorbent material in order to minimize echo).

The eyes of the patient and the not inspected ear should be closed. The audiometrician/audiologist/instructor starts to "whisper" 5 or 10 words with a distance of 1 m from the patient. When the patient correctly repeats all the words, the distance will be increased by 2 m and 5 or 10 words are whispered again. The procedure will be repeated until only 80% of the words are correctly repeated. This distance (in m) is the so called "hearing range". Depending on the distance of hearing, the qualitative level of hearing loss (mild, moderate, or severe) is determined and the type of hearing loss (sensorineural or conductive) is determined according to which syllables were often falsely interpreted by the patient [Kartika, 2007] [Radhi, 2012b].

Disadvantages of the voice test are the non existence of quantitative information of the hearing loss in dB, the absence of sound level control of the instructor, the inability of SI-measurement in noise, and the low contextual cues of words that are used.

As mentioned in Chapter 1, the Gadjah Mada phonetically balanced list was the first speech audiometric test introduced in Indonesia in 1973 [Soewito, 1985]. Soewito improved this test by updating the words on the test lists of mono- and bisyllables, extending this test with lists of acronyms, phrases and numbers, and providing recordings of the test lists. He found out that the new phonetically balanced lists with numbers and phrases can be used to determined SRT, while lists with monosyllables and bisyllables can be used to measure the speech discrimination score (SDS). These new developed lists are the most common speech audiometry tests used in Indonesia, especially in clinics with an ENT-department. However, audiometricians/audiologists nowadays mostly apply the bisyllabic lists to measure both SRT and SDS. Due to the recorded speech material, the diagnostic quality is more precise than the voice test. The disadvantage, that only lists with bisyllabic words are used, is the limited number of test lists (n=10). It requires a high control on the used test list, so that the same test list is not applied to the same patient in a short period of time. No contextual cues exist in this test as only words are used. However, the words are meaningful and can be easily memorized.

The IndoHINT developed by [Abiratno, 2016a] is until today the only possible test to assess speech intelligibility in noise in Indonesian language. Speech intel-

Table 2.1.: Example of the renewed phonetically balanced word lists [Soewito, 1985].

Monosyllables	Bisyllables	Numbers	Phrases
Lem *glue*	Sabun *soap*	13	Apa kabar *how are you?*
Bu *ma'am*	Kuda *horse*	67	Besar kecil *big small*
Skor *score*	Dingin *cold*	81	Cocok tanam *cropping*
Gong *gong*	Banyak *many*	55	Gerak badan *body movement*
Jin *genie*	Gula *sugar*	70	Hasil bumi *crops*
Wol *wool*	Pipi *cheek*	92	Iri hati *envy*
Zat *substance*	Besar *big*	28	Juru tulis *secretary*
Pot *pot*	Enak *delicious*	34	Kapal terbang *airplane*
Mi *noodle*	Lidah *tongue*	76	Lemah lembut *gentle*
Sah *legal*	Kembar *twins*	49	Matahari *sun*
Plek *one on one*	Umur *age*		Meja tulis *desk*
Tos *high five*	Salon *salon*		Nenek moyang *ancestor*
Krim *Cream*	Tikus *mouse*		Obat tidur *sleeping pills*
Spoor *Track*	Panah *arrow*		Panjang pendek *long short*
Klab *club*	Becak *trishaw*		Papan tulis *whiteboard*
Loop *loop*	Nasi *rice*		Roti tawar *white bread*
Stem *stem*	Ilmu *knowledge*		Soal jawab *questioning*
Drop *drop*	Kamar *room*		Suka rela *voluntary*
Cek *check*	Telor *egg*		Tahun baru *new year*
Bin *bin*	Tempat *place*		Uang kecil *small change*
Ros *rose*			
Lang *hose*			
Step *step*			
Dan *and*			
Sel *cell*			

ligibility measurement in quiet background is also possible with the IndoHINT. The IndoHINT uses everyday sentences and therefore has highly contextual cues. Example of the IndoHINT sentences [Abiratno, 2016a]: "Ayah bangun jam lima pagi" (*Father stands up at five o'clock in the morning*), "Pisau itu sangat tajam" (*The knife is very sharp*), "Kakek sedang makan roti" (*Grandfather is eating bread*).

The IndoHINT has not been applied yet in the clinics. Despite 27 available test lists, the repetition of the same test list with the same patient within a short period of time should not be avoided since sentences are easily memorized. HINT has been developed for many languages [Soli and Wong, 2008], but it is not possible to execute HINT in a closed-set response (without instructor who masters the language).

2.3. Indonesian as a language

The Indonesian language has been announced as the unity language in Indonesia since October 28th, 1928 during the Second Youth Congress in Jakarta. Indonesian as a national language for 236 million of Indonesian population is regulated in the Constitution of Indonesia (*Ind: Undang-Undang Dasar Republik Indonesia 1945*, UUD'45), which was authorized on August 18th,1945.

Before Indonesian was declared as official language, people spoke local languages. Abiratno in her dissertation reported 748 local languages that are used in Indonesia [Abiratno, 2016a]. According to Statistics Indonesia [Statistik, 2011], the number of regional/local languages almost reached 2,500, which is almost twice as the number of ethnic groups (1,340 groups are registered). The high number of local languages occurs due to the geographical situation of Indonesia with more than 16,000 islands spread from 95°to 141°E.

Based on an official publication of the Indonesian population census in 2010, the Indonesian language was reported as the second most used daily language in Indonesia with only 19.9%, after Javanese with 31.8%. In comparison to the census from 1990 with 10.7%, more and more people used Indonesian as main daily language at home. The Indonesian language as normal daily communication language has been reported to be spoken only by the majority in 5 of 34 provinces: Jakarta, West Papua, Riau, North Sumatra und East Borneo [Statistik, 2011]. However, more than 96% of the Indonesian population are able to speak Indonesian and only around 3% are not.

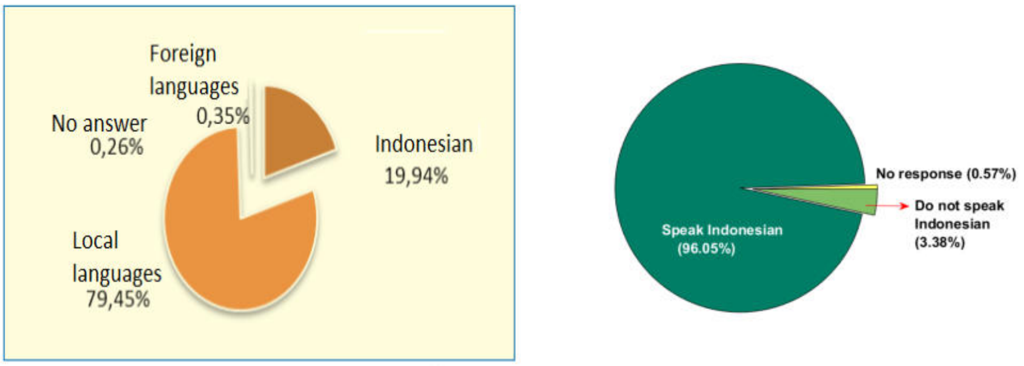


Figure 2.3.: Two of the results of the Indonesian population census in 2010 [Statistik, 2011] [Statistik, 2012].
(a) Indonesian population based on the daily language spoken at home [Statistik, 2011].
(b) Indonesian population based on their ability to speak Indonesian.

The ratio of Indonesian population who can not speak Indonesian to who can speak Indonesian is higher for age group of 75-89 years (almost 1) and age group above 90 years (slightly more than 1) than the younger age groups [Statistik, 2012]. This may occur, because the transition from colonial language (Dutch) to Indonesian language at schools, on public media, and in the Indonesian parliament did not happen directly after the declaration of Indonesian as the national language in the Youth Congress in 1928. The interest of using Indonesian language, however, increased regarding the increasing number of written official speech, articles, and books in Indonesian language. Until March 8^{th}, 1942, Dutch was a mandatory language at schools. At the time of colonialism of Japan (1942-1945), Dutch language was forbidden and Indonesian language became compulsory and was introduced all over Indonesia. Social contact between Indonesian language and the regional languages occured and they influenced and/or complemented one another [Muljana, 2008].

There are two main reasons why Indonesian as a relative new language was adapted mainly from Malay language:

1. Intralinguistical factor: in comparison to other local languages, such as Javanese or Sundanese, which is more popular than Malay, Malay is simple and has no "politeness level", i.e. younger generation may use the same language when talking to older generation and vice versa. Malay can be understood, even when the speaker does not master the language fluently.

2. Extralinguistical factor (social and cultural background): Malay is well-known among Indonesian archipelago and had been introduced in Indonesia during the victory era of Srivijaya Kingdom (8^{th}th-12^{th} century) through trading traffic [Armandhani, 2012] [Muljana, 2008].

Table 2.2.: Indonesian consonants

Articulation Manner / Articulation place	Bilabial	Labiodental	Alveolar	Palatal	Velar	Glotal
Stop						
voiceless	p		t	c	k	
voiced	b		d	ɟ	g	
Fricative						
voiceless		f	s	ʃ	x	h
voiced			z			
Nasal						
voiced	m		n	ɲ	ŋ	
Trill						
voiced		r				
Lateral						
voiced			l			
Semivowel						
voiced	w			j		

The Indonesian spelling system has changed several times since 1928. The valid spelling used nowadays is the Enhanced Indonesian Spelling System (*Ind: Ejaan*

Yang Disempurnakan, EYD) or the Perfected Spelling System (PSS). PSS, compared to the van Ophuijsen or Soewandi Spelling System (the previous spelling systems), is basically combination of 2 spelling systems:

1. *Phonetic spelling as PSS basis*: every alphabet represents a phoneme after sounds measured. Moreover, punctuation marks are reduced, which are mostly unnecessary.

2. *Etymological spelling*: spelling system that regulates the spelling of every word that derived/adapted from (foreign) languages [Muslich, 2008].

In PSS, there are 2 kinds of phonemes:

1. vowel phonemes: Indonesian language has 6 vowels: in total 2 high vowels /i/ and /u/,3 mid vowels /e/,/ə/,/o/, and one low vowel /a/.

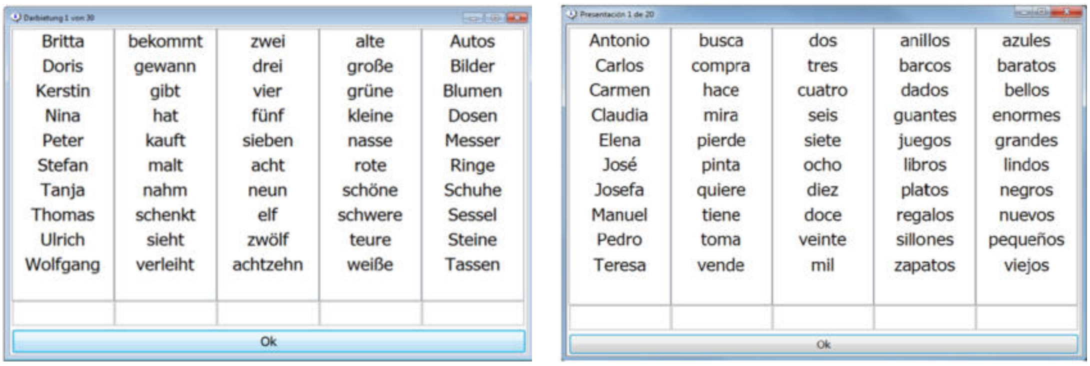
2. consonant phonemes: There are 22 consonants in Indonesian languages. (See Table 2.2) [Muslich, 2008] [Moeliono and Dardjowidjojo, 1988].

2.4. Matrix test

The matrix sentence test was first developed by Björn Hagerman in 1982 for the Swedish language for speech intelligibility measurements in background noise [Hagerman, 1982]. The matrix test is based on a fixed "base matrix" which contains 50 words of 5 word groups (10 names, 10 verbs, 10 numbers, 10 adjectives and 10 objects). A sentence is generated by taking one word randomly from each word group and combining them into a grammatically correct sentence [Hagerman, 1982]. The generated sentences always have the same syntactical structure. Each of 50 words occurs only once in a list of 10 sentences. The sentences are semantically unpredictable, so they might be less redundant than the everyday sentences and are difficult to memorize. A 50-word base matrix enables generation of 100000 different sentences.

In 1999, matrix test was also developed for the German language and is known as OLSA (Oldenburger Satztest). Compared to the matrix test for Swedish, the recording was refined by recording sentences instead of words and including co-articulation between the words to keep the speech sounds natural [Wagener et al., 1999c]. Adaptive procedure proposed by Brand and Kollmeier [Brand and Kollmeier, 2002] was used for the measurements to obtain 50% of speech intelligibility [Wagener et al., 1999a] [Wagener et al., 1999b].

Since then, the matrix sentence test has been developed for other languages. The order of the word groups depends on the grammar of the language, for example: name - verb - number - adjective - noun in English or name - verb -



(a) Oldenburg Sentence Test (OLSA)   (b) Spanish matrix test

Figure 2.4.: The order of respective word groups of the matrix sentence tests depends on the grammar of the language. In German, the word order is name - verb - number - adjective - noun. In Spanish, the word order is name - verb - number - noun - adjective.

number - noun - adjective in Spanish (see Figure 2.4). The noise signal is typically generated from the speech material of the test so that the long term spectrum of speech signals and noise are very similar.

During the EU-funded projects NATASHA and HearCom (www.hearcom.eu), the matrix sentence test was successfully developed for eight languages. The follow-up project "HurDig" lead by Prof. Dr. Dr. Birger Kollmeier emphasized the research focus on the development of hearing screening and diagnostic speech intelligibility tests, not only for European languages (e.g., Spanish and Italian) but also for non-European languages (e.g., American English and Persian). Together in some bilateral projects, this project yielded additional six languages for the matrix sentence test. All available languages for the matrix test (finished until 2014) are listed in Kollmeier et al., 2015 [Kollmeier et al., 2015].

According to [Wagener et al., 1999c], a matrix sentence test has some advantages, such as:

1. Optimized for speech intelligibility measurement in noise.

2. Steep discrimination function.

3. All test lists contain the same 50 words, i.e. the homogeneity of the speech material is very high.

4. The sentences are difficult to memorize so that the number of test lists available is unlimited.

5. Selected words are well-known and daily-used.

6. Closed-set response format enables measurement in listener's native language, even when the test instructor does not master the language.

The disadvantage of a matrix sentence test is the learning / training effect on inexperienced and experienced listeners. The SRT-value decreases significantly when listener gets used to the words of the base matrix. [Hagerman, 1982] [Wagener et al., 1999a]. At least two training lists of 20 sentences needed to be executed before getting the real SRT [Warzybok et al., 2015a] [Wagener et al., 1999b].

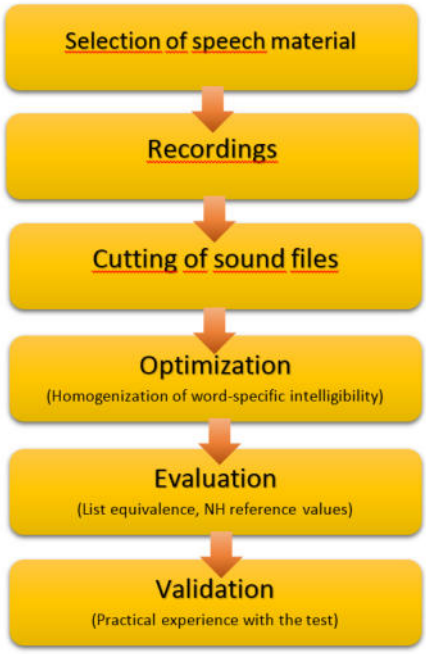


Figure 2.5.: Matrix sentence test development procedure [Kollmeier et al., 2015].

Figure 2.5 shows several steps in the development process as suggested by [Kollmeier et al., 2015]. It starts with the selection of the speech material. Fifty words consisting 10 names, 10 verbs, 10 numerals, 10 nouns, 10 adjectives are denoted as base matrix. The chosen words based on [Akeroyd et al., 2015] have to be highly frequent words and semantically neutral. The phoneme distribution of the matrix sentence test does not differ significantly from the phoneme distribution of language-specific references. For the recording, 100 sentences are generated in such an order, that every word is connected once with each word of the neighboring word group. These sentences are recorded by a female speaker as a compensation between male and infant voice [Kollmeier et al., 2015]. In the third step, cutting the sound files, the words are cut including the co-articulation of the word of the next word group. After cutting, new sentences are resynthesized. During the resynthesis, each word has to be followed by the word of the matched co-articulation. The test-specific noise is also generated as suggested in [Akeroyd et al., 2015] and [Kollmeier et al., 2015]. The optimization aims to increase the homogeneity of the word-specific intelligibility. In the evaluation stage, the normative reference value is obtained and the test list equivalence is analyzed. The validation is the last stage, where matrix sentence tests will be compared with other existing audiometric tests in a clinical population [Kollmeier et al., 2015].

3. The Indonesian Matrix Sentence Test INDMatrix

The development of the Indonesian matrix sentence test followed the international standard procedure of multilingual matrix sentence tests [Kollmeier et al., 2015] and ICRA recommendations [Akeroyd et al., 2015]. The development process of the INDMatrix consists of several steps as shown in Figure 2.5. Each of these steps will be described in the subsequent sections.

Section 3.1 describes the selection of the speech material. Recording of speech material, cutting procedure, and the resynthesis of test sentences will be explained in Section 3.2. The remaining steps will be explained in detail (including methods and results) in seperate chapters: The optimization measurements to reach high homogeneity of speech items is presented in Chapter 4. The evaluation measurements to obtain normative reference values and proof test list equivalence is described in Chapter 5. Chapter 6 presents the last step of the test development, namely the clinical validation. The measurement results are discussed in Chapter 7.

Evaluation and validation measurements were conducted as a multicenter study involving two Indonesian hearing centers in addition to Carl von Ossietzky University of Oldenburg:

1. **A. Kasoem Hearing and Speech Center** is a family company, one of the distributors of hearing-aids and cochlea-implants in Indonesia. Besides offering services in hearing aids and cochlea implants, Kasoem offers audiological diagnostics, consultation, and therapy. The A. Kasoem Hearing and Speech Center trains audiometricians and (auditory-verbal) therapists on audiological measurements or auditory-verbal therapy methods. It is responsible for audiometrical setup calibration for many ENT-clinics and ENT departments of hospitals, including Cipto Mangunkusumo General Hospital. The headquarter of A. Kasoem Hearing and Speech Center is located in Jakarta. However, the evaluation and validation measurements involved the branch in Bandung under the supervision of Dr. dr.[1] Ratna

[1] dr.: the title for practicing physician.

Anggraeni, Sp. THT-KL(K)[2], M. Kes[3].

2. **Cipto Mangunkusumo General Hospital** in Jakarta is one of the largest national hospitals. As an academical hospital, Cipto Mangunkusumo General Hospital cooperates with the medical faculty of the University of Indonesia. Every clinical research project involving Cipto Mangunkusumo Hospital involves the medical faculty of the University of Indonesia. The clinical validation took place in the Division of Neurootology in the Department of ENT under the supervision of dr. Ronny Suwento, Sp. THT-KL(K), dr. Widayat Alviandi, Sp. THT-KL(K), and Prof. Dr. dr. Jenny Bashiruddin, Sp. THT-KL(K).

3.1. Speech material

The speech material of the Indonesian matrix sentence test was selected considering the ICRA recommendations [Akeroyd et al., 2015]. The chosen words are highly frequent according to the frequency dictionary of the Indonesian language [Quasthoff et al., 2015]. Most of the words belong to the first 10,000 common words in Indonesian. The words are semantically neutral. Sentences that can be resynthesized by the selected words are grammatically correct.

Table 3.1.: Fifty-word base matrix of the Indonesian matrix sentence test.

Index	Name a	Verb b	Numeral c	Object d	Adjective e	Translation
0	Agus	melihat	delapan	bola	bagus	*Agus sees eight nice balls*
1	Arif	membawa	dua	buah	baru	*Arif brings two new fruits*
2	Ayu	membayar	empat	buku	berat	*Ayu pays four heavy books*
3	Fajar	memilih	enam	cangkir	bersih	*Fajar chooses six clean cups*
4	Ika	mencari	lima	gitar	kecil	*Ika looks for five small guitars*
5	Made	mendapat	sembilan	kaleng	keras	*Made gets nine hard cans*
6	Maya	mengangkat	sepuluh	kotak	kotor	*Maya lifts ten dirty boxes*
7	Putri	mengganti	seratus	kunci	kuning	*Putri changes one hundred yellow keys*
8	Putu	menjual	tiga	sendok	mahal	*Putu sells three expensive spoons*
9	Sari	menyimpan	tujuh	sofa	putih	*Sari stores seven white sofas*

All names were chosen from the top 1,000 first names of the Indonesian population calculated from 56,719 names of applicants for the Indonesian Treasury Department in 2010 [Hana, 2010] and cross-checked with the names in the Indonesian frequency dictionary [Quasthoff et al., 2015]. There are 3 male, 5 female, and 2 neutral names (gender independent).

All names, nouns and adjectives are bisyllabic. Sentences consist of subject-verb-object and are categorized as active transitive sentences (the verb represents

[2]Sp. THT-KL(K): the title for specialist physician in ENT
[3]M. Kes: Master of Health Science

action that has to be followed by an object). Indonesian verbs for the active transitive sentences start mostly with the prefix *meng-* (some are followed by suffix *-i* or *-kan*). The derivation of the prefix "meng-" depends on the first phoneme of the verbs' infinitive (verbs in the original form). This process is called morphophonemics [4] [Moeliono and Dardjowidjojo, 1988].

1. Prefix *meng-* do not change the form when added to the infinitive verb forms with the first phoneme /a/, /i/, /u/, /e/, /o/, /e/, /k/, /g/, or /h/.
   - meng- + angkat (*to lift*) = mengangkat
   - meng- + ganti (*to change*) = mengganti

   Exception: meng + kalah = mengalah

2. Prefix *meng-* changes to *me-* when added to infinitive verb forms with first phoneme /l/, /m/, /n/, /ɲ/, /ŋ/, /r/, /j/, or /w/.
   - meng- + lihat (*to see*) = melihat

3. Prefix *meng-* changes to *men-* when added to infinitive verb forms with the first phoneme /d/, /t/, /c/, /ɟ/, or /ʃ/.
   - meng- + dapat (*to get*) = mendapat
   - meng- + cari (*to search*) = mencari

4. Prefix *meng-* changes to *mem-* when added to infinitive verb forms with the first phoneme /b/, /p/, or /f/.
   - meng- + bayar (*to reply*) = membayar
   - meng- + potong (*to cut*) = memotong

5. Prefix *meng-* changes to *meny- /meɲ-/* when added to infinitive verb forms with the first phoneme /s/.
   - meng- + simpan (*to save/keep/store*) = menyimpan

6. Prefix *meng-* changes to *menge-* when added to monosyllabic infinitive verb forms.
   - meng- + cek (*to check* = mengecek
   - meng- + rem (*to brake* = mengerem

According to [Soewito, 1985],the percentage of correctly recognized words increases with increasing number of syllables in a word. Therefore, only active

[4] Morphophonemics is the study of the relationship between morphology and phonology. Morphophonemics investigates phonological variations within morphemes.

transitive verbs with prefix *meng-* (without suffix) were selected to keep the number of the syllables in the verbs minimal (3 syllables are the minimum number of syllables of a derivated verbs) but grammatically correct. To avoid the explodation of the phoneme $/m/$ in the phoneme distribution, the number of the phoneme $/m/$ was minimalized in the other word groups.

Indonesian natural numbers (larger or equal 1) are generally multisyllabic. Only numbers 1 to 7 are bisyllabic. To keep the number of syllables balanced within the group of numerals, 6 bisyllabic numbers and 4 trisyllabic numbers were chosen.

The phoneme content of the 50-word base matrix was analyzed and checked by an Indonesian computational linguist , Totok Suhardiyanto, from the Department of Linguistics, Faculty of Humanities of the University of Indonesia. The phoneme distribution of the base matrix, as shown in Figure 3.1 (blue line), was compared to three different language-specific references (grey lines): the phoneme distribution according to the prepatory work for the new phonetically balanced word lists [Soewito, 1985], the phoneme distribution of the IndoHINT, and the Handbooks of Linguistics and Communication Science (in German: *Handbücher zur Sprach- und Kommunikationswissenschaft* HSK). These language-specific references used different methods of selecting the speech corpus. Soewito's distribution based on 50,000 words recorded in public places (groceries, train stations, offices, (junior and senior) high schools and universities) and news reading from national radio and TV. HSK used the Indonesian dictionary as corpus. Abiratno represented the phoneme distribution calculated from 750 test sentences.

Chi-square tests were performed to examine the relation between phoneme distribution of INDMatrix and the language-specific references. There was no significant difference found between the phoneme distribution of the INDMatrix and the phonetically balanced word lists, the HSK, and the IndoHINT ($\chi^2 = 7.229, \chi^2 = 5.303, \chi^2 = 4.294$, respectively).

3.2. Recording and generation of masking noise

The recording of the speech material and the generation of the masking noise was conducted according to the procedure of Wagener et al. [Wagener et al., 1999c]

The recording session took place in a sound-attenuated room fulfilling ISO 8253-3, 2012 in the House of Hearing (Haus des Hörens) at Marie-Curie-Straße 2 in Oldenburg on July 8^{th}, 2016. The recording was carried out using a Neumann 184 microphone with a cardioic characteristic (Georg Neumann GmbH, Berlin, Germany) and a Fireface UC soundcard using a sampling rate of 44.1 kHz and saved on a PC hard disc as *.wav-files, 32 bits, using Adobe Audition 2.0 (Adobe

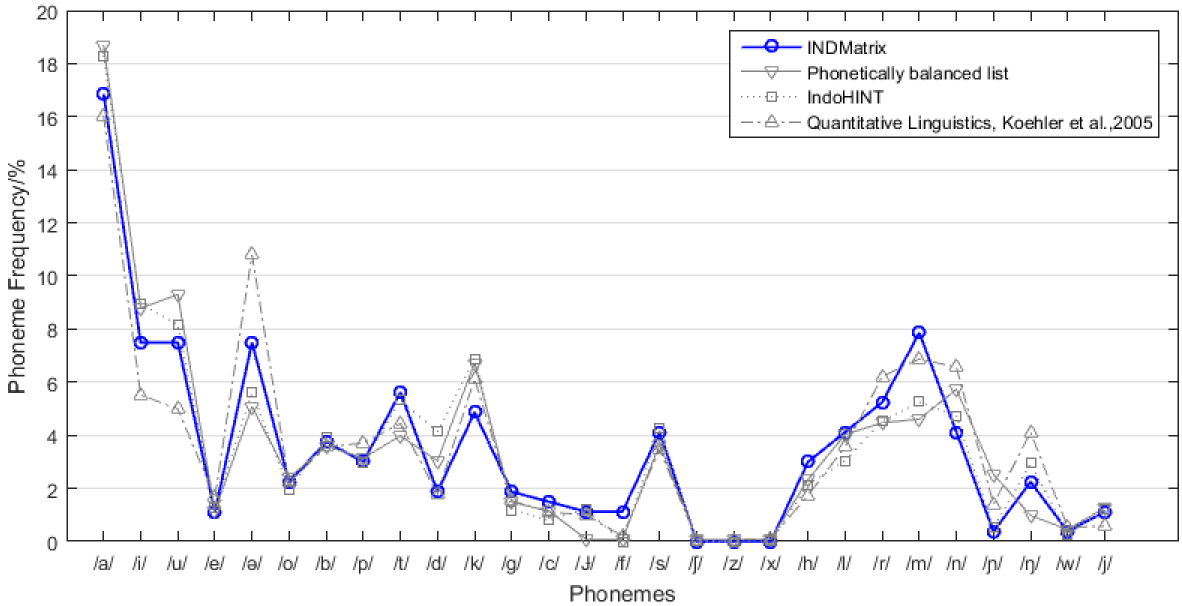


Figure 3.1.: Phoneme distribution of the Indonesian matrix test in comparison to some Indonesian phoneme distribution references.

Systems Incorporated, San José, USA).

The speaker read 100 sentences at an intermediate speaking tempo 4 times. In these 100 sentences, there were 10 realizations of each word from the base matrix combined in a way that all possible transition to the following word were included, i.e., each name was followed once by every verb, each verb is followed once by every number, etc. The average speaking rate was about 127 wpm. The speaker was asked to keep the same speech effort during the whole recording session and to avoid any exaggerated pronounciation, which could lead to unnatural speech cues.

3.2.1. Speaker profile

The speaker, Anne Ivana Samanhudi, chosen according to the ICRA recommendation [Akeroyd et al., 2015] is a female, accent-free Indonesian native speaker, 25 years old, at the time of recording. The speaker grew up, attended school, and worked in Jakarta. She finished her bachelor degree in psychology at the University of Indonesia, Depok. Depok belongs to one of the metropole regions of Java, where most of the people speak Indonesian on a daily basis. The speaker is also able to use one local language, but not as a native language and she admitted not to use this language or any regional languages on a daily basis. This dialect was used neither at school nor at home. By the time of recording, the speaker has been living in Hamburg for less than 1 year. However, the Indonesian language is still used in daily conversations with friends and family, so that the influence of the local dialect as well as German accent on her Indonesian pronounciation can be excluded. Another consideration was that Anne I. Samanhudi worked also as a piano teacher for 6 years before studying in Hamburg, so it can be assumed that she is trained to control the speed of speech and the intonation.

3.2.2. Cutting speech material

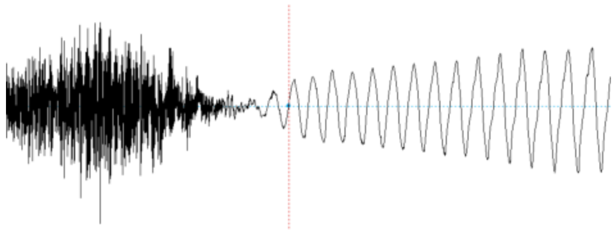


Figure 3.2.: Example of cutting the words "Agus" and "membayar" (*pays*). At the transition, co-articulation to the word "membayar" was included in the previous word "Agus". None of co-articulation of the word "Agus" was included in "membayar".

The speech material was first high-passed filtered with a cut-off frequency of 60Hz. The start and the end of the sentences were marked using the phonetic software "Praat". If the sentence started with fricatives, such as "s", the boundary was placed directly at the sounds beginning. If the sentence started with plosives like "p", "b", "d", or vowels like "a" or "e", the boundary was set about 20 ms before the sentence beginning. Using a script in the software "Praat" the boundaries are automatically shifted to the nearest zero-crossing and using MATLAB the recording was cut into sentences.

From several recordings of a given sentence, the best version was picked out, considering pronounciation, tempo and intonation. The root mean square (RMS) of the best version of all sentences was computed. The level of all sentences was then set to the average RMS (-30 dB). The chosen 100 sentences were segmented into five single words. Cutting was done according to the procedure of [Wagener et al., 1999c], i.e., the co-articulation was included at the end of the word but avoided at the beginning of the word, as shown in Figure 3.2. This procedure allowed a more natural sound, when it comes to resynthesize of new sentences. It was also important to cut the words at the zero beginning to avoid clicking/clipping sound: the word should start with 0°phase and end with 180°phase. This should avoid a clicking sound or any other undesired sound when generating new sentences. Smacks or any other sounds which did not belong to the pronounciation of the word were cut off. Words which due to cutting and editing were not good enough, were exchanged with another realization outside the 100 chosen sentences.

After cutting, 300 sentences were resynthesized. During the resynthesis, each word has to be followed by the word of the matched co-articulation. For example: the realization of word "membawa" (*brings*) recorded together with the word "lima" (*five*) can be only followed by the word "lima" and not by any other word. Within 300 sentences each word realization appeared three times. They were grouped into lists of 10 sentences (in total 30 test lists), where one test list included the whole speech material (50 words). Using MATLAB, overlaps of 0-37 ms (average: 1 ms) between the words were adjusted, so that the transition of one word to the neighboring word sounded natural. All new generated sentences were controlled by Sabine Hochmuth and Anna Warzybok.

In order to implement the sentences in the Software "Oldenburg Measurement Applications" (OMA) from Hörtech gGmbH, all stimuli (the generated sentences and noise) had to be converted into 16 bits.

3.2.3. Test-specific noise

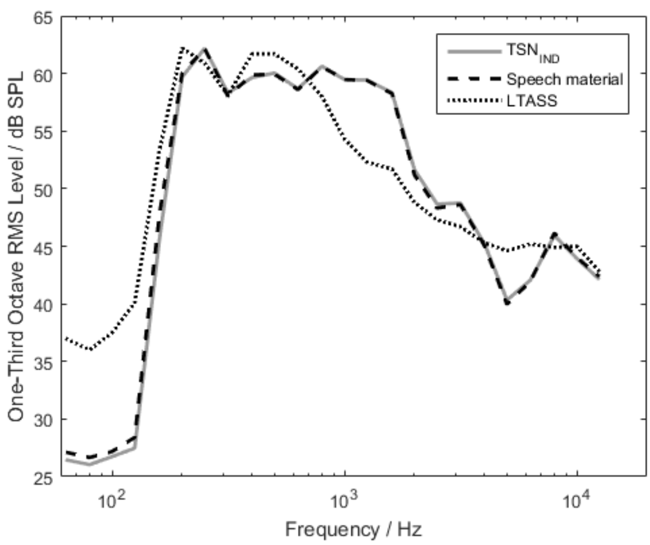


Figure 3.3.: Long-term averaged spectrum of the interfering noise (grey solid line) is very similar to the long-term averaged spectrum of the speech material of INDMatrix (dashed line), because the noise is generated by multiple superimposition of the speech material. The long-term average spectrum of the Indonesian female speaker is compared to female LTASS (dotted line) [Byrne et al., 1994].

The test-specific noise was generated, as recommended by [Kollmeier et al., 2015] and [Akeroyd et al., 2015], by randomized superimposing the speech material, with random initial delay and random delay of maximum 2 s between the sentence repetitions. As a result of the superimposition, a speech-shaped noise was generated. The long-term average spectrum matches the long-term average spectrum of the entire speech material (see Figure 3.3). The superimposition was performed 30 times, so that no strong fluctuation exists in the noise. The total duration of the noise was 3 minutes. In this way, an optimal spectral masking and a steep discrimination function can be achieved.

In comparison to the female long-term average speech spectrum (LTASS) [Byrne et al., 1994], the fundamental frequencies of the Indonesian female speaker resembled the fundamental frequencies of female LTASS. The RMS level of the middle frequencies (ca. 1-2 kHz) of the Indonesian female speaker is up to 10 dB higher than the corresponding RMS level of the female LTASS. In the low frequencies (ca. 60-100 Hz), however, about 10 dB weaker.

4. Optimization of Speech Material

The aim of the optimization procedure was to obtain highest homogeneity of the intelligibility among speech items. A reliable speech test is characterized by a high slope of test-specific function. The higher the slope, the more precisely SRTs can be estimated or speech intelligibility can be measured. The probabilistic model of Kollmeier [Kollmeier, 1990] describes the test-specific cumulative discrimination function S_{50test} as a convolution of the average item-specific intelligibility function (in case of matrix test: average word-specific speech intelligibility function) and the distribution of the word-specific SRTs, as shown in Figure 4.1.

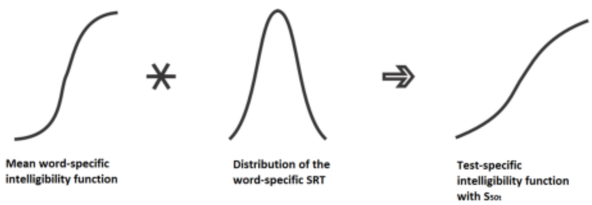


Figure 4.1.: Probabilistic model according to Kollmeier, 1990 [Kollmeier, 1990], to compute the slope of the test-specific intelligibility function.

By definition, there are two ways to obtain a steeper test-specific function:

1. **Change the word-specific slope,** for example, by changing the speaker. The higher the word-specific slope, the steeper the test-specific discrimination will be.

2. **Narrow the SRT distribution,** by applying level correction and adjusting the realization level. The broader the distribution of the word-specific SRT, the flatter the slope of the test-specific function becomes.

To simplify the calculation of the test-specific function, we assume that application of level correction shifts the SRT, but does not change the word-specific slope. By adjusting the RMS level of speech items, i.e., the RMS level of less-intelligible words will be increased and the sound level of better-intelligible words will be decreased, the standard deviation of the SRT distribution becomes smaller and as the result, the slope of the test-specific discrimination function becomes steeper.

According to the probabilistic model (Figure 4.3), the slope of the test-specific intelligibility function for the matrix test based on words of known word-specific intelligibility functions (SRT_{word} and S_{50word}) can be approximated by formula (4.1):

$$S_{50test} \approx \frac{S_{50median}}{\sqrt{1 + \frac{16 S_{50median}^2 \sigma_{SRT}^2}{(ln(2e^{1/2} - 1 + 2e^{1/4}))^2}}} \tag{4.1}$$

where :

$S_{50median}$: the median slope of the word-specific intelligibility functions,

σ_{SRT} : the standard deviation of the word-specific SRTs across the words.

The equation 4.1 is slightly different from the original equation, like described in [Warzybok et al., 2015b]. Due to the fact, that some realizations have a very steep slope, the median of the word-specific slope is considered instead of the mean of word-specific slope what leads to a better test-specific approximation.

The word-specific intelligibility function (SI_{word}) can be described using the logistic function (Equation 4.2):

$$SI_{word}(SNR) = \frac{1}{1 + exp(4 S_{50word}(SRT_{word} - SNR))} \tag{4.2}$$

where SRT_{word} indicates the word-specific SRT in dB SNR and S_{50word} indicates the slope at the SRT_{word} in dB^{-1} (where 1 stands for 100%).

Speech intelligibility measurement can be understood as a binomial experiment [Wesselkamp, 1994], where the number of correctly repeated words is counted as 1 and the number of incorrectly repeated words is counted as 0. The parameters SRT_{word} and S_{50word} are estimated in the equation 4.2 using the maximum-likelihood method. Thus, the discrimination function is adjusted to the data by varying the value of the slope and the SRT until the smallest standard deviation is reached [Brand and Kollmeier, 2002].

To characterize the intelligibility function of each word realization, measurements were executed at several fixed SNRs which covered 0 to 100% of speech intelligibility. The upper and lower boundary of SNRs were tested during a pilot test. Between both boundaries, several SNRs are determined. The stepsize should not be larger than $3dB$. According to Brand [Brand, 1998], at least 25 measured values between 20% and 50% of speech intelligibility and at least 25 measured values between 50% and 80% of speech intelligibility are necessary to obtain a reliable logistic function of each realization.

4.1. Participants

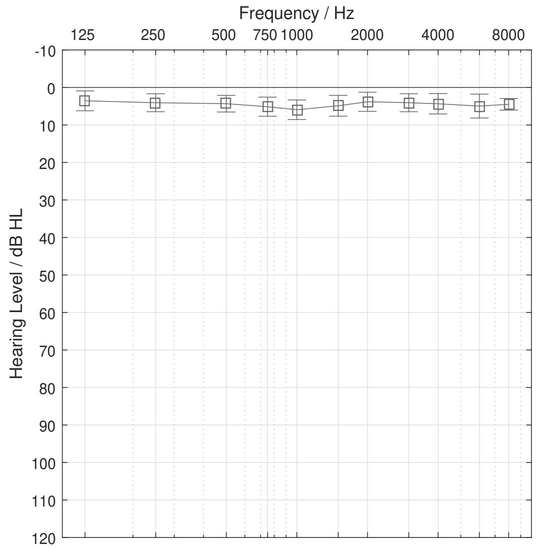


Figure 4.2.: Average of the pure-tone threshold of 20 listeners participated in the optimization measurement. Error bars indicate the standard deviations across listeners.

For the optimization measurement 20 listeners (7 male, 13 female, average age: 21.8 years) participated in the experiments conducted at the University of Oldenburg, Campus Wechloy. The following requirements were fulfilled by the subjects:

1. Normal-hearing native Indonesian speaker
2. Born and grown up in Indonesia
3. Not older than 30 years old
4. Finished education until at least senior high school (secondary school) in Indonesia
5. Living in Germany for less than 5 years.

All participants received payment for the measurements. The average pure-tone thresholds of the listeners is shown in Figure 4.2. All thresholds did not exceed 15 dB HL. Participants were also asked to fill in a questionnaire about their language profile (see Table 4.1).

There was no restriction of listener's place of origin, beside being born in Indonesia. As shown in Figure 4.3, most listeners were coincidently born in the island with the highest population density (Java).

Table 4.1.: Optimization measurement listeners' language profile(N=20)

Variable	Mean	SD	Range
Age (years)	21.8	2.4	18-28
Gender (male/female)			
Male	N=7		
Female	N=13		
Education			
Senior high school	N=17		
Diploma	N=0		
Bachelor	N=3		
Duration of residence in Germany (years)	2.6	1.6	0.6-5.1
Age of language acquistion			
Indonesian(N=20)	0.2	0.8	0-4.0
Javanese(N=5)	1.6	2.6	0-6.0
Sundanese(N=2)	6.5	0.7	6.0-7.0
English(N=2)	13.5	2.1	12.0-15.0
Khek(N=1)	0		0
Daily exposure of native languages before DE-residence (%)			
Indonesian(N=20)	86	16	50-100
Other languages(N=9)	20	15	2-50
Daily exposure of native languages during DE-residence (%)			
Indonesian(N=20)	56	21	20-90
Other languages(N=9)	7	6	0-20

Figure 4.3.: Demography of listeners' place of birth (N=20). The pins point to cities where listeners were born and the number on the pin represents the number of listeners born in the city. Maps: maps.google.com.

4.2. Material and methods

The optimization measurements were executed in a sound-attenuated booth fulfilled DIN ISO 8253-1/DIN EN 26189/ISO 6189 (the international standard for audiometry) in the Nessy Building, University of Oldenburg, Campus Wechloy.

The speech stimuli were presented monaurally over free-field equalized audiometric headphones (Sennheiser HDA 200) through a mobile audiometer with an integrated soundcard (ear 3.0 from AURITEC GmbH). The measurements were administered using the Oldenburg Measurement Application software from HörTech gGmbH. The measurement setup was calibrated to dB SPL with an artificial ear from Brüel&Kjær (B&K) Type 4153, a B&K 4134 0.5-inch microphone, a B&K 2669 preamplifier and a B&K 2610 measuring amplifier. The calibration of the system used the test-specific noise generated from the INDMatrix speech material. The speech signal and the test-specific noise were calibrated digitally to same overall level (80 dB SPL). An SNR of 0 dB means the noise and the speech signal have the same digital long-term RMS level. Breaks between sentences were not taken into account for the calculation.

All participants were firstly measured with pure-tone audiometry to confirm that the pure-tone threshold did not exceed 15 dB HL at third-octave frequencies from 125 to 8000 Hz.

For simplification and time efficiency, 30 test lists of 10 sentences were combined into 10 test lists of 30 sentences for the optimization measurements. However, the analyses were done for test lists of 10 sentences. Every participant was trained with two test lists: the first one was presented in quiet of a speech level of 60 dB SPL and the second one was presented in noise at a fixed SNR of -4 dB. The noise level was set to 65 dB SPL. The data from the training lists were discarded from the assessment.

Based on the pilot study with one listener and randomly selected lists, 0% and 100% of speech intelligibility was reached at -15 and 0 dB SNR, respectively, when measured at fixed SNR with noise level of 65 dB SPL. To ensure that all speech intelligibility from 0 to 100% is covered for all listeners and all speech items, the dB SNR-range was fixed from -20 to 2.5 dB SNR with a stepsize of 2.5 dB. Ten measurement points for every listener: -20, -17.5, -15, -12.5, -10, -7.5, -5, -2.5, 0, 2.5 dB SNR. Each listener was measured once at each SNR and each test list was presented only once to the listener. The test lists presented to one listener were randomized in such a way that after ten listeners every test list has been measured once at all 10 SNRs. This yielded to the same number of measurement point of every word realization.

All lists including training were executed using an open-set response format. The optimization measurement took about 1.5 hours per listener.

4.3. Results

The word-specific intelligibility function is plotted for every word realization. The psychometric curve is fitted to the distribution of the correct or incorrect answer. In Figure 4.4, the distribution of the correct answer is shown by the upside-down histogram and the distribution of the incorrect answer is shown by the (normal, upwards-directed) histogram. Psychometric curves of all realizations are shown

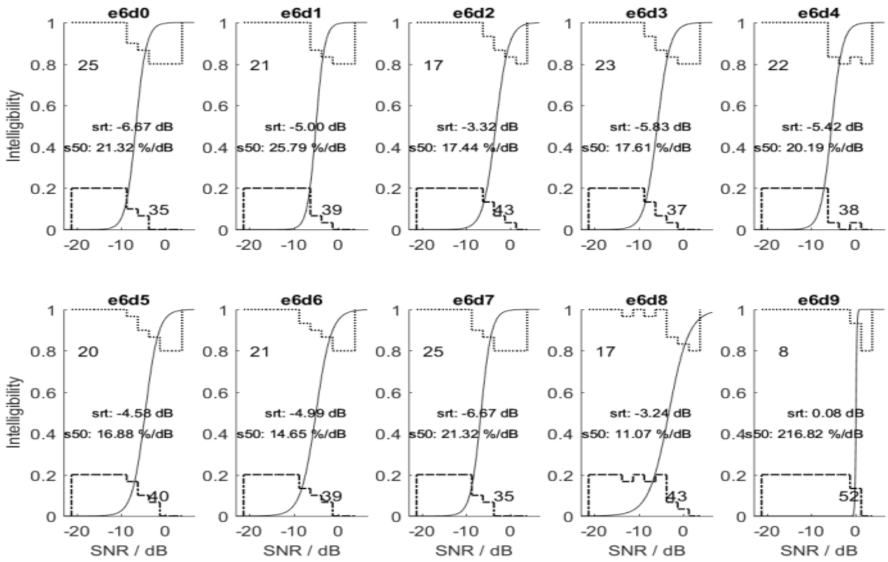


Figure 4.4.: Word-specific intelligibility function for all realizations of the word "kotor" (*eng:*dirty).

in Section E. As an example, all realizations of the word "kotor" are shown in Figure 4.4. For the word realization e6d9, an adequate fit of the psychometric curve is not possible. The SRT of e6d9 shows a large discrepancy to the median SRT across all realizations. Based on these two reasons, e6d9, as well as e8d1 and e8d8 were excluded from the calculation.

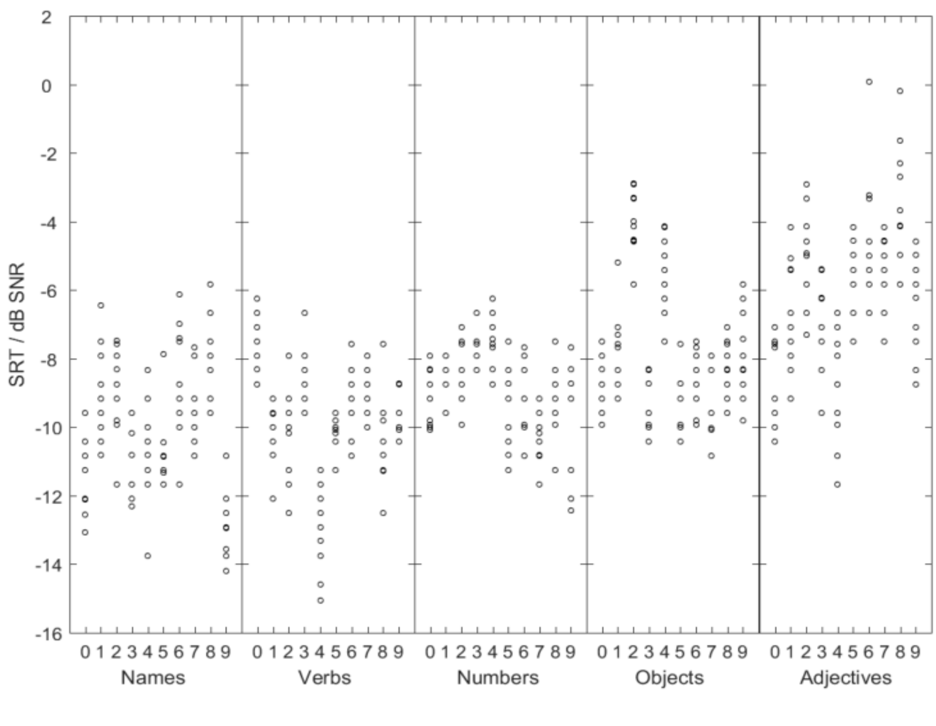


Figure 4.5.: SRT of every word realization obtained from the optimization measurement.

All test lists that contained the excluded word realizations were excluded from the final speech material. By that, 22 out of 30 test lists of 10 sentences with 485 word realizations remained. This resulted according to the probabilistic model [Kollmeier, 1990] in a test-specific SRT and slope of -8.6 dB SNR and 13.4%/dB, respectively.

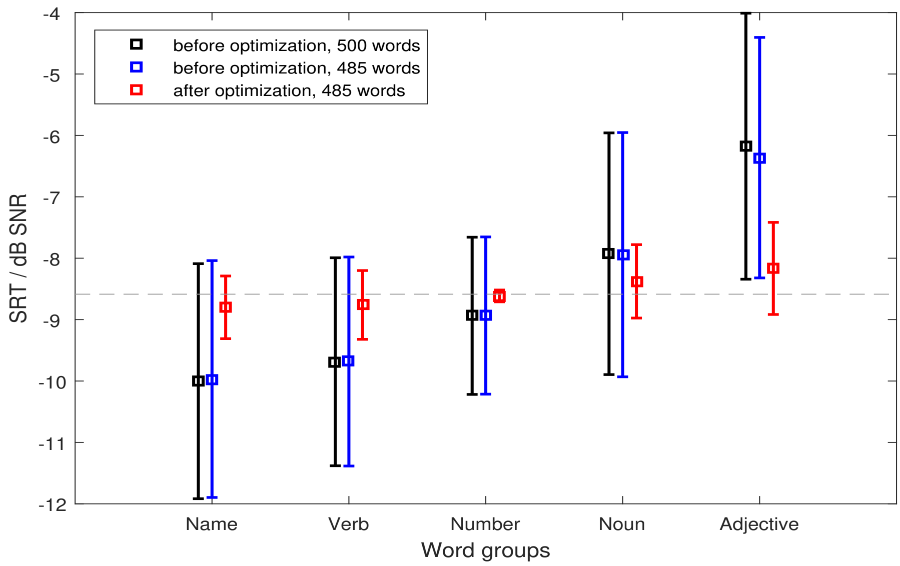


Figure 4.6.: Results of the optimization measurement presented as mean value of the SRTs of each word group. The error bars show the standard deviation of the SRT. Black represents SRT of the word group before the optimization including 500 realizations. Blue represents SRT of the word group after exclusions of some realizations. Red represents SRT of the word group after level adjustment with 485 remaining realizations.

The SRT of all word realizations are shown in Figure 4.5. The respective SRT is plotted against the corresponding representation of each word group. Figure 4.6 shows the mean value of the SRT of each word group with standard deviation (SD) before and after optimization (after application of level correction of ±3 dB). The adjectives are more less-intelligible than the other word groups and the names are the most intelligible. With increasing SRT from names to adjectives, there is a tendency that the sentences were pronounced gradually softer. The black squares and error bars represent the mean SRT and SD across all word realizations within a word group obtained from the optimization measurements. The blue squares and bars represent the mean SRT and SD of word groups after word exclusion, and the red squares and bars represent the mean SRT and SD of word groups after level correction limited to ±3 dB. The exact values from Figure 4.6 as well as the corresponding values for slope are summarized in Table 4.2.

Table 4.2.: SRT and slope of each word group before and after optimization.

Word Group	Before Optimization-500 words S_{50}[%/dB] Median	SD	SRT[dB SNR] Mean	SD	Before Optimization-485 words S_{50}[%/dB] Median	SD	SRT[dB SNR] Mean	SD	After Optimization-485 Words S_{50} [%/dB] Median	SD	SRT[dB SNR] Mean	SD
Names	15.8	103.7	-10.0	1.9	16.8	105.7	-10.0	1.9	16.8	105.7	-8.8	0.5
Verbs	21.3	130.6	-9.5	1.7	21.3	130.4	-9.7	1.7	21.3	130.4	-8.8	0.6
Numbers	21.3	145.2	-8.9	1.3	21.3	146.5	-8.9	1.3	21.3	146.5	-8.6	0.1
Nouns	21.3	158.8	-7.9	2.0	21.3	161.1	-7.9	2.0	21.3	161.1	-8.4	0.6
Adjectives	20.0	92.6	-6.2	2.2	18.8	90.0	-6.4	2.0	18.8	90.0	-8.2	0.8

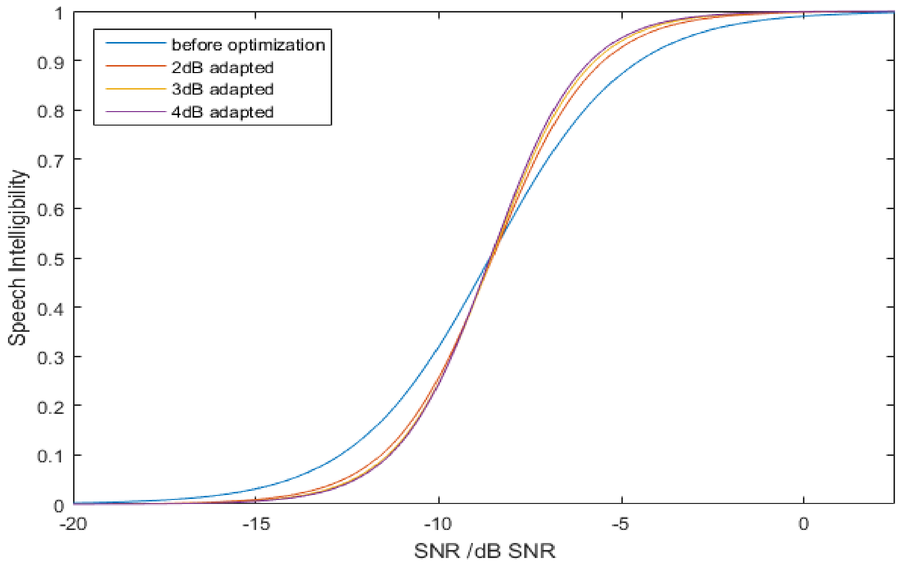


Figure 4.7.: Test-specific intelligibility function obtained from the optimization measurement (blue line). The predicted test-specific intelligibility function according to the probabilistic model [Kollmeier, 1990] with maximal adaptation/correction level of ±2dB (red line), ±3 dB (yellow), and ±4 dB (purple).

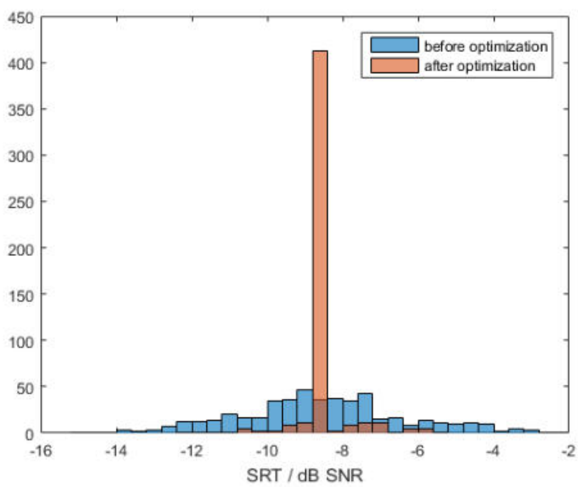


Figure 4.8.: Distribution of word-specific SRTs before and after optimization.

To increase the steepness of the test-specific intelligibility function and to optimize the homogeneity, the level of each word realization was adjusted towards the mean value of the SRT across speech items obtained from the optimization measurement (-8.6 dB). A maximal correction level (L_{corr}) of ±3 dB was chosen to obtain a high slope according to the probabilistic model and natural sound of the speech, at the same time. Figure 4.7 shows the predicted test-specific intelligibility function according to the probabilistic model [Kollmeier, 1990] with different maximal correction levels and in Table 4.3 the corresponding predicted test-specific SRTs and slopes.

Table 4.3.: Probability model [Kollmeier, 1990] to predicted test-specific SRT and slope after optimization.

Optimization	SRT [dB SNR]	S_{50test} [%/dB]
Before optimization	-8.6 ± 2.2	13.4
2 dB adaptation	-8.5 ± 1.0	18.0
3 dB adaptation	-8.6 ± 0.6	19.3
4 dB adaptation	-8.6 ± 0.3	20.0

The optimization of the speech material due to level adjustment decreased the standard deviation from 2.2 dB to 0.6 dB (see Figure 4.8), i.e., the SRT distribution bell curve becomes shallower. This yielded a steeper predicted test-specific intelligibility function with S_{50test} increased to 19.3%/dB (see Figure 4.7, yellow line).

5. Evaluation

The goals of the evaluation measurements were the establishment of reference SRT values for normal-hearing listeners, the verification of the properties of the optimized speech material, the assessment of test list equivalence, the examination of training effect, the investigation of the test-retest reliability, open and closed-set format, and the investigation of applicability of the test for measurements with bilingual listeners of the Indonesian language. Beside that the benefit of listening between the gaps was analyzed by measuring SRT in fluctuating noise.

In the previous section 4, the test-specific intelligibility function with the parameters SRT_{50p} of -8.5 dB SNR and S_{50p} of 19.3%/dB was predicted according to the probabilistic model of [Kollmeier, 1990]. If the correction level L_{corr} were determined and applied correctly, the remeasured test-specific intelligibility function with the optimized speech material should be equal or close to the predicted test-specific intelligibility function [Ozimek et al., 2010].

5.1. Participants

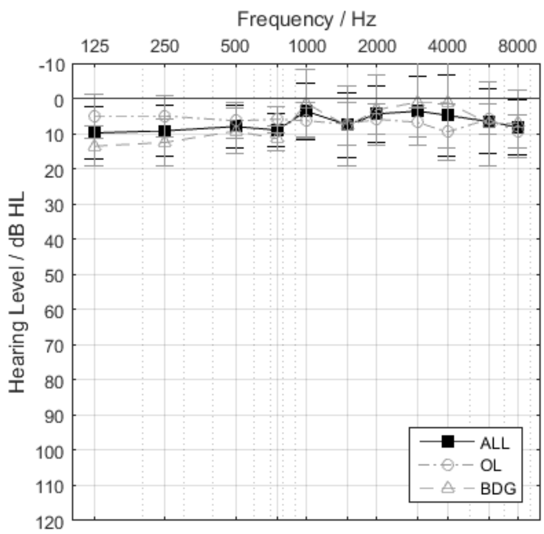


Figure 5.1.: Pure-tone threshold averaged across 25 listeners participating in the evaluation measurement (black squares). The average of pure-tone threshold is plotted separately for listeners in Oldenburg (gray circles) and listeners in Bandung (gray triangles).

The evaluation measurement involved the measurement sites: the University of Oldenburg and the A. Kasoem Hearing and Speech Center in Bandung. For the measurements, the A. Kasoem Hearing and Speech Center accepted the same ethical approval as used for measurements in Oldenburg. In total, 25 NH listeners participated in this measurement ($N_{Oldenburg}$=11, $N_{Bandung}$=14). Participants should fulfill the following requirements: normal-hearing Indonesian native speaker, born and grown up in Indonesia, not older than 32 years old, at least finished senior high school in Indonesia, less than 5-year long stay in Germany/other foreign countries. Due to the difficulty of finding Indonesian participants in Oldenburg, one participant who has stayed for 9 years in Germany was included in the measurement.

Eventhough the pure-tone audiometry setup in Oldenburg and Bandung was not the same, all participants were measured by the same instructor. The average of the pure-tone audiogram across all 25 participants is shown in Figure 5.1 (black line). Average pure-tone audiograms are also plotted separately for participants in Oldenburg (gray circles) and in Bandung (gray triangles). As in the optimization measurements, the participants were asked to fill in a questionnaire about their language profile (see Table 5.1). The average age of all listeners was 24.4 years (male: 14, female: 11). There was no restriction of participants' place of origin as shown in Figure 5.2. All participants received incentive according to cooperation partners' policy.

Figure 5.2.: Demography of evaluation measurement listeners' place of birth (N=25). The pins point to cities listeners were born and the number inside the pin represents the number of listeners born in the city. Image: maps.google.com.

5.2. Material and methods

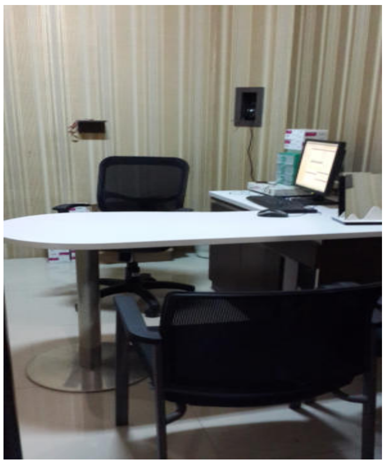
The evaluation measurements in Oldenburg were executed in the same sound-attenuated booth as during the optimization measurements. The measurements in Bandung were executed in two sound-attenuated booths, as shown in Figure

Table 5.1.: Evaluation measurement listeners' language profile (N=25)

Variable	Mean	SD	Range
Age (years)	24.4	2.7	20-32
Gender (male/female)			
Male	N=14		
Female	N=11		
Education			
Senior high school	N=14		
Bachelor	N=9		
Master	N=2		
Duration of residence in Germany, N=11 (years)	4.0	1.9	2.1-9.1
Age of native language acquistion			
Indonesian(N=25)	0.8	1.9	0-6.0
Javanese(N=3)	0	0	0
Sundanese(N=5)	1	2.2	0-5.0
Balinese(N=1)	0	-	0
English(N=1)	6	-	6
Manadonese(N=1)	0	-	0
Mandarin(N=1)	12	-	12
Japanese(N=1)	22	-	22
Karo(N=1)	6	-	6
Daily exposure of native languages (%)			
Indonesian(N=25)	80.2	24.6	10.0-100.0
Javanese(N=3)	31.7	7.6	25.0-40.0
Sundanese(N=5)	48.0	4.5	40.0-50.0
Other languages(N=6)	25.7	33.6	0-100.0

5.3. Due to the availability of the room, most of the measurement took place in booth 5.3a.

(a)



(b)

Figure 5.3.: Two rooms used for the measurements in the A. Kasoem Hearing and Speech Center, Bandung. Room 5.3a is used normally for auditory behavior observation (pure-tone audiometry with small children). Room 5.3b is used for hearing-aid consultation and fitting.

The measurement setup in Oldenburg was the same as for the previous measurements (see Section 4.2). For the measurements in Indonesia a mobile setup was used including a laptop with the software Oldenburg Measurement Application from Hörtech gGmbH, Sennheiser Headphones HDA200, and a mobile audiometer ear3.0 from AURITEC GmbH. Both setups were calibrated in the same manner as in the optimization measurements with the same calibration equipment and procedure.

For the evaluation measurements, 22 test lists of 10 sentences available after optimization were merged into 11 test lists of 20 sentences. In the test list equivalence measurement, each test list was adaptively measured in an open-set response format at two thresholds corresponding to 20% and 80% of speech intelligibility (SRT_{20} and SRT_{80} , respectively, pair of compromise [Brand and Kollmeier, 2002]). Each listener performed each test list at both threshold(22 test measurements each with 20 sentences) to estimate list-specific SRTs and S_{50}. To avoid fatigue the test list equivalence measurements were divided into 2 sessions on 2 different days. The difference between the first and the second session varied from 1 day to 1.5 weeks. At the beginning of the first session, every listener was measured with pure-tone audiometry to confirm that the pure-tone average (PTA) from the octaves between 0.5 and 4 kHz did not exceed 20 dB HL. All stimuli were presented monaurally on the better ear.

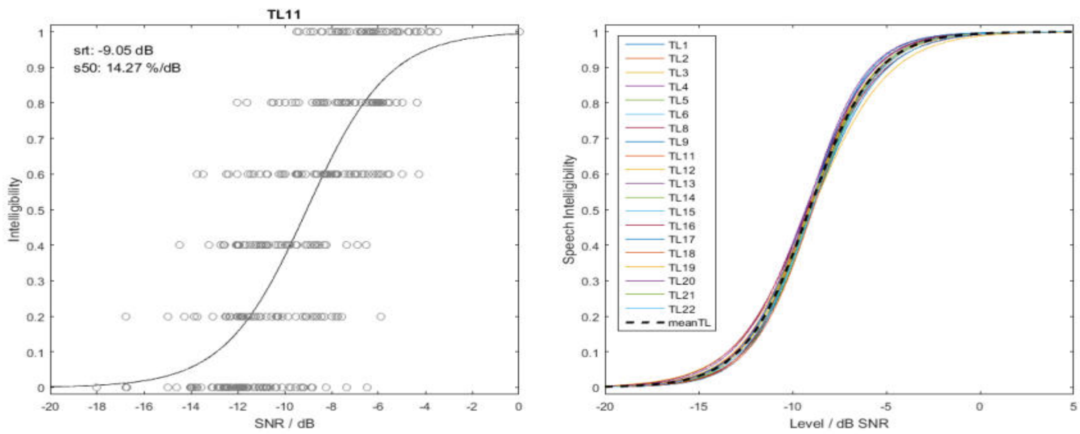
The noise level was set to 65 dB SPL for all listeners. After pure-tone audiometry, one training list was performed using the closed-set response format at a fixed SNR of 0 dB (Training 1). Afterwards, three test lists were measured using adaptive level control to reach 50% speech intelligibility with the open-set response format (Training 2, 3, and 4). Half of the test list equivalence measurements as described above (11 measurements in a randomized order) closed the first session.

The second session started with one training list in a closed-set response format at a fixed SNR of 0 dB (Training 5), continued with one list adaptively measured in an open-set response format (Training 6), one list adaptively measured in a closed-set response format, one list adaptively measured in quiet and one list adaptively measured in fluctuating noise (ICRA5-250 as used in Wagener et al. [Wagener et al., 2006]). Then the other half of the test list equivalence measurement was executed (11 measurements in randomized order).

5.3. Results

Test list equivalence and reference data

The analysis was done for lists of 10 sentences, i.e., 22 test lists of 10 sentences. As mentioned in the previous section, each test list was measured at 2 thresholds (SRT_{20} ans SRT_{80}). One sentence presented at a certain SNR was considered as one measurement point and its SI was obtained from the number of correctly repeated words in the respective sentence. This means, for each test list 20 measurement points were obtained for each listener. A reliable fit of list-specific intelligibility function could be obtained when 500 measurement points obtained from all 25 listeners were combined together (Figure 5.4a). List-specific SRTs and S_{50} are shown in Table 5.2. Test-specific SRT of -9.1 ± 0.2 dB SNR (SD across test lists) and test-specific slope of 14.4 ± 0.9 %/dB (SD across lists) were calculated as the average of the list-specific SRTs and slopes, respectively.



(a) (b)

Figure 5.4.: (5.4a) List-specific intelligibility function of test list (TL) no. 11 fitted to 500 points obtained from all listeners. (5.4b) List-specific intelligibility functions for all remaining test lists are represented by solid colored lines, test-specific intelligibility function is depicted by black dashed line.

In order to proof statistically that all test lists were equal in intelligibility, individual results were taken into account, i.e., a listener- and list-specific intelligibility function was fitted based on 20 points obtained from a given listener for each test list (see Figure 5.5). List-specific SRTs and slopes were averaged across the 25 individually fitted SRTs and slopes are shown in Figure 5.6. A one-way repeated measures ANOVA applied on these data revealed a significant effect of test lists ($F(21,504)=3.459$, $p<0.001$). According to pairwise comparisons (see Appendix F), test list no. 7 and no. 10 differed significantly from the remaining lists and were therefore excluded from the INDMatrix. The list-specific functions

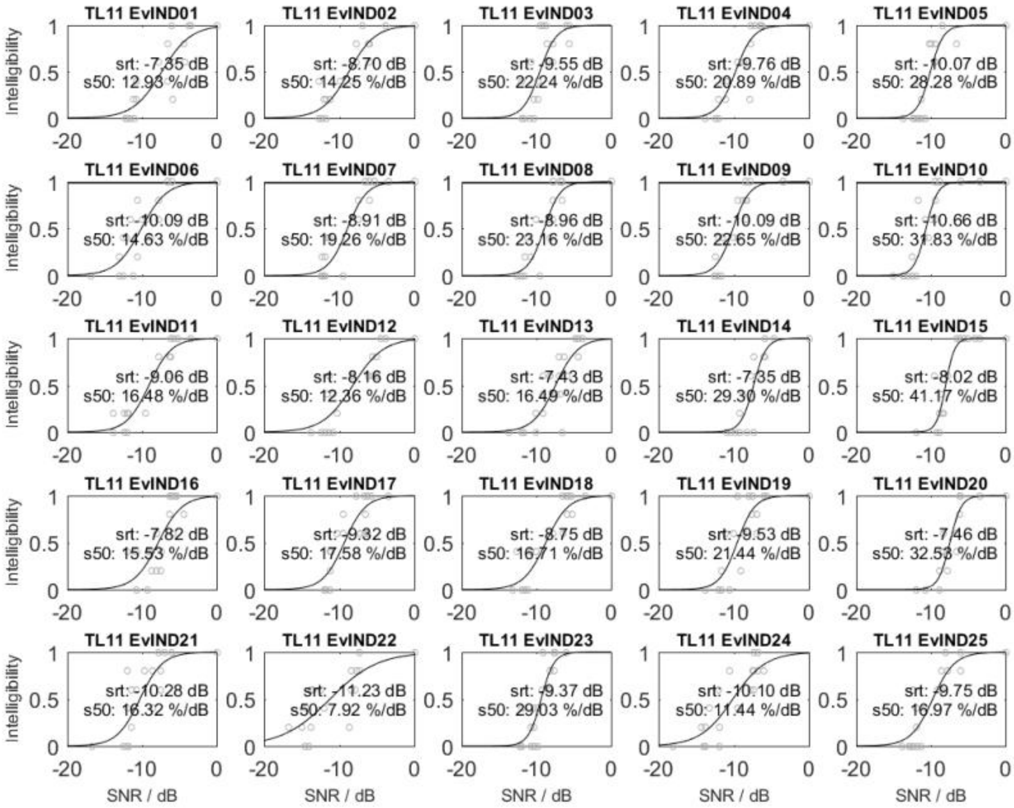


Figure 5.5.: An example of individually fitted list-specific discrimination functions (here from list no. 11). These functions were used for the statistical analysis.

of the 20 remaining test lists of 10 sentences are shown in Figure 5.4b. After exclusion of two test lists, the test-specific SRT remained the same, SD across test lists decreased from 0.2 to 0.1 dB. The new test-specific slope was 14.5±1.0 %/dB (SD across test lists).

Training effect

Average SRTs across listeners of the adaptively measured training lists are shown in Figure 5.7. Black squares represent the average SRTs measured in Oldenburg, gray circles represent the average SRTs measured in Bandung. Training 1 and 5 are not included into the plot due to presentation of these lists at a fixed SNR. Training 2 to training 4 were obtained during the first measurement session, while training 6 was obtained during the second session. In addition, the test-specific SRT is shown, which was obtained from the list-equivalence measurements (SRT_{eval}). In training 6, listeners were better trained due to 11 measurements of test list equivalence that were executed after training 4 in the first session. Average SRT of training 2 and 4 differ about 0.5 dB SNR. Between training 4 and

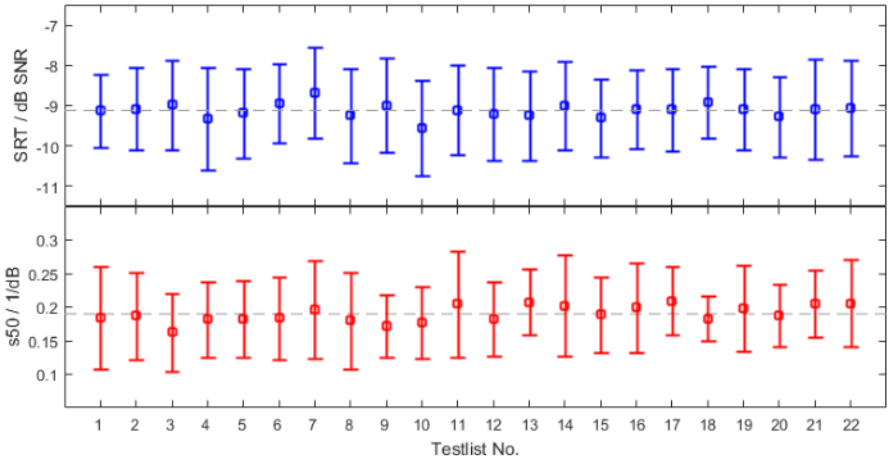


Figure 5.6.: Mean SRT (upper panel) and and slope S_{50} (lower panel) across listeners for each test list. Error bars indicate standard deviations across listeners.

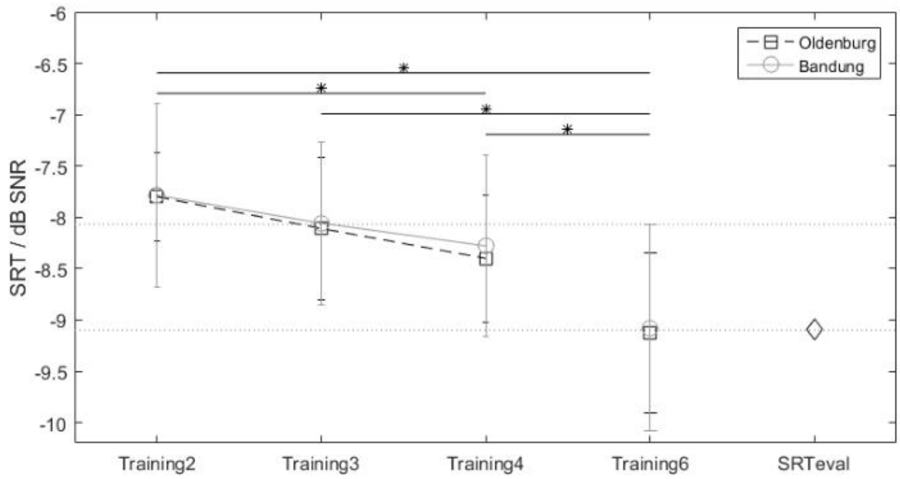


Figure 5.7.: Average SRTs of the training lists (Training 2-6 on the X-Axis) across listeners in Oldenburg (black squares) and in Bandung (gray circles) in comparison to the test-specific SRT obtained from the test list equivalence measurements SRT_{eval}

6, there is also a SRT difference of about 0.8 dB SNR. However, when listeners are well-trained, the SRT of the training list (see training 6) and SRT_{eval} are similar.

A two-way repeated measures ANOVA was performed on the adaptively measured training data with training as within-subject factor and measurement place as between-subjects factor. The ANOVA revealed a significant effect of training ($F(3,69) = 40.100$, $p<0.001$), but no interaction between training lists and measurement places ($p=0.979$) was found. No significant difference between measurements in Oldenburg and Bandung ($p=0.836$) was found in the measurement. To

Table 5.2.: Test list-specific SRTs and slopes S_{50} based on pooled SI data of all listeners for each list.

Test list no.	SRT [dB SNR]	S_{50} [%/dB]
1	-9.1	14.00
2	-9.1	13.93
3	-9.0	12.28
4	-9.3	13.69
5	-9.2	14.48
6	-8.9	14.42
7	-8.7	14.49
8	-9.2	12.96
9	-8.9	13.59
10	-9.6	13.50
11	-9.0	14.27
12	-9.1	13.90
13	-9.2	16.11
14	-8.9	14.86
15	-9.3	14.92
16	-9.1	15.58
17	-9.1	16.20
18	-8.9	15.38
19	-9.1	14.69
20	-9.3	14.70
21	-9.0	14.85
22	-9.0	14.60
Average	-9.1 ± 0.2	14.4 ± 0.9

determine which training lists differed in SRT, multiple pairwise comparisons were done with Bonferroni correction. Significant differences were found between training 2 and 4 ($p<0.001$), training 2 and 6 ($p<0.001$), training 3 and 6 ($p<0.001$), and training 4 and 6 ($p<0.001$).

A lot of lists would be required to be able to reach SRT_{eval} of -9.1 dB SNR. Therefore, the NH reference value for the adaptive measurements for clinical purpose of -8.2 ± 0.8 dB SNR (SD across listeners) was calculated as the average of training 3 and training 4.

Test-Retest Reliability

In Figure 5.8, SRT measured in noise (SRT_N) between the first (on X-Axis) and the second (on Y-Axis) session is compared. From the first session, the average SRT of training 3 and 4 is calculated for each listener and denoted as "SRT_N Test". From the second session, the SRT of training 6 is denoted as "SRT_N Retest". The gray solid line is a reference line with a slope of 1. The gray dashed line represents the bias/deviation of the correlated data of the reference line (0.9 dB). There is a significant correlation between SRT_N Test and SRT_N Retest

(R^2=0.51, p<0.0001).

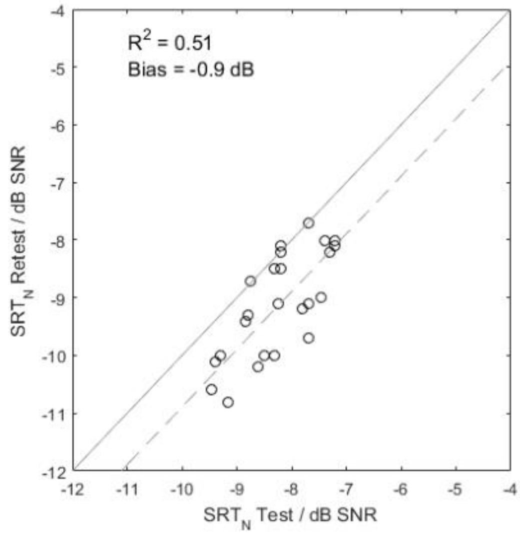


Figure 5.8.: SRT in noise for test (SRT_N Test) compared to retest SRT (SRT_N Retest).

Jansen et al. [Jansen et al., 2012] defined test-retest reliability as the root mean square of the within-subject standard deviations of repeatedly measured adaptive SRTs. In their study, all training lists were obtained within one session. According to this definition and avoiding the training effect, i.e., excluding the first two measurements, only training 3 and training 4 would be taken into account. The within-subject variability of 0.4 dB was found. By adding training 6 into the calculation (test-retest on different days), within-subject variability increased to 0.7 dB.

Open and closed-set response format

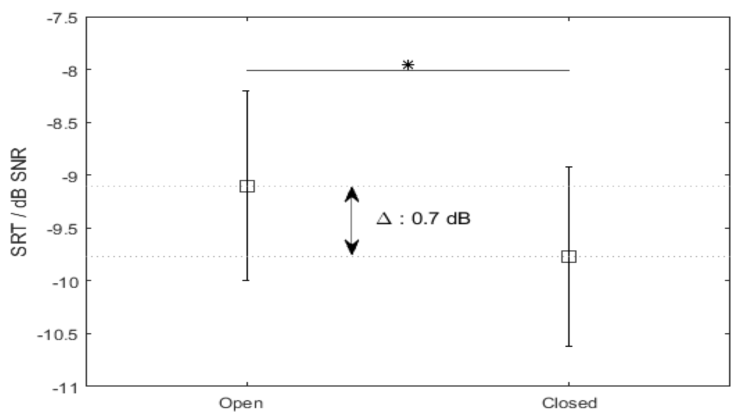


Figure 5.9.: Mean and SD of SRT for open and closed-set response format.

Figure 5.9 shows the average SRTs across listeners using the open-set (training 6) and closed-set response format. The error bars represent the corresponding standard deviation of the SRT. The average SRT was 0.7 dB higher using the open-set response format than when the closed-set response format was used. A one-way repeated measures ANOVA indicated a significant effect of response format ($F(1,24)=42.267$, $p<0.001$).

Mono- and bilingual listeners

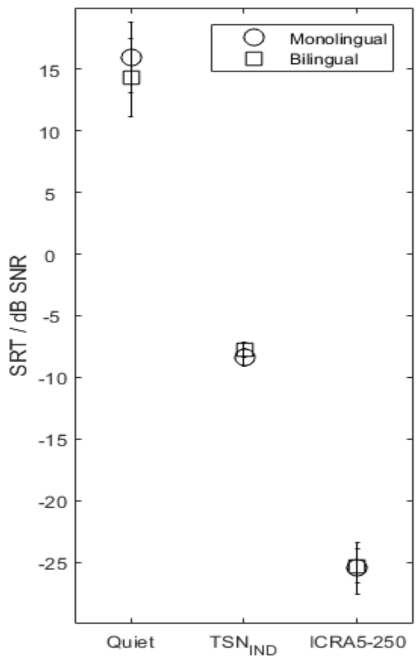


Figure 5.10.: The average and SD of SRT for speech intelligibility measurement in quiet, test-specific noise (TSN_{IND}), and modulated noise (ICRA5-250) for mono- (circles) and bilingual (squares) listeners.

In order to evaluate the applicability of the INDMatrix on bilingual listeners, listeners were divided into 2 groups based on the questionnaire: monolingual ($N_{monolingual}$=13) and bilingual ($N_{bilingual}$=12) listeners. Monolingual listeners were defined as those who reported to use Indonesian as native language only. Bilingual listeners were defined as those who reported to have another native language (regional or foreign) besides Indonesian. All bilingual listeners reported 6 years of age as latest age of acquisition of the Indonesian language (see Table 5.1).

Mean SRTs of speech intelligibility measurement in quiet, stationary noise (TSN_{IND}), and modulated noise (ICRA5-250) are presented in Figure 5.10. SRTs of monolingual listeners are indicated by circles and SRTs of bilingual listeners by squares. Error bars represent the standard deviations across listeners. A statistical analysis of SRTs measured in quiet, TSN_{IND} and ICRA5-250 using two-way repeated measures ANOVA revealed a significant effect of the SRTs ($F(2,46)=3554.270$, $p<0.001$). Multivariate test revealed no correlation between SRT and type of listeners (monolingual or bilingual, $F(2,22)=1.920$, $p=0.170$). Multiple pairwise comparisons with Bonferroni correction indicated no significant SRT difference between monolingual and bilingual listeners ($F(1,23)=0.217$, $p=0.646$).

6. Clinical Validation

The aim of the clinical validation was to validate the INDMatrix test with HI listeners, and compare their performance to pure-tone audiogram and the phonetically balanced word lists in terms of test specificity and sensitivity. Sensitivity is defined as the ability of the test to correctly identify patients with disease, while specificity defines the test ability to correctly identify patients without disease, in this case: hearing loss. Beside that, benefit of "listening between the gaps" measured in modulation noise (ICRA5-250) for HI listeners was observed.

6.1. Participants

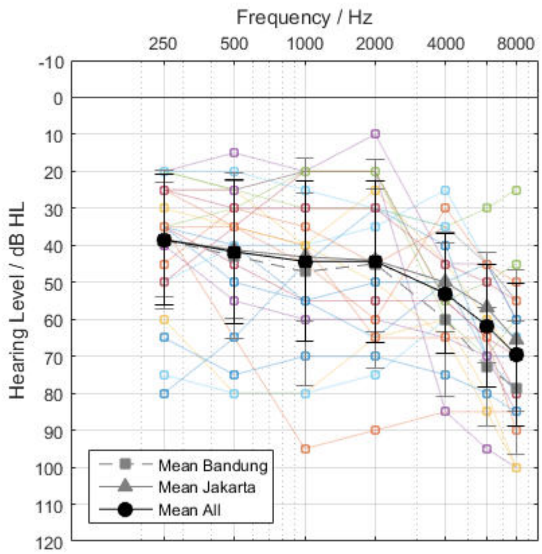


Figure 6.1.: Average thresholds of all validation measurement listeners are presented by black circles. Average thresholds of 7 listeners measured in Bandung are presented by gray squares and average thresholds of 16 listeners measured in Jakarta by gray triangles. Error bars represent standard deviations across listeners.

Twenty three patients with sensorineural hearing loss(SNHL) participated in the validation measurements in two citites of Indonesia ($N_{Bandung}$=7, $N_{Jakarta}$=16). All patients were older than 18 years. Complications besides SNHL, for example: tinnitus, otitis media, conductive hearing loss, etc. were allowed for the measurements. Patients with severe and profound HL were not included in the test because the OMA software only allowed noise signal presentation at maximum of 90 dB for stationary noise (TSN_{IND})and at maximum of 85 dB for modulated noise (ICRA5-250). Listeners were those who diagnosed with SNHL during the clinical ENT-diagnostics. Some listeners were invited by the clinical partners for this study. Pure-tone audiograms from all SNHL-listeners are shown in Figure 6.1. The thin solid colored lines represent the individual thresholds of the measured ear. The circles connected by a black solid line and the error bars represent the pure tone audiogram averaged across all listeners and the corresponding SD, respectively. The triangles connected by the gray solid line represent the average of listeners in Jakarta, the squares connected by the gray dashed line the average of listeners in Bandung. Pure-tone audiograms were measured by the audiometricians working at the measurement sites. The hearing level average from each meaurement site did not differ from the overall average. Listeners were asked to fulfill a questionnaire about their language profile and hearing loss history. Table 6.1 shows the listeners' profile for the validation measurement. All listeners confirmed that they have not had any ear surgery before. Figure 6.2 shows the distribution of the birth place of listeners participating in the validation study.

All subjects received incentive according to the policy from each measurement site.

Figure 6.2.: Demography of place of birth of listeners participating in the validation measurements (N=23). The pins point to cities listeners were born and the number inside the pin represents the number of listeners born in the responding city. Image: maps.google.com

6.2. Material and methods

The validation measurements took place in Bandung at the A. Kasoem Hearing and Speech Center and in Jakarta at the Cipto Mangunkusumo General Hospital. The ethical approval for the measurements at Cipto Mangunkusumo General Hospital can be seen in Appendix D. The validation measurements in

Table 6.1.: Profile of validation measurement listeners (N=23).

Variable	Mean	SD	Range
Age (years)	56.3	15.0	19-72
Gender (male/female)			
Male	N=11		
Female	N=12		
Education			
Elementary school	N=2		
Junior high school	N=1		
Senior high school	N=10		
Diploma	N=2		
Bachelor	N=6		
Master	N=1		
Doctoral	N=1		
Occupation			
Student	N=1		
Employee	N=4		
Housewife	N=6		
Teacher	N=1		
Retiree	N=6		
Entrepreneur	N=5		
Daily exposure of native languages (%)			
Indonesian(N=23)	69.8	23.0	25.0-100.0
Javanese(N=5)	31.0	18.2	5.0-50.0
Sundanese(N=6)	52.5	21.6	25.0-75.0
Bataknese(N=3)	26.7	20.8	10.0-50.0
Padangnese(N=2)	27.5	31.8	5.0-50.0
Other languages(N=4)	22.5	5.0	20.0-30.0
Age of hearing aids (years)			
Bilateral(N=1)	1	-	1
Right ear HA(N=4)	7.5	12.3	1.0-26.0
Left ear HA(N=0)	0		0

Bandung were executed in the same sound-attenuated booth as in the evaluation measurements (Figure 5.3a). In Jakarta, the measurements were executed in a sound-attenuated booth fulfilling DIN ISO 8253-1/ DIN EN 26189/ ISO 6189, as shown in Figure 6.3. The measurement setup was the same as explained in the evaluation measurements (Section 5.2).

For the validation measurement, lists of 20 sentences were used. Before measurements began, the individual thresholds were considered to set the noise level. For listeners with PTA (Pure-Tone Average: average threshold of hearing level at 0.5, 1, 2, and 4 kHz) smaller than 45 dB HL, the noise level was set to 65 dB SPL. In opposite to the validation of the German matrix test [Wardenga et al., 2015], noise level was individually increased for listeners, whose PTA was higher than 45 dB HL. The increased noise level was presented through the headphones to the listener before the training session to adjust the noise level to a comfortable

Figure 6.3.: Measurement room used in Cipto Mangunkusumo General Hospital. This room is normally used for balance tests and tympanometry.

level, if necessary. This was done to ensure that the sentences were masked by the test-specific noise.

The measurement began with two training lists. The first one was presented in a closed-set response format at fixed SNR of 2 dB and the second one was presented adaptively to reach 50% speech intelligibility (SRT). The measurement itself consisted of one test list presented in the test-specific noise, one test list presented in quiet and one test list presented in modulated noise (ICRA5-250, [Dreschler et al., 2001], [Wagener et al., 2006]) (see Section A.3). These three lists were presented adaptively to reach 50% speech intelligibility in an open-set response format. The validation measurement session (without pure-tone audiometry) took about 30 minutes. The order of measurements was the same for all listeners. However, the choice of the test lists was randomized. All listeners completed the speech intelligibility measurements in test-specific noise, but only 19 out of 23 listeners finished the tasks completely. Twenty subjects finished until speech intelligibility measurement in quiet.

For listeners who were not used to use computer, a laminated printed screenshot of the base matrix was handed out to the listener as a subtitute of a closed-set response format on a desktop. The listener's task was to repeat the presented sentence or understandable words to the instructor. The instructor faced the screen with the closed-format INDMatrix and selected the words repeated by the subjects. If the training list in closed-set response format was not possible, for example: due to vision problem, training list was executed in open-set response format.

Beside the measurements with INDMatrix, the SRT of other speech audiometric tests in quiet and in noise of each listener (if available and not older than 2 months) was noted in the validation measurement report (see Section A.3). Both measurement sites used the phonetically balanced word lists of Soewito (1985) for speech intelligibility measurements in quiet. No speech intelligibility measurements in noise were applied in the clinics at the time of the measurements.

6.3. Results

Correlation between INDMatrix and pure-tone audiometry

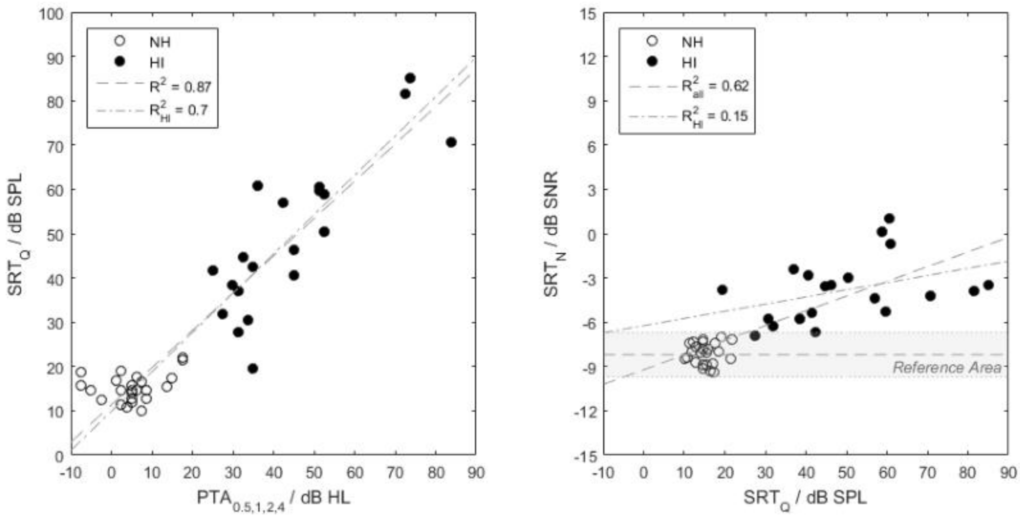


Figure 6.4.: The left panel represents SRTs measured in quiet (SRT_Q) as a function of the PTA of 4 octave-frequencies (0.5, 1, 2, and 4 kHz). The right panel represents SRTs measured in test-specific noise (SRT_N) as function of SRTs in quiet. The filled circles represent HI listeners. The unfilled circles represent NH listeners. The dashed lines indicate the regression line across all listeners (NH and HI) with the corresponding R^2. The diagonal dashed-dotted lines indicate the regression line across HI listeners with the corresponding R^2_{HI}. The reference area (filled area) refers to the average SRT (horizontal dashed line) of -8.2 dB SNR ± 2x SD across listeners (0.8 dB SNR) obtained from the evaluation measurements.

The left panel in Figure 6.4 represents the correlation between measured SRT in quiet (SRT_Q) and the pure-tone average (PTA) of 4 octave-frequencies (0.5, 1, 2, and 4 kHz). The filled circles represent HI listeners. The unfilled circles represent data of NH listeners. There was a strong and significant correlation found between speech intelligibility in quiet and the PTA ($R^2 = 0.87$) among all listeners. This correlation decreased to 0.7 when observing only HI listeners. The right panel of Figure 6.4 shows the correlation between measured SRT in the test-specific noise (SRT_N) and in quiet (SRT_Q). The reference area was established based on the evaluation measurement as mean SRT (-8.2 dB SNR, horizontal dashed line) ± 2x SD (0.8 dB SNR) among NH listeners (see Section 5.3, Training effect, for details). The correlation between measurements in noise and in quiet among all listeners was lower ($R^2 = 0.62$) than the correlation between PTA and measurement in quiet, but still significant ($p<0.0001$). Analyzing the data of HI listeners only, there was also no correlation between SRTs measured in noise and in quiet ($R^2 = 0.15$, p=0.0882).

Table 6.2.: Results of PTA, SRT measurements in quiet and in the test-specific noise using the INDMatrix for NH and HI listeners.

	NH listeners	HI listeners
Pure-tone Average / PTA [dB HL]	N=25	N=23
Average	5.2	45.9
SD	6.6	16.7
Range	-7.5 - 17.5	25 - 83.8
INDMatrix in quiet [dB SPL]	N=25	N=20
Average	15.2	49.3
SD	3.1	17.5
Range	10.1 - 21.9	19.5 - 85.2
INDMatrix in noise [dB SNR]	N=25	N=23
Average	-8.1	-3.4
SD	0.7	2.6
Range	-9.4 - -7	-6.9 - 4.2

Table 6.2 shows the results of PTA, INDMatrix measurements in quiet, and INDMatrix measurements in noise between NH and HI listeners. HI listeners showed higher averages (PTA and SRTs), larger SD, and wider range at all measurements than NH listeners.

Test sensitivity and specificity

In Figure 6.5, SRTs in noise (SRT_N) are shown as a function of PTA for all listeners. Unfilled circles refer to 25 NH listeners. Filled circles refer to 23 HI listeners. The horizontal dashed line represents the upper bound of the reference area (average SRT + 2x SD) equal to -6.7 dB (see Figure 6.4, right panel). The vertical dashed line refers to the maximal hearing threshold according to WHO's definition of normal hearing [WHO, 2017], i.e. 25 dB HL. Both the vertical and horizontal lines devide the plot into four areas. The ratio between filled circles in the top right area (N=21) and all HI listeners (N=23) points out the test sensitivity (91.3%). The ratio between unfilled dots (N=25) in the bottom left area and all NH listeners (N=25) points out the test specificity (100%).

Correlation between INDMatrix and phonetically balanced word lists

In Figure 6.6, SRTs in quiet (SRT_Q) measured with the INDMatrix are compared to SRTs in quiet (SRT_Q) measured with the phonetically balanced lists. The

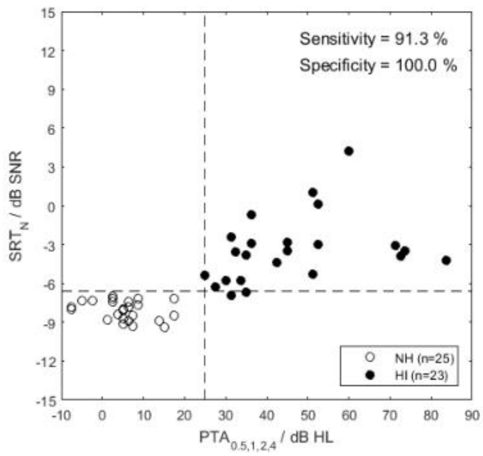


Figure 6.5.: SRT measurements in noise are plotted as a function of PTA to determine test sensitivity and specificity. Unfilled circles represent the NH listeners. Filled circles represent the HI listeners. The test specificity is calculated as the ratio of the number of unfilled circles in the bottom left area to the number of all unfilled circles. The test sensitivity is calculated as the ratio of the number of filled circles in the top right area to the number of all filled circles.

diagonal solid line is the reference line $y = x$ ($slope = 1$). If all points are on the diagonal, there is no difference between SRT in quiet measured with the INDMatrix and the phonetically balanced lists. A strong correlation ($R^2 = 0.93$) was found for SRTs in quiet measured with the INDMatrix and the phonetically balanced lists. A bias of 8.7 dB from the reference line (dashed line) indicated an 8.7 dB higher SRT in quiet measured with the INDMatrix than with the phonetically balanced word lists.

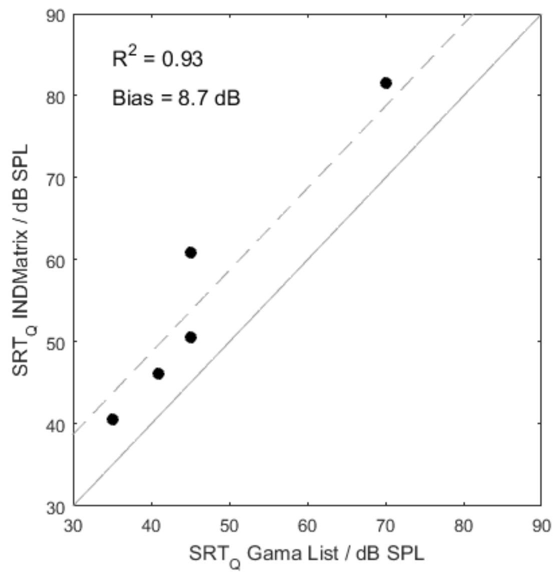


Figure 6.6.: Comparison of two speech intelligibility tests (INDMatrix and new phonetically balanced list) in quiet.

Modulation Benefit

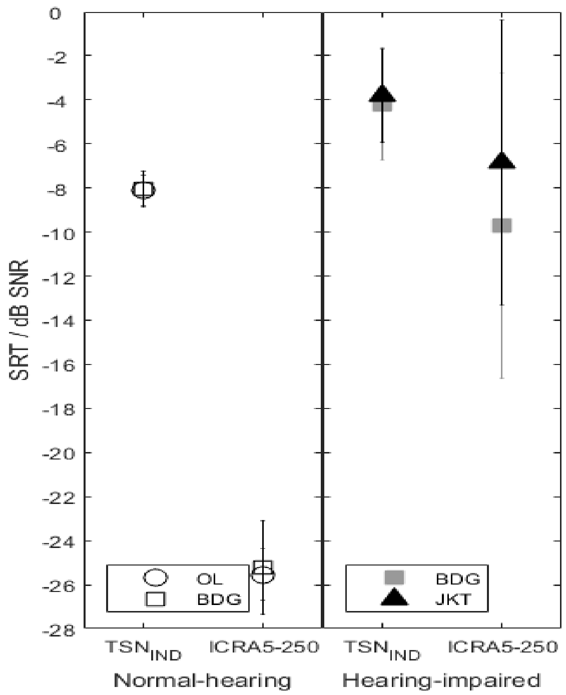


Figure 6.7.: Modulation benefit for NH listeners (left panel) and HI listeners (right panel). Average SRT of each measurement site is plotted separately.

In order to observe the modulation benefit, the average SRT measured in the stationary test-specific noise (TSN_{IND}) was compared to the average SRT measured in modulated noise (ICRA5-250). In Figure 6.7, the left panel shows average SRT of measurements in stationary and modulated noise for NH listeners, the right panel for HI listeners. Average SRTs are plotted separately for every measurement site. Circles represent the measurements in Oldenburg. Squares represent the measurements in Bandung. Triangles reprensent measurements in Jakarta. The SRTs measured in the stationary test-specific noise (TSN_{IND}) and in modulated noise (ICRA5-250) for NH listeners as well as the SRTs measured in the test-specific noise for HI listeners were very similar between measurement sites. The difference between SRT measurements in modulated noise for HI listeners in Bandung and Jakarta was larger. However, a one-way repeated measures ANOVA revealed no significant effect between SRTs and measurement sites (Bandung and Jakarta, $F(2,34)=0.955$, $p=0.395$). No significant SRT difference was found between the measurement sites (Bandung and Jakarta, $F(1,17)=0.940$, $p=0.346$).

The benefit in modulated noise compared to stationary noise was larger for NH listeners than for HI listeners (16.3 ± 1.3 dB SNR and 5.5 ± 3.3 dB SNR, respectively). SD across listeners in the modulated noise condition was larger

(SD_{all}:9.9 dB, SD_{NH}: 1.7 dB, SD_{HI}:6.6 dB) than in the stationary noise condition (SD_{all}:3.0 dB, SD_{NH}: 0.9 dB, SD_{HI}:2.2 dB). HI listeners' ability of "listening in the gaps" was not only worse than for NH listeners but also can be considered as unpredictable due to the large SD. A two-way repeated measures ANOVA was performed with the SRTs in stationary and modulated noise as within-subject factor and type of listeners (NH or HI) as between-subjects factor. The ANOVA proofed a significant effect of the SRT measurements (in stationary and modulated noise, $F(1,42)=348.954$, $p<0.001$) and a correlation between SRT measurements and listeners (NH and HI, $F(1,42)=134.242$, $p<0.001$). Pairwise comparison revealed a significant SRT difference for NH and HI listeners ($p<0.001$).

7. Discussion

The Indonesian matrix sentence test was introduced as a new speech recognition test similar to existing tests of the same structure in Swedish, German, Danish, Italian and Spanish [Hagerman, 1982], [Wagener et al., 1999c], [Wagener et al., 2003], [Puglisi et al., 2015], [Hochmuth et al., 2012]. In this chapter, the results of the evaluation and validation measurements will be discussed. Futhermore, the INDMatrix will be compared to the results of the existing matrix sentence tests and to the existing speech recognition test in noise available for the Indonesian language (the IndoHINT).

7.1. Evaluation

In comparison to the probabilistic model, the test-specific SRT (-9.1 dB SNR) obtained in the evaluation measurements was slightly lower than the predicted test-specific SRT (-8.6 dB SNR). This may be caused by several factors: (1) Two different group of listeners for the optimization and evaluation measurements; (2) The test-specific SRT was obtained in the test list-equivalence measurements that occured after extensive training sessions.

The optimization of the speech material resulted in an increase of the test-specific slope of 1.1%/dB (13.4%/dB before optimization, 14.5%/dB after optimization). The desired improvement of 5.9%/dB according to the probabilistic model could not be reached. One possible reason may be a different set of listeners was used in the optimization measurement compared to the evaluation measurement. This was also the case in other matrix tests, such as in [Wagener et al., 1999c], [Wagener et al., 1999a], [Wagener et al., 2003], and [Ozimek et al., 2010], but the differences between the test-specific slope after optimization and the predicted slope were small or can be neglected. Such a large difference between the test-specific slope and the predicted slope, however, was similar to the finding in Spanish matrix test [Hochmuth et al., 2012], where an increase of slope of only 2.2%/dB out of a desired improvement of 5.1%/dB was reached. Similar to Spanish matrix test [Hochmuth et al., 2012], the word group of the names were considered to be more intelligible than other word groups and therefore the names were attenuated in level. The unattenuated names during the optimization measurements may have triggered the listener's attention for the entire sentence at the time when the remaining words of the sentence were presented. Thus, reduction of names' level probably has resulted in a less dominant presentation of the names and reduced the role of temporal processing trigger. Nevertheless, the obtained test-specific slope is in the range of slopes of other matrix tests (see Figure 7.1).

In comparison to the intelligibility function of the IndoHINT with SRT_{test} of -5.8 dB SNR and S_{50test} of 10.0%/dB, the intelligibility function reached by the INDMatrix is steeper (14.5%/dB) and has a lower SRT (-9.1 dB SNR). This may occur due several factors: (1) the familiarity of the INDMatrix using 50 word-base matrix is higher than the familiarity of the IndoHINT using daily sentences; (2) The INDMatrix and the IndoHINT used different speakers. The steeper slope of the INDMatrix offers more precise diagnostics, because a small difference in SNR causes a larger difference in speech intelligibility.

7.1.1. Monolingual and bilingual listeners

There was no significant difference in SRT average found between monolingual and bilingual listeners/speakers. This indicates that the INDMatrix is applicable to every Indonesian, as long as Indonesian is one of the native languages. Other native language, either a regional or an international language, besides Indonesian seem not to influence the SRT.

7.1.2. Comparison to matrix tests developed for other languages

Test-list equivalence

Test lists which SRT deviated too much from the average list-specific SRT were excluded, so that no significant difference was found between test lists. After excluding two test lists, the evaluation measurements confirmed the equivalence of the test lists. The INDMatrix shows a very small difference between lists (0.1 dB) which is in line with the standard deviations across test lists of matrix tests established for other languages that usually range between 0.1 and 0.3 dB (see Table 7.1).

Even though the test-specific SRT was not obtained from measurements at fixed SNRs like in previous matrix test studies (e.g: [Wagener et al., 1999b], [Puglisi et al., 2015], [Hochmuth et al., 2012], [Jansen et al., 2012], [Warzybok et al., 2015b], etc.), the adaptively measured test-specific function (test-specific slope and SRT) of the INDMatrix was comparable to matrix tests in other languages (see Table 7.1 and Figure 7.1). Figure 7.1 shows that SRTs of the matrix test in other languages range between -10 and -6 dB SNR and the slopes range be-

Table 7.1.: Test-specific SRT and S_{50} of matrix tests established in other languages [Kollmeier et al., 2015].

Language	SRT obtained from measurements at fixed SNR [dB SNR]	S_{50} [%/dB]
Italian	-7.3 ± 0.2(SD across lists)	13.3 ± 1.2 (SD across lists)
	-7.4 ± 0.9 (SD across listeners)	14.3 ± 3.6 (SD across listeners)
Norwegian	-6.0 ± 0.8 (SD across listeners)	14.0 ± 3.4 (SD across listeners)
Polish	-9.6 ± 0.2 (SD across lists)	17.1 ± 1.6 (SD across lists)
Russian	-9.5 ± 0.2(SD across lists)	13.8 ± 1.6 (SD across lists)
	-9.5 ± 0.7 (SD across listeners)	14.0 ± 3.4 (SD across listeners)
Swedish	-8.1 ± 0.3 (SD across lists)	16.0 ± 3.4 (SD across lists)
Indonesian	-9.1 ± 0.1 (SD across lists)	14.5 ± 1.0 (SD across lists)
	-8.2 ± 0.8 (SD across listeners)	

tween 10.2 and 17.5%/dB. The intelligibility function of the Indonesian language represented by the dashed-dotted black line lies within this range.

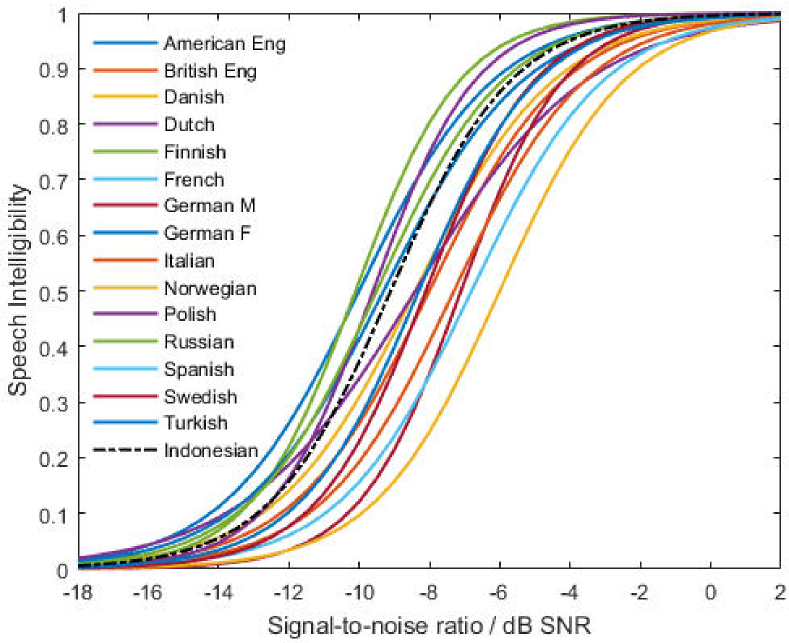


Figure 7.1.: Test-specific intelligibility function of matrix tests established for other languages

Training effect

The assessment of the training effect was conducted slightly different than in previous studies [Wagener et al., 1999b], [Warzybok et al., 2015b], [Hochmuth et al., 2012], [Ozimek et al., 2010], [Zokoll et al., 2015]. In this study, the training session in the evaluation measurements was always started with one test list in a closed-set response format at a fixed SNR of 0 dB, i.e., almost 100% intelligibility.

This way, an accelerated familiarity of the speech material was expected, with the perspective of reducing the necessary training measurements for the test. The significant difference found between Training 2 and 4, between Training 2 and 6, between Training 4 and 6 can be interpreted as follows: the familiarity to the speech material is important, but the familiarity to the adaptive procedure plays a role as well. In the Spanish test evaluation, an improvement of 0.8 dB from the first to the third training list was found, although the training session was completely measured with a closed-set response format and adaptive procedure [Hochmuth et al., 2012]. This indicates that one adaptively measured test list in a closed-set response format will not be sufficient for the training session. Therefore, the standard suggestion of 2 training lists in an open-set response format, firstly at an SNR of high intelligibility (for example: at 0 dB SNR) and then adaptively to reach 50% of speech intelligibility still offers an optimal solution [Kollmeier et al., 2015]. By conducting both training lists in an open-set response format, the difficulty of measuring listeners with visual impairment can be avoided.

Test-retest reliability

The test-retest reliability of the repeatedly measured adaptive SRTs within one session was 0.4 dB. The test-retest reliability increased to 0.7 dB by including SRTs from another test session. Nevertheless, the test-retest reliability of the adaptively measured INDMatrix for NH listeners of 0.7 dB is still high. This result is close to the test-retest reliability of matrix tests in other languages, such as in French, Italian, Russian, and German (see Table 7.2). The test-retest reliability for French, Italian, and Russian matrix test was calculated according to Jansen's definition of test-retest variability [Jansen et al., 2012]. For German matrix test, speech intelligibility was measured adaptively on different days using 30 sentences. In the clinical practice, the test-retest reliability refers to the consistency of the test results between 2 test sessions. In this case, conducting the INDMatrix with the same condition (same noise level, measured after a training session) and calibrated setup resulted in a more or less same SRT, even though it was measured in a different session. This occurs due to the familiarity of the speech material. The test-retest reliability of the INDMatrix is higher than the test-retest reliability of the IndoHINT score. The difference of the IndoHINT score between the first and second measurement ranged between -2 to 5 dB SNR depending on the tested test list. However, there was no significant difference between the first and second measurement detected for the IndoHINT ($p>0.05$, [Abiratno, 2016a]).

Table 7.2.: Comparison of test-retest reliability for normal-hearing listeners of matrix tests in different languages.

Language	Test-Retest Reliability [dB]	Reference
French	0.4	[Jansen et al., 2012]
Italian	0.5 (open-set)	[Puglisi et al., 2015]
	0.6 (closed-set)	
Russian	0.6 (open-set)	[Warzybok et al., 2015b]
	0.5 (closed-set)	
German	0.5 (normal-hearing)	[Wagener, 2004]
	0.7 (hearing-impaired)	

Open vs closed-set response format

The average SRT measured with the open-set response format was significantly higher than measured with the closed-set response format (0.7 dB). This means that the visual cue improved the SRT result. Significant differences in response format were also observedin other languages as shown in Table 7.3. However, no difference was found due to the visual cue in German and Polish matrix test [Brand et al., 2004], [Ozimek et al., 2010].

Table 7.3.: Comparison between SRTs measured in open-set and closed-set response format for matrix tests in different languages.

Language	Open-set [dB SNR]	Closed-set [dB SNR]	Reference
Indonesian	-9.1 ± 0.9	-9.8 ± 0.8	
Italian	-6.7 ± 0.7	-7.4 ± 0.8	[Puglisi et al., 2015]
Russian	-8.8 ± 0.8	-9.4 ± 0.8	[Warzybok et al., 2015b]
Spanish	-6.2 ± 0.8	-7.2 ± 0.7	[Hochmuth et al., 2012]
Turkish	-7.2 ± 0.8	-7.9 ± 0.8	[Zokoll et al., 2015]

7.2. Validation

Even though the PTA was strongly correlated with speech intelligibility measurements in quiet, speech intelligibility in quiet was not correlated with speech intelligibility in noise (correlation across all listeners $R^2 = 0.62, p < 0.0001$ and for hearing-impaired only $R^2 = 0.15, p = 0.0882$, see Figure 6.4). The correlation coefficients indicate that pure-tone audiometry can not predict the speech intelligibility in noise. Therefore, pure-tone audiometry alone is not enough to assess one's hearing deficit.

The correlation across HI listeners was lower than the correlation across all listeners, because of the broader distribution of HI listeners in comparison to NH listener. Broad distribution of the results of HI listeners reflected a poor temporal coding and selective auditory attention due to the loss of auditory nerve fibers.

No significant differences in SRT between the measurement sites indicates that the INDMatrix is also applicable everywhere, as long as the technical requirements are fulfilled.

7.2.1. Test sensitivity and specificity

With a high test sensitivity of 91.3% and specificity of 100%, the INDMatrix can distinguish between NH and HI quite precise. Nevertheless, these values should be considered as an initial evidence of a positive assessment of the test development. A larger group of patients is suggested to determine a more adequate test specificity and sensitivity.

The IndoHINT mentioned a test sensitivity of 84.0% and specificity of 77.4%, but the test sensitivity and specificity were defined to distinguish mild and moderate SNHL, instead of distinguishing NH and HI listeners [Abiratno, 2016a]. The validation measurements in this study were not designed to analyze mild and moderate SNHL separately. However, such comparisons with well-defined groups of patients could be an interesting topic for further studies.

7.2.2. Comparison to phonetically balanced word lists

Despite of the bias of 8.7 dB, a high correlation of speech intelligibility measurements in quiet between the INDMatrix and the phonetically balanced list was observed. However, the number of subjects that were measured with the phonetically balanced list was very low (n=5) due to financial limitations of diagnostics and treatment given by the state insurance. That means speech audiometry with the phonetically balanced word lists and pure-tone audiometry could not be performed at the same day. Only patients with private insurance or self-paying patients could receive all necessary diagnostics and treatment on the same day. Thus, more data is necessary to draw firm conclusions.

7.3. Technical notes

Technical difficulty could be faced during measurement with elderly HI patients who speak mostly regional languages. A test instructor who masters the regional language is suggested to -if necessary- explain the test instructions or answer patient's questions in the regional language. Apparently, the phonetically balanced lists seem viable for these patients. The words in the INDMatrix are easy and well-known. In this study, Indonesians who only speak regional languages were excluded from the study. However, investigating how SRT influences these listeners can be an interesting topic for further studies.

Furthermore, wrong choices of introduction words to the study caused patient's rejection of the study participation willingness. Words, such as "research" and "study", count as frightening and therefore should be avoided during the introduction and in the invitation or advertisement of the study.

In general, through the experience with HI patients in this study, it is suggested to use the open-set response format in Indonesia. By executing the open-set response format, measurement time can be kept short and the number of handled patients can be increased. Furthermore, a lot of elderly people in Indonesia are not used to use a computer, so that a closed-set format can be highly confusing and irritating. This would also help visual-impaired patients. However, the closed-set response format is conceivable to be applied for Indonesian in foreign countries and also to elderly people. Most Indonesians living in foreign countries are not illiterates and are better equipped in order to keep in touch with their families, relatives, or acquaintances with high-technology objects such as laptops, smartphones, or tablets.

8. Conclusions

The Indonesian matrix sentence test (INDMatrix) followed the international recommendations for the contruction of multilingual speech tests. The obtained results are comparable with matrix tests established for other languages. With a test-specific SRT of -9.1 ± 0.1 dB SNR (SD across test lists) and slope of 14.5 ± 1.0 %/dB, the INDMatrix shows a steep speech recognition function, which is in agreement with matrix tests developed for other languages. There are 20 lists of 10 sentences available for the final version of INDMatrix. No significant SRT differences between measurement sites indicate that as long as an adequate audiometry room is available and the measurement setup is calibrated, the INDMatrix is suitable for reliable measurements of speech recognition in quiet and in noise. Furthermore, the test is characterized by a high sensitivity of 91.3%, that is higher than the sensitivity of the Indonesian version of the Hearing in Noise Test (84.0%). The INDMatrix is also validated with hearing-impaired listeners and is applicable for monolingual, bilingual, and multilingual listeners with the same normative reference values, so that the INDMatrix can be applied for clinical purposes.

9. Acknowledgement

I would like express my gratitude to my supervisors: Prof. Dr. Dr. Birger Kollmeier and Dr. rer. nat. Anna Warzybok of Department of Medical Physics an Acoustics, University of Oldenburg / Cluster of Excellence Hearing4All for the useful comments, remarks, advice, encouragement, engagement through the learning process of this master thesis project and also for the financial support. Special thanks goes also to Dr. rer. nat. Sabine Hochmuth for knowledge sharing and all support on the way.

Furthermore, I would like to thank everyone who have helped me for this project: Anne Ivana Samanhudi, my high school friend, for being the speaker of INDMatrix, and Totok Suhardiyanto, Ph.D, Indonesian phonetician, for the precious time on checking the speech material. Also, I would like to acknowledge Dr. rer. nat. Daniel Berg from Hörtech gGmbH to have supported me in IT area and provided the INDMatrix in the OMA software.

Not forget to mention the cooperation partners in Indonesia. My sincere thanks goes to Dr. dr. Ratna Anggraeni, Sp.THT-KL(K), M.Kes. as the head of A. Kasoem Hearing and Speech Center in Bandung for the supervision during the project, for organizing rooms and subjects for the evaluation and validation measurement, to dr. Madhita Kasoem, M. Sc. who gave me the permission for the measurements in Bandung, and to the entire staff in A.Kasoem, Bandung. Also, I would like to thank the PIs from Cipto Mangunkusumo: dr. Ronny Suwento, Sp.THT-KL(K), Prof. Dr. dr. Jenny Bashiruddin, Sp.THT-KL(K), dr. Widayat Alviandi, Sp.THT-KL(K) for the support and supervision during the validation measurement, and the team in the neurootology division: Mujiono (audiometrician), dr. Guntur Surya, dr. Dita Meutia, Tatik and also Dr. dr. C.H. Soejono, SpPD, K-Ger, MEpid, FACP, FINASIM as president director and Dr. dr. Andri Maruli Tua Lubis, Sp.OT(K) from the research division who gave the permission to do the validation measurement in Cipto Mangunkusumo General Hospital.

I place on record, my biggest thank you to all participants in Oldenburg, Bandung, Jakarta to have willingly shared their precious time to take part in the measurements.

I am also grateful to all colleagues in the SPRAUD working group: Dr. rer. nat. Thomas Brand, Jana Müller, Christopher Hauth. I am extremely thankful and indebted to them for sharing ideas and experiences, for brain storming, and a lot of constructive inputs.

Finally, I must express my very profound gratitude to my parents, to my brother and to my boyfriend for the tremendous and indefatiguable support and endless encouragement throughout the years of my master study and through the entire process of researching and writing this thesis that helped me keeping on going and putting pieces together. This accomplishment would not have been possible without them.

A. Measurement reports

A.1. Optimization measurement report

Optimization Measurement Report
Indonesian Matrix Sentence Test

SubjectID : OptIND

Pat.ID	OptIND		
Age		Gender	
Date of measurement			
Place of measurement	W30-333 (Tür)		

List of levels (dB SNR)	
L1 :	-20
L2 :	-17,5
L3 :	-15
L4 :	-12,5
L5 :	-10
L6 :	-7,5
L7 :	-5
L8 :	-2,5
L9 :	0
L10 :	2,5

Certain remarks

____

____

____

Part 1 - Audiogram

- ☐ Available from (not older than 2 months):
- ☐ New audiogram
- ☐ Requirement fulfilled, better ear: R ☐ L ☐

Part 2 - Training lists (Better ear preferred)

Measurement	List No.	SNR	Speech Level	Noise Level	Intelligibility in %
Training 1	TL4	60 dB	60 dB	0	
Training 2	TL2	-4 dB	61 dB	65 dB	

Part 3- Test lists at fixed SNR (Better ear preferred)

Measurement	List No.	SNR	Speech Level (dB)	Noise Level (dB)	Intelligibility in %
1	TL9	L2	47,5	65	
2	TL2	L9	65	65	
3	TL8	L7	60	65	
4	TL1	L3	50	65	
5	TL5	L5	55	65	
6	TL7	L1	45	65	
7	TL3	L4	52,5	65	
8	TL10	L8	62,5	65	
9	TL4	L10	67,5	65	
10	TL6	L6	57,5	65	

Messdatum : ____________

Fragebogen zum Indonesischen Matrixtest

Adaptiert von LEAP-Q (Marian, Blumenfeld & Kaushanskaya, 2007)

Pat.ID : ____________ Age : ____ M / W

1. Wo sind Sie geboren (Stadt)? ____________
2. Wo sind Sie aufgewachsen (Stadt, falls mehrere bitte auch die Zeitangabe schreiben)?

   ____

   ____
3. Was ist Ihr höchster Bildungsgrad?

   ☐ Grundschule (SD) ☐ Mittelschule (SMP) ☐ Oberschule (SMA) ☐ Diploma (D3)

   ☐ Bachelor (S1) ☐ Master/Diplom (S2) ☐ Doktor (S3) ☐ Staatsexamen
4. Sprechen Sie andere Muttersprachen außer Indonesisch? ☐ Nein ☐ Ja, ____________
5. Wann haben Sie Ihre Muttersprache/n gelernt?
   a. Indonesisch, seit ____________
   b. ____________ , seit ____________
   c. ____________ , seit ____________
6. Zeitpunkt des Zuzugs nach Deutschland (MM/YYYY) ____________

   Wenn Sie in einem anderen Land gelebt haben, wo, wann und wie lange haben Sie dort gelebt?

   ____
7. Wie oft sprechen Sie **vor Ihrem Aufenthalt in DE** Ihre Muttersprache/n?
   a. Indonesisch : ________ %
   b. ____________ : ________ %
   c. ____________ : ________ %
8. Wie oft sprechen Sie **derzeit** Ihre Muttersprache/n?
   a. Indonesisch : ________ %
   b. ____________ : ________ %
   c. ____________ : ________ %

A.2. Evaluation measurement report

SubjectID : EvIND

Evaluation Measurement Report
Indonesian Matrix Sentence Test

Noise level:	65 dB SPL

Pat.ID	EvIND		
Age		Gender	
Date of measurement			
Place of measurement			

Certain remarks

Part 1 - Audiogram

☐ Available from (not older than 2 months):
☐ New audiogram
☐ Requirement fulfilled, better ear: R ☐ L ☐

Part 2 - Training lists (Better ear preferred)

Measurement	List No.	Procedure	Set Response	Intelligibility / %	SRT / dB SNR
Training 1		constant SNR	closed		0
Training 2		adaptive	open	50	
Training 3		adaptive	open	50	
Training 4		adaptive	open	50	

Part 3- Evaluation measurement Day 1

Measurement	List No.	Procedure	Set Response	Intelligibility / %	SRT / dB SNR
1		adaptive	open	20	
2		adaptive	open	80	
3		adaptive	open	20	
4		adaptive	open	80	
5		adaptive	open	20	
6		adaptive	open	80	
7		adaptive	open	20	
8		adaptive	open	80	
9		adaptive	open	20	
10		adaptive	open	80	
11		adaptive	open	20	

SubjectID : EvIND

Evaluation Measurement Report
Indonesian Matrix Sentence Test

Noise level: 65 dB SPL

Pat.ID	EvIND		
Age		Gender	
Date of measurement			
Place of measurement			

Certain remarks

Part 1 - Audiogram

☒ Available from (not older than 2 months):

☐ Requirement fulfilled, better ear: R ☐ L ☐

Part 2 - Training lists + Extra measurements (Better ear preferred)

Measurement	List No.	Procedure	Set Response	Intelligibility / %	SRT / dB SNR
Training 1		constant SNR	closed		0
Training 2		adaptive	open	50	
Training 3		adaptive	closed	50	
Quiet		adaptive	open	50	
ICRA 5-250		adaptive	open	50	

Part 3- Evaluation measurement Day 2

Measurement	List No.	Procedure	Set Response	Intelligibility / %	SRT / dB SNR
1		adaptive	open	80	
2		adaptive	open	20	
3		adaptive	open	80	
4		adaptive	open	20	
5		adaptive	open	80	
6		adaptive	open	20	
7		adaptive	open	80	
8		adaptive	open	20	
9		adaptive	open	80	
10		adaptive	open	20	
11		adaptive	open	80	

Tanggal pemeriksaan : _______________ Tempat pemeriksaan: ____________
Measurement date *Measurement place*

Kuesioner untuk Evaluasi Tes Matrix Bahasa Indonesia

Questionnaire for Evaluation of the Indonesian Matrix Sentence Test

Kode pasien : ______________________________ Usia : _______ L / P
Patient-ID *Age*

1. Kota kelahiran : ________________________
   Place ofbirth
2. Kota di mana Anda tumbuh (jika lebih dari satu, tolong sertakan tahun tinggal):
   Place where you grew up (in case of more than one city, please write down also the time information of residence)

   ____

   ____
3. Pendidikan tertinggi :
   Highest education level

   ☐ SD *Elementary school* ☐ SMP *Junior high school* ☐ SMA *Senior high school* ☐ D3 *Diploma*

   ☐ S1*Bachelor* ☐ S2 *Master* ☐ S3 *Doctorate*
4. Apakah ada bahasa ibu lain (bahasa daerah / bahasa asing) selain Bahasa Indonesia?
   Are there any other native languages (regional or foreign languages) beside Indonesian?

   ☐ Tidak ☐ Ya:____________________
   No *Yes*
5. Kapan Anda mulai mengenal / belajar bahasa ibu – bahasa ibu tersebut?
   When did you first know/learn these native languages?
   a. Bahasa Indonesia, sejak ___________
   b. Bahasa _______________,sejak ___________
   c. Bahasa _______________ ,sejak ___________
6. Bagaimana persentasi penggunaan bahasa ibu – bahasa ibu tersebut sehari-hari?
   How often do you use these native languages in daily life?
   a. Bahasa Indonesia : ___________ %
   b. Bahasa ________ : ___________ %
   c. Bahasa ________ : ___________ %

A.3. Validation measurement report

Validation Measurement Report
Indonesian Matrix Sentence Test

SubjectID : ValIND

Noise level: 65 dB SPL

Pat.ID	ValIND		
Age		Gender	
Date of measurement			
Place of measurement			

Certain remarks

____

____

Part 1 - Audiogram

☐ Available from (not older than 2 months):

☐ Measured ear: R ☐ L ☐

Part 2 - Other available diagnostics

☐ Questionnaire

☐ Speech audiometry in quiet ________ , SRT: ________

☐ Speech audiometry in noise ________ , SRT: ________

Part 3 - Measurement lists

Measurement	List No.	Procedure	Set Response	Intelligibility / %	SNR/dB
Training 1	TL5	constant SNR	closed		2
Measurement 1	TL6	adaptive	open	50	
Measurement 2	TL7	adaptive	open	50	
Quiet	TL8	adaptive	open	50	
ICRA5	TL9	adaptive	open	50	

Tanggal pemeriksaan : _______________ Tempat pemeriksaan: ____________
Measurement date *Measurement place*

Kuesioner untuk Validasi Klinis Tes Matrix Bahasa Indonesia

Questionnaire for clinical validation of the Indonesian Matrix Sentence Test

Kode pasien : ______________________________ Usia : _______ L / P
Patient-ID *Age*

1. Kota kelahiran : ________________________
   Place ofbirth
2. Kota di mana Anda tumbuh (jika lebih dari satu, tolong sertakan tahun tinggal):
   Place where you grew up (in case of more than one city, please write down also the time information of residence)
   ____
   ____
3. Pendidikan tertinggi :
   Highest education level
   ☐ SD *Elementary school* ☐ SMP *Junior high school* ☐ SMA *Senior high school* ☐ D3 *Diploma*
   ☐ S1 *Bachelor* ☐ S2 *Master* ☐ S3 *Doctorate*
4. Pekerjaan : ____________________________
   Occupation
5. Apakah ada bahasa ibu lain (bahasa daerah / bahasa asing) selain Bahasa Indonesia?
   Are there any other native languages (regional or foreign languages) beside Indonesian?
   ☐ Tidak *No* ☐ Ya: ____________________ *Yes*
6. Bagaimana persentasi penggunaan bahasa ibu – bahasa ibu tersebut sehari-hari?
   How often do you use these native languages in daily life?
   a. Bahasa Indonesia : ___________ %
   b. Bahasa ________ : ___________ %
   c. Bahasa ________ : ___________ %
7. Apakah Anda menggunakan alat bantu dengar? ☐ Tidak *No* ☐ Ya *Yes*
   Are you using hearing aid(s)?
   Jika ya: ☐ Telinga kanan, sejak_________ ☐ Telinga kiri, sejak ___________
   If yes *Right ear, since* *Left ear, since*
8. Apakah Anda menderita tinitus (suara berdenging/berdesir)?
   Are you suffering from tinnitus?
   ☐ Tidak *No* ☐ Ya: *Yes* ☐ Telinga kanan *Right ear* ☐ Telinga kiri *Left ear*
9. Apakah Anda pernah menjalankan operasi telinga? ☐ Tidak *No* ☐ Ya *Yes*
   Have you ever had any ear surgery?
   Jika ya: *If yes* ☐ Telinga kanan, sebab ____
   Right ear, because
   ☐ Telinga kiri, sebab ____
   Left ear, because

B. Research banner

Registration number: THT-PKRS/002/rev00/2016/Br

C. Letter of intent

CARL VON OSSIETZKY universität OLDENBURG
FAKULTÄT VI
MEDIZIN UND GESUNDHEITSWISSENSCHAFTEN

CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG · 26111 OLDENBURG

Dr. dr. Ratna Anggraeni, Sp.THT-KL(K)., M.Kes
Dr. Madhita Hatta Kasoem, M.Sc
PT. Kasoem Hearing and Speech Center
Jl.Cikini Raya no.18
Jakarta Pusat

Prof. Dr. rer. nat. Dr. med.
Birger Kollmeier
Abteilung Medizinische Physik
Department für Medizinische Physik und Akustik

TELEFONDURCHWAHL
+49 (0)441 798-5466
Sekretariat +49 (0)441 798-5470
Fax +49 (0)441 798-3902
EMAIL
birger.kollmeier@uni-oldenburg.de
INTERNET
http://medi.uni-oldenburg.de

OLDENBURG, 4. Dezember 2016

POSTANSCHRIFT
Universität Oldenburg
Medizinische Physik
D-26111 Oldenburg

PAKETANSCHRIFT
Ammerländer Heerstraße 114 - 118
D-26129 Oldenburg

Cooperation project "Development of the Indonesian Matrix sentence test"

Dear Dr. dr. Ratna Anggraeni and Dr.Madhita Hatta Kasoem,

I appreciate your willingness to cooperate with us on the development of the Indonesian Matrix test. With this letter, I would like to confirm the conditions of our intended collaboration in this validation project. If you agree on the conditions stated below, please confirm by signing a copy of this letter and sending it back to me. Thank you very much in advance!

The following conditions will apply for our collaboration:

Each partner within this cooperation will pay salaries and other expenses for their own personnel. Each partner has the right to publish the contents and results of the current cooperation, but has to mention the contribution of the respective other partners in an appropriate way. The University Oldenburg will cover all expenses related to the recordings and have the ownership rights for the audio speech material. Kasoem Hearing and Speech Center represented by Dr. dr. Anggraeni and Dr. Kasoem has the nonexclusive, nontransferable rights for using the test methods free of charge for the duration of the cooperation project "Development of the Indonesian Matrix sentence test". Subsequently, HörTech gGmbH will provide one free, unrestricted software license for the Oldenburg Measurement Applications software including the Indonesian Matrix sentence test module (including possible future updates) to the Kasoem Hearing and Speech Center represented by Dr. dr. Anggraeni and Dr. Kasoem. HörTech gGmbH intends to market the Indonesian Matrix sentence test software and the audio test material. Dr. dr. Anggraeni and Dr. Kasoem's team at Kasoem Hearing and Speech Center will support this marketing effort, e.g. by referencing the HörTech software in publications on studies where this software was used. Both parties agree to exchange scientific information and results, to encourage the exchange of specialists and to support the transfer of scientific and technological achievements, to initiate the organization of scientific conferences and symposiums, and to simulate the publication of joint study results. Both parties agree that Dr. dr. Anggraeni and Dr. Kasoem's team should actively participate in the efforts to establish the matrix test and its distribution in Mainland Indonesia while The University Oldenburg / HörTech gGmbH lead the distribution of the matrix test for the rest of the world.

December 4, 2016, Oldenburg

Date, place (Signature Birger Kollmeier, Universität Oldenburg)

December 13-2016, Bandung

Date, place (Signature Ratna Anggraeni, Kasoem Hearing and Speech Center)

December 22, 2016 Jakarta

Date, place (Signature Madhita H. Kasoem, Kasoem Hearing and Speech Center)

CARL VON OSSIETZKY universität OLDENBURG
FAKULTÄT VI
MEDIZIN UND
GESUNDHEITSWISSENSCHAFTEN

CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG · 26111 OLDENBURG

dr.Ronny Suwento, Sp.THT-KL(K), Prof.Dr.dr.Yenni Bashiriddin, Sp.THT-KL(K),
and dr.Widayat Alviandi, Sp.THT-KL(K)
Department of Ear,Nose,Throat- Head and Neck Surgery
RSUPN Dr. Cipto Mangunkusumo
Jl. Diponegoro No.71
Jakarta Pusat, Indonesia

Prof. Dr. rer. nat. Dr. med.
Birger Kollmeier
Abteilung Medizinische Physik
Department für Medizinische Physik und Akustik

TELEFONDURCHWAHL
+49 (0)441 798-5466
Sekretariat +49 (0)441 798-5470
Fax +49 (0)441 798-3902
EMAIL
birger.kollmeier@uni-oldenburg.de
INTERNET
http://medi.uni-oldenburg.de

OLDENBURG, 10. November 2016

POSTANSCHRIFT
Universität Oldenburg
Medizinische Physik
D-26111 Oldenburg

PAKETANSCHRIFT
Ammerländer Heerstraße 114 - 118
D-26129 Oldenburg

Cooperation project "Development of the Indonesian Matrix sentence test"

Dear dr.Ronny Suwento, Sp.THT-KL(K), Prof.Dr.dr.Yenni Bashiriddin, Sp.THT-KL(K), and dr.Widayat Alviandi,

I appreciate your willingness to cooperate with us on the development of the Indonesian Matrix test. With this letter, I would like to confirm the conditions of our intended collaboration in this validation project. If you agree on the conditions stated below, please confirm by signing a copy of this letter and sending it back to me. Thank you very much in advance!

The following conditions will apply for our collaboration:

Each partner within this cooperation will pay salaries and other expenses for their own personnel. Each partner has the right to publish the contents and results of the current cooperation, but has to mention the contribution of the respective other partners in an appropriate way. The University Oldenburg will cover all expenses related to the recordings and have the ownership rights for the audio speech material. RSUPN Dr.Cipto Mangunkusumo represented by dr.Suwento, Prof.Dr.dr.Bashiriddin, and dr.Alviandi has the nonexclusive, nontransferable rights for using the test methods free of charge for the duration of the cooperation project "Development of the Indonesian Matrix sentence test". Subsequently, HörTech gGmbH will provide one free, unrestricted software license for the Oldenburg Measurement Applications software including the Indonesian Matrix sentence test module (including possible future updates) to the RSUPN Dr.Cipto Mangunkusumo represented by dr.Suwento, Prof.Dr.dr.Bashiriddin, and dr.Alviandi. HörTech gGmbH intends to market the Indonesian Matrix sentence test software and the audio test material. dr.Suwento,et al.'s team at RSUPN Dr.Cipto Mangunkusumo will support this marketing effort, e.g. by referencing the HörTech software in publications on studies where this software was used. Both parties agree to exchange scientific information and results, to encourage the exchange of specialists and to support the transfer of scientific and technological achievements, to initiate the organization of scientific conferences and symposiums, and to simulate the publication of joint study results. Both parties agree that dr.Suwento,et al.'s team should lead the efforts to establish the matrix test and its distribution in Mainland Indonesia while The University Oldenburg / HörTech gGmbH lead the distribution of the matrix test for the rest of the world.

November 10, 2016, Oldenburg
Date, place (Signature Birger Kollmeier, Universität Oldenburg)

Dec, 23, 2016, Jakarta
Date, place (Signature Ronny Suwento, RSUPN Dr. Cipto Mangunkusumo)

Dec, 28, 2016 Jakarta
Date, place (Signature Yenni Bashiriddin, RSUPN Dr. Cipto Mangunkusumo)

Jan, 3, 2017, Jakarta
Date, place (Signature Widayat Alviandi, RSUPN Dr. Cipto Mangunkusumo)

D. Ethical approval

CARL VON OSSIETZKY universität OLDENBURG
FAKULTÄT VI
MEDIZIN UND GESUNDHEITSWISSENSCHAFTEN

CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG · 26111 OLDENBURG

Frau
Dr. Zokoll-van der Laan
Medizinische Physik
- hier -

Prof. Dr. Christiane Thiel
Department für Psychologie

KOMMISSION FÜR FOR-SCHUNGSFOLGENAB-SCHÄTZUNG UND ETHIK

TELEFONDURCHWAHL
+49 (0)441 798 3641
FAX
(0441) 798 3848
EMAIL
christiane.thiel@uni-oldenburg.de

POSTANSCHRIFT
D-26111 Oldenburg

OLDENBURG, 16.07.2015

ems
european medical scho
oldenburg-groningen

Stellungnahme der Kommission für Forschungsfolgenabschätzung und Ethik zum Antrag „Amendment:"Mulitlingual Audiobogical Diagnostics, dabei Probandentests (Optimierung, Evaluation und Validierung von kognitiven Tests sowie Hör- und Sprachverständlichkeitstests)" (Drs.63/2015)

Sehr geehrter Frau Dr. Zokoll-van der Laan,

die Ethikkommission hat in ihrer Sitzung vom 8.7.2015 über oben eingereichten Antrag beraten. Die Kommission ist mehrheitlich der Meinung, dass Ihrem Antrag stattgegeben werden kann. Am Antrag beteiligte Personen waren nicht in diese Entscheidung eingebunden.

Die zustimmende Bewertung ergeht unter der Annahme gleichbleibender Gegebenheiten. Die Verantwortlichkeit des jeweiligen Wissenschaftlers bleibt im vollen Umfang erhalten.

Für Ihr Vorhaben wünsche ich Ihnen viel Erfolg.
Mit freundlichen Grüßen

C. Thiel

Prof. Dr. Christiane Thiel

UNIVERSITAS INDONESIA
FAKULTAS KEDOKTERAN

Gedung Fakultas Kedokteran UI
Jl. Salemba Raya No.6, Jakarta 10430
PO.Box 1358
T. 62.21.3912477, 31930371, 31930373, 3922977, 3927360, 3153236,
F 62 21 3912477, 31930372, 3157288,
E. humas@fk.ui.ac.id, office@fk.ui.ac.id
fk.ui.ac.id

Nomor : 938 **/UN2.F1/ETIK/2016**

KETERANGAN LOLOS KAJI ETIK

ETHICAL APPROVAL

Komite Etik Penelitian Kesehatan Fakultas Kedokteran Universitas Indonesia dalam upaya melindungi hak asasi dan kesejahteraan subyek penelitian kedokteran, telah mengkaji dengan teliti protokol berjudul:

The Ethics Committee of the Faculty of Medicine, University of Indonesia, with regards of the Protection of human rights and welfare in medical research, has carefully reviewed the research protocol entitled:

"Indonesian Matrix Test: Evaluation Measurement and Validation".
No. protokol: 16-10-378

Peneliti Utama
Principal Investigator : **Felicia Primadita, B.Eng**

Nama Institusi
Name of the Institution : **Dept. of Medical Physics and Acoustics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg**

dan telah menyetujui protokol tersebut di atas.
and approved the above mentioned protocol.

Jakarta, 31 OCT 2016
Ketua
Chairman

Prof. Dr. dr. Rianto Setiabudy, SpFK

* *Ethical approval* berlaku satu tahun dari tanggal persetujuan.
** Peneliti berkewajiban
1. Menjaga kerahasiaan identitas subyek penelitian.
2. Memberitahukan status penelitian apabila
   a. Setelah masa berlakunya keterangan lolos kaji etik, penelitian masih belum selesai, dalam hal ini *ethical approval* harus diperpanjang.
   b. Penelitian berhenti di tengah jalan.
3. Melaporkan kejadian serius yang tidak diinginkan (*serious adverse events*).
4. Peneliti tidak boleh melakukan tindakan apapun pada subyek sebelum protokol penelitian mendapat lolos kaji etik dan sebelum memperoleh *informed consent* dari subjek penelitian.
5. Menyampaikan laporan akhir, bila penelitian sudah selesai.
6. Cantumkan nomor protokol ID pada setiap komunikasi dengan KEPK FKUI-RSCM.

Semua prosedur persetujuan dilakukan sesuai dengan standar ICH-GCP.
All procedure of Ethical Approval are performed in accordance with ICH-GCP standard procedure.

E. Optimization: Psychometric function of all realizations

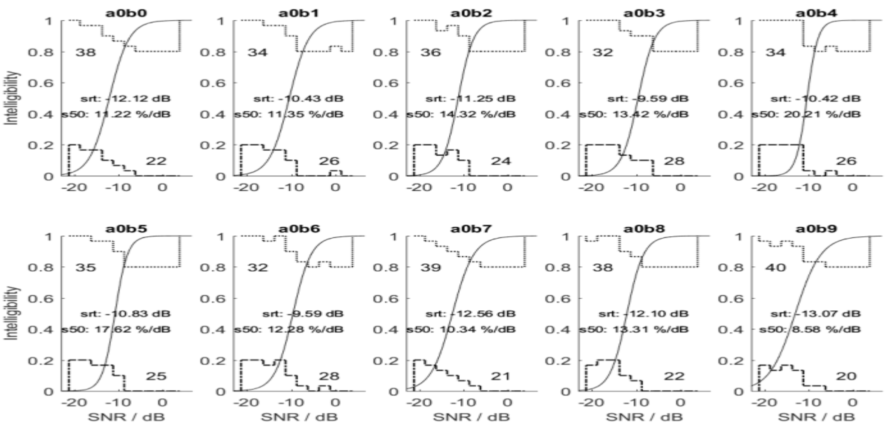


Figure E.1.: Agus

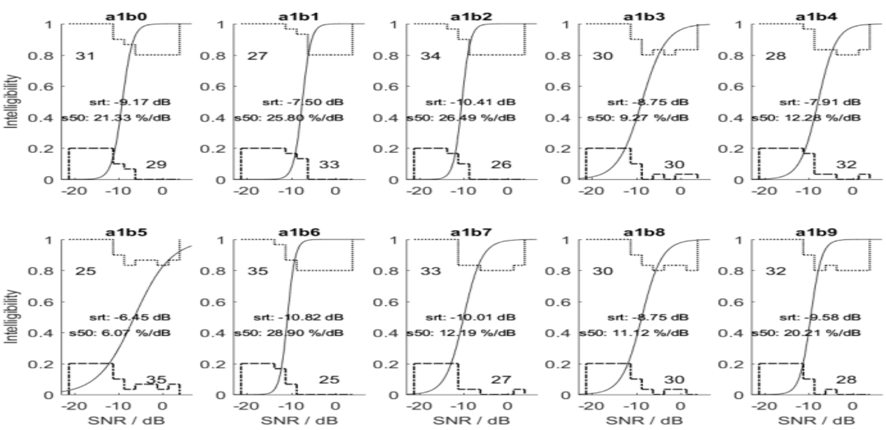


Figure E.2.: Arif

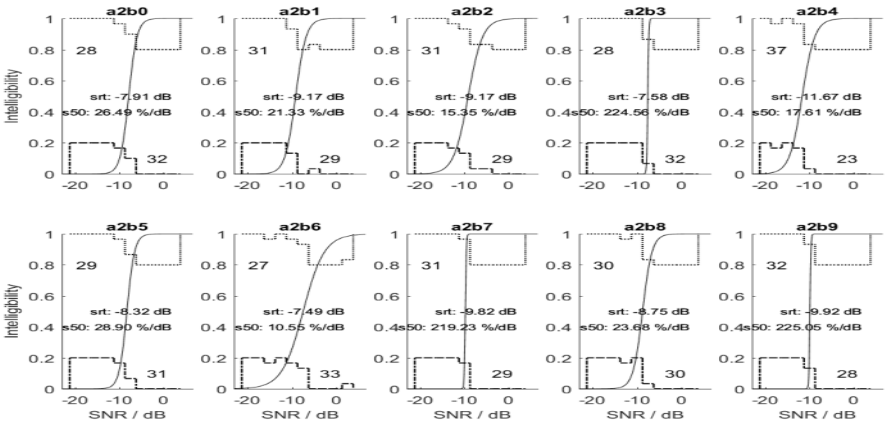


Figure E.3.: Ayu

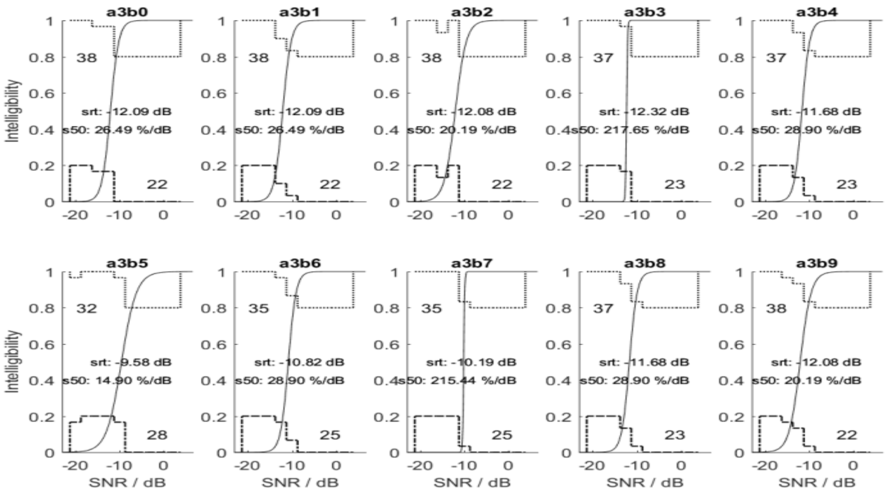


Figure E.4.: Fajar

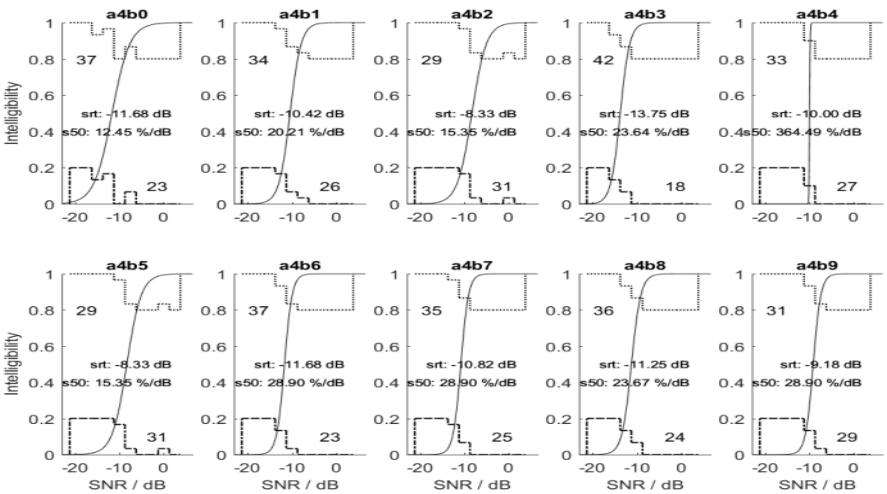


Figure E.5.: Ika

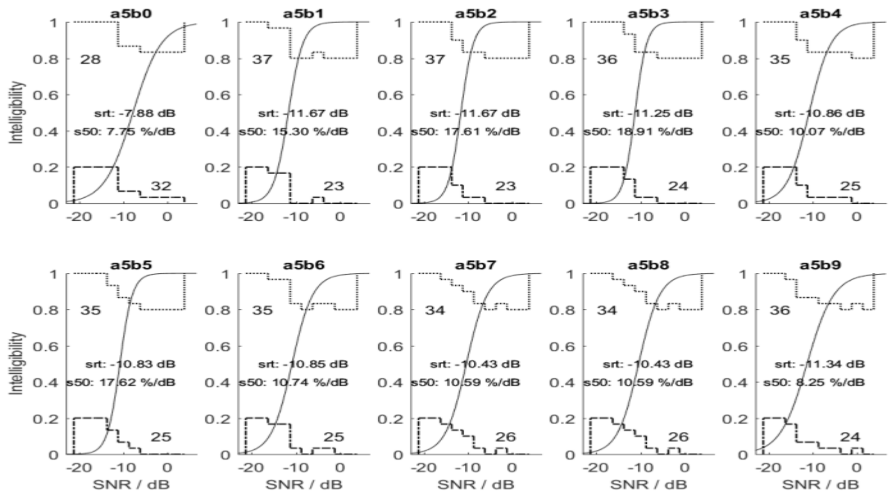


Figure E.6.: Made

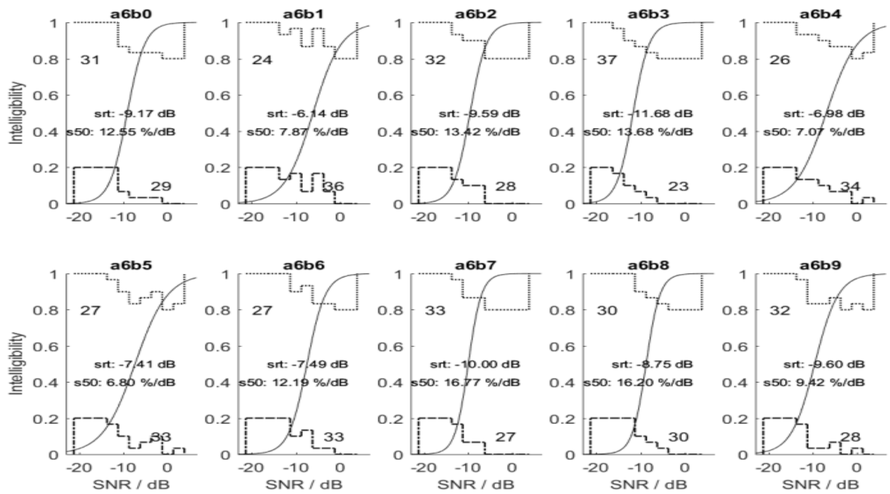


Figure E.7.: Maya

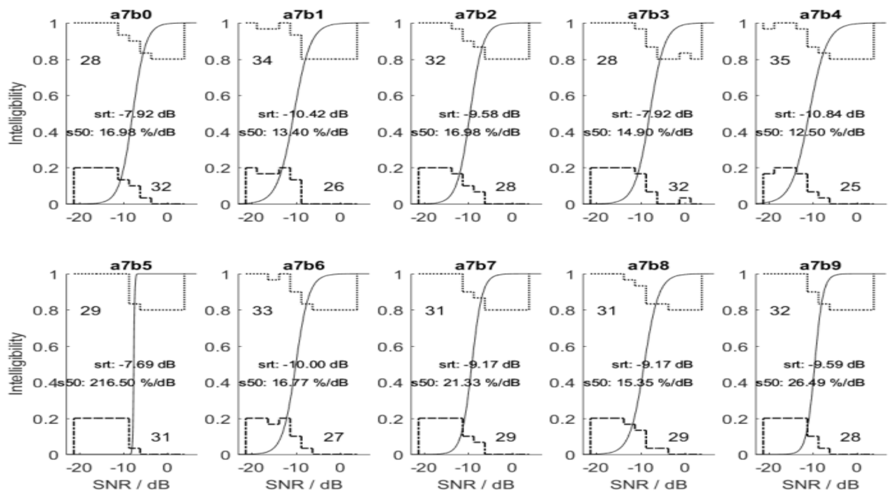


Figure E.8.: Putri

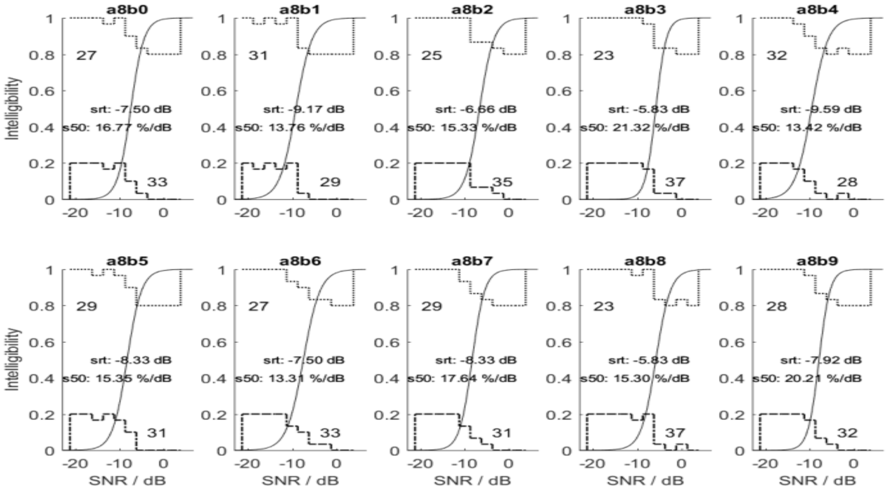


Figure E.9.: Putu

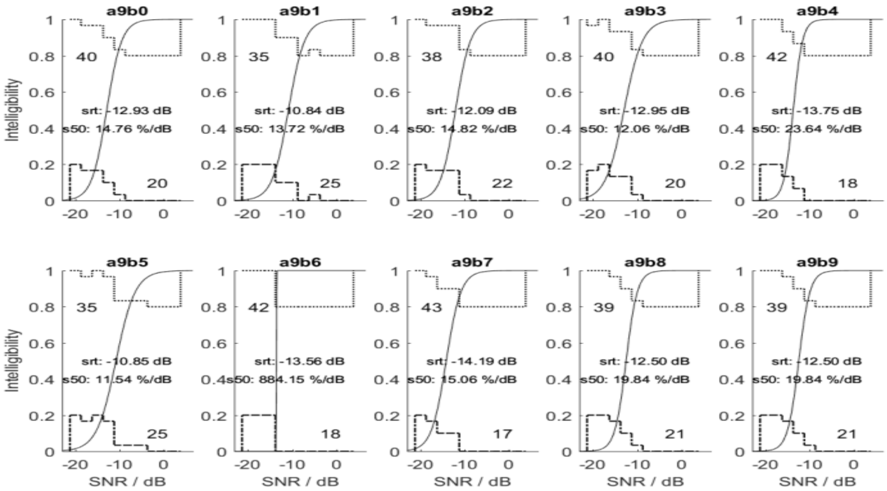


Figure E.10.: Sari

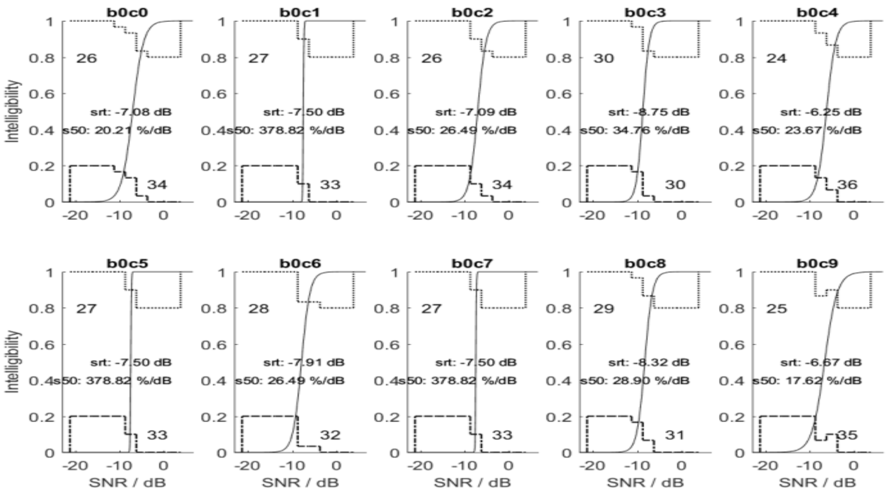


Figure E.11.: Melihat (to see)

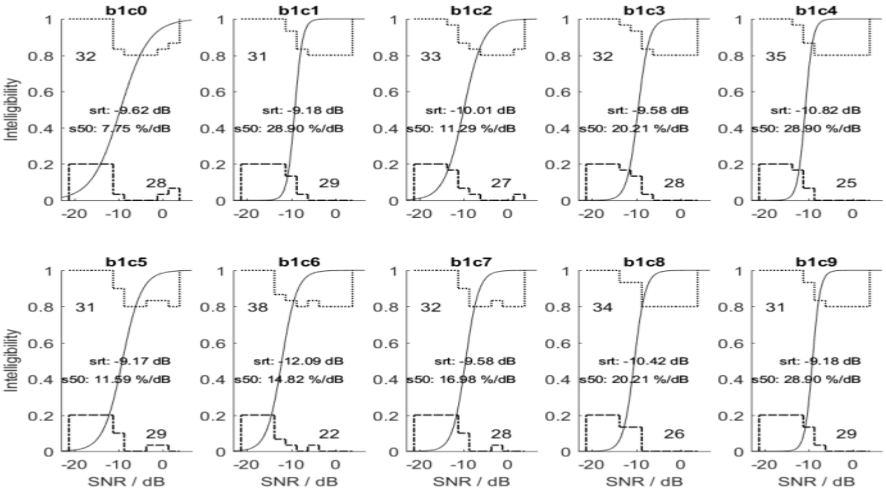


Figure E.12.: Membawa (to bring)

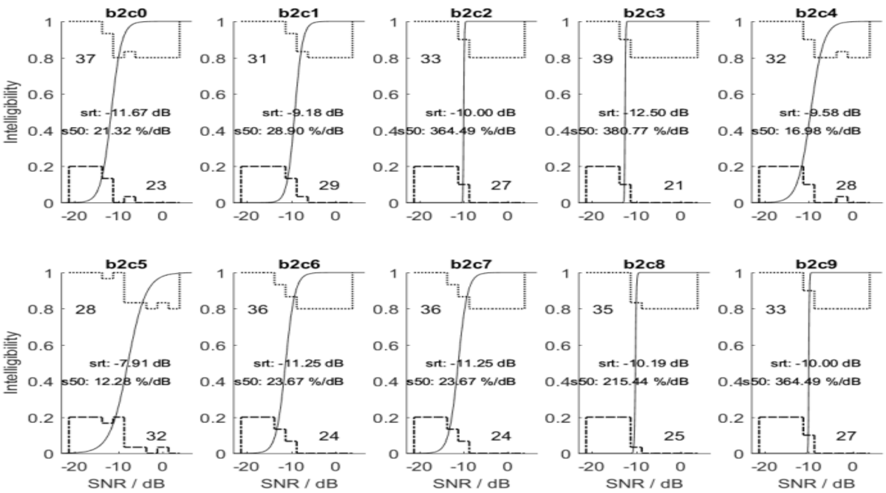


Figure E.13.: Membayar (to pay)

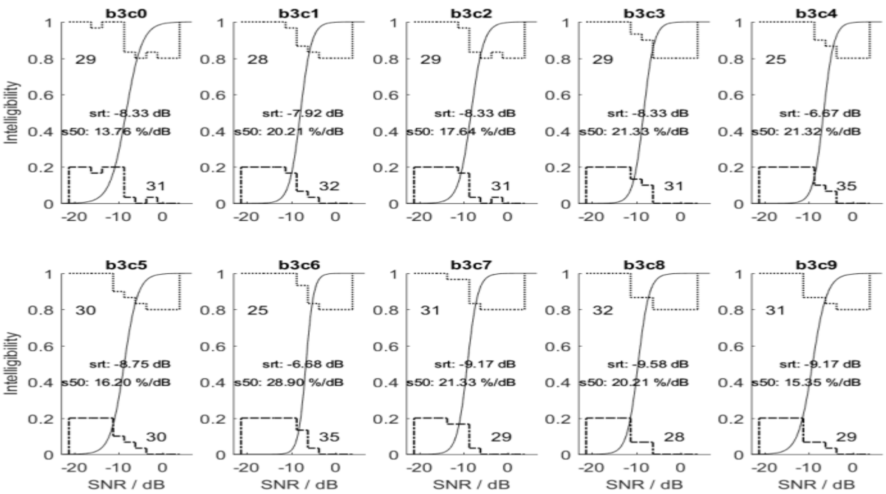


Figure E.14.: Memilih (to choose)

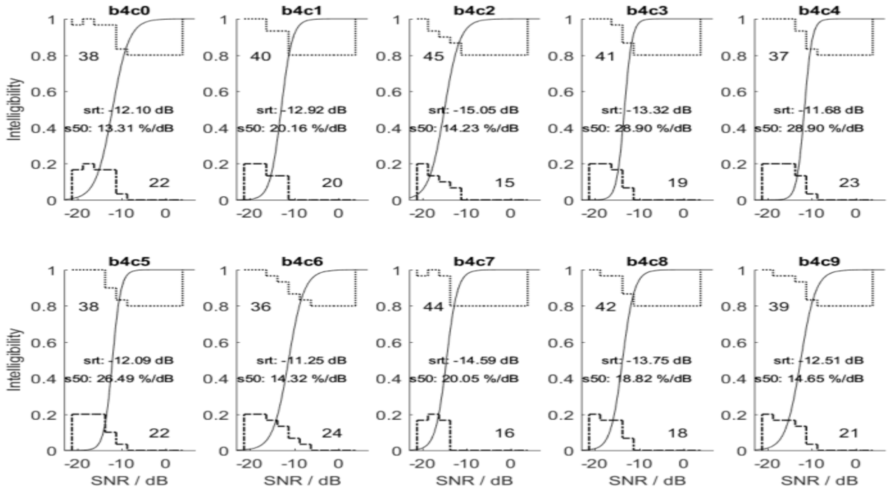


Figure E.15.: Mencari (to look for)

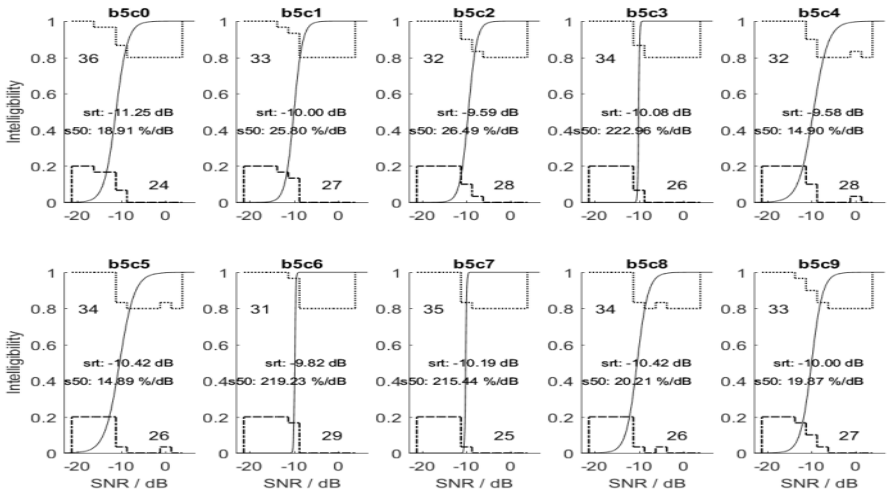


Figure E.16.: Mendapat (to get)

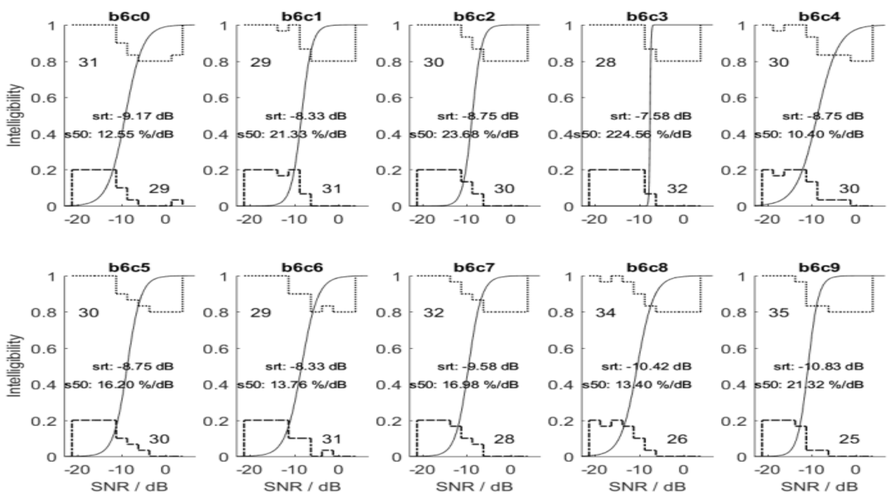


Figure E.17.: Mengangkat (to lift)

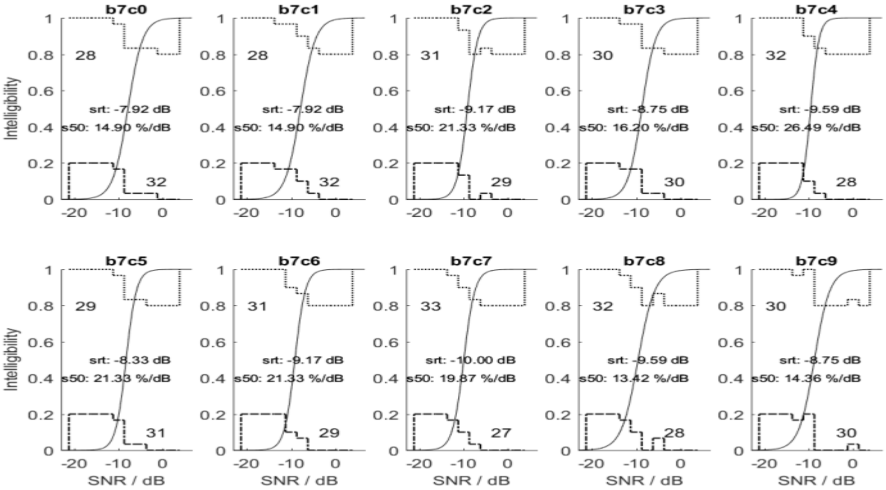


Figure E.18.: Mengganti (to change)

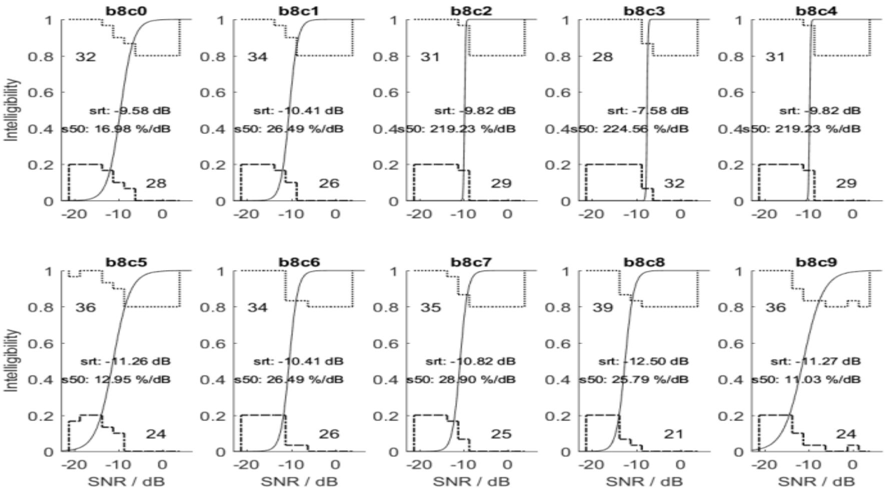


Figure E.19.: Menjual (to sell)

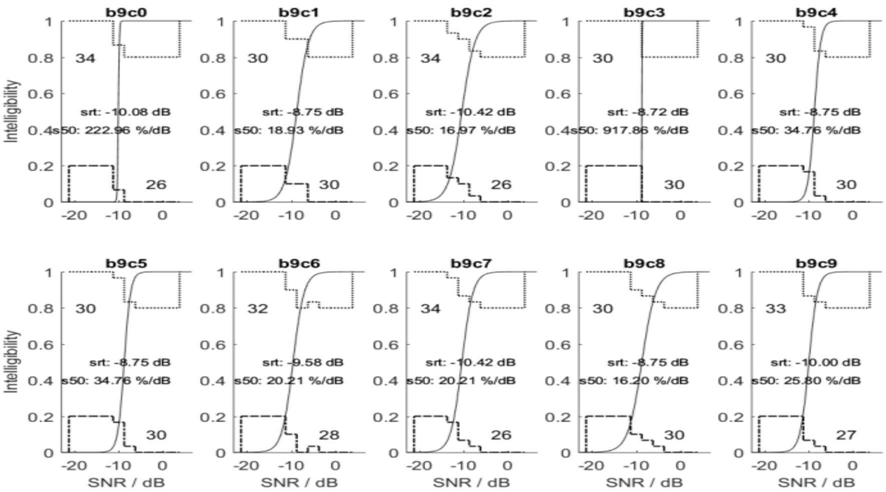


Figure E.20.: Menyimpan (to store)

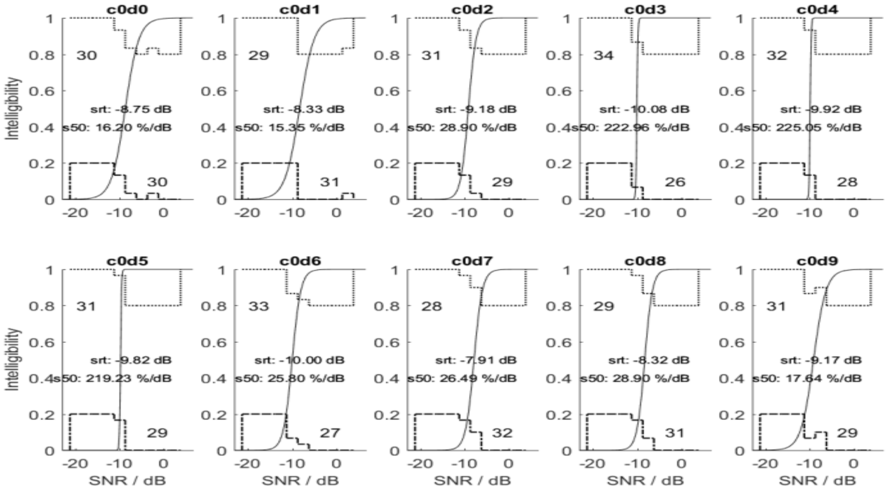


Figure E.21.: Delapan (eight)

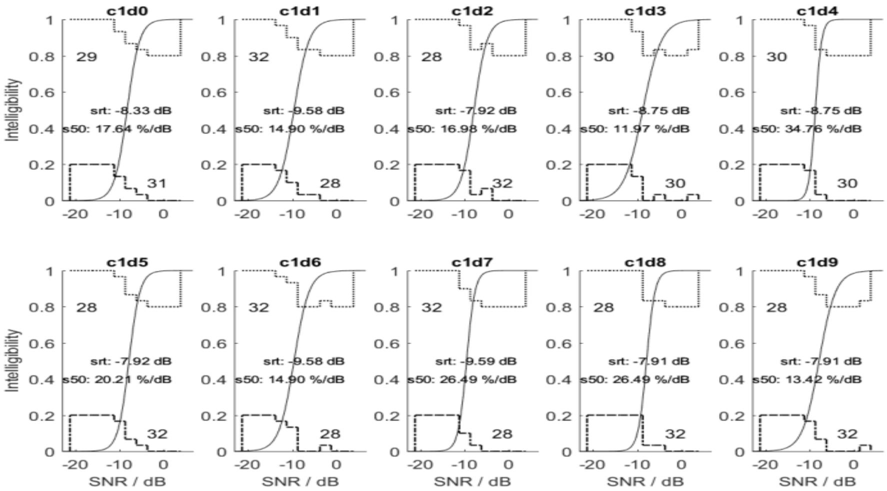


Figure E.22.: Dua (two)

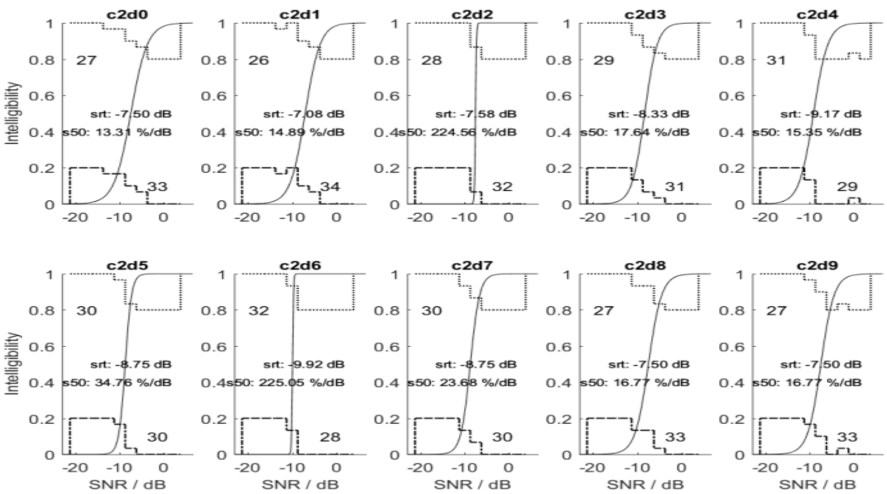


Figure E.23.: Empat(four)

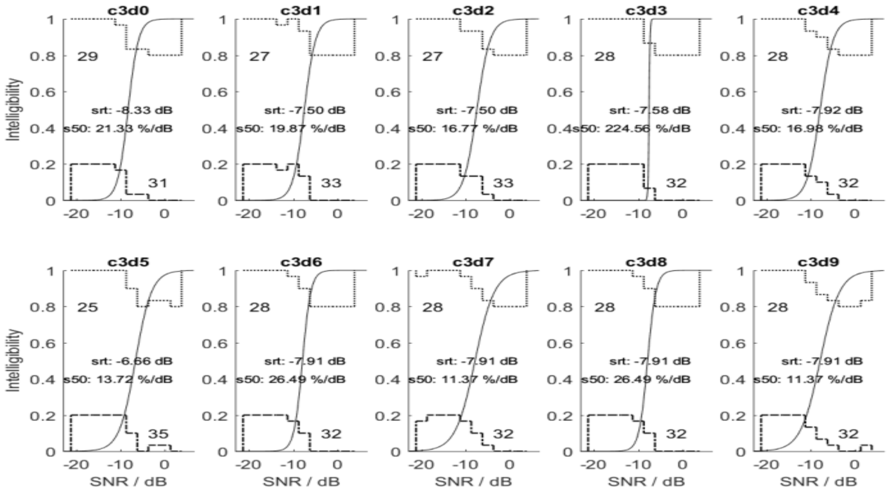


Figure E.24.: Enam (six)

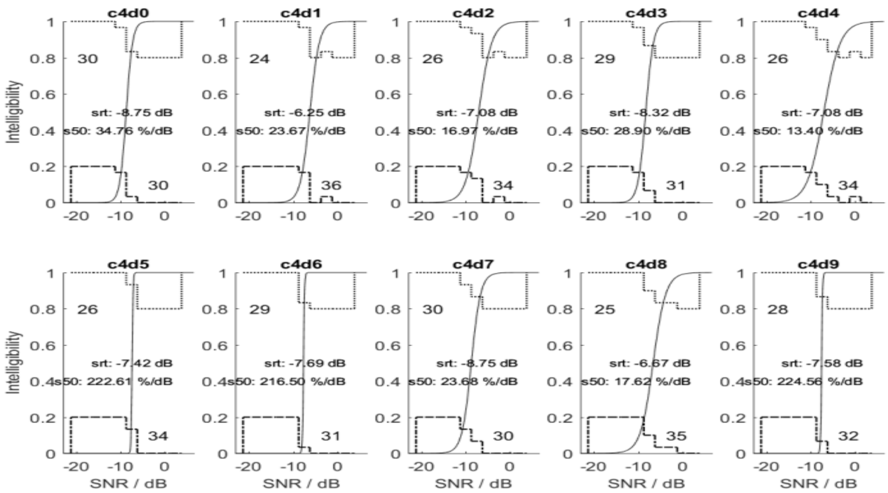


Figure E.25.: Lima (five)

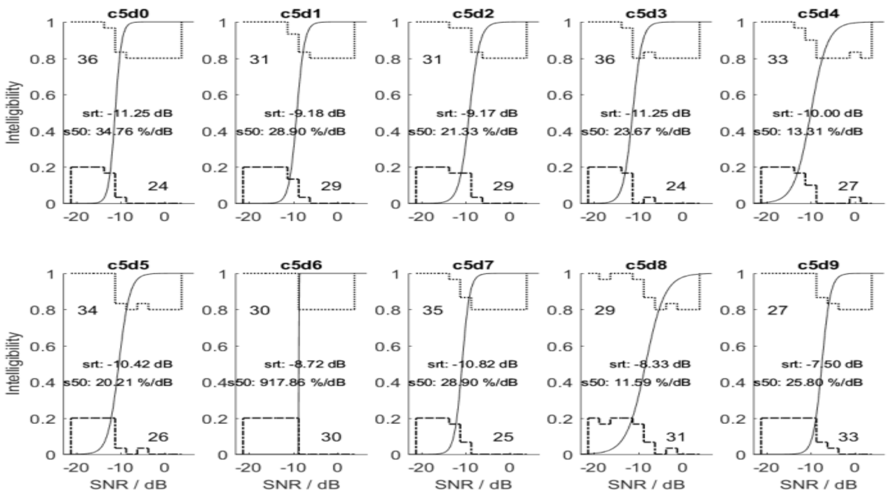


Figure E.26.: Sembilan (nine)

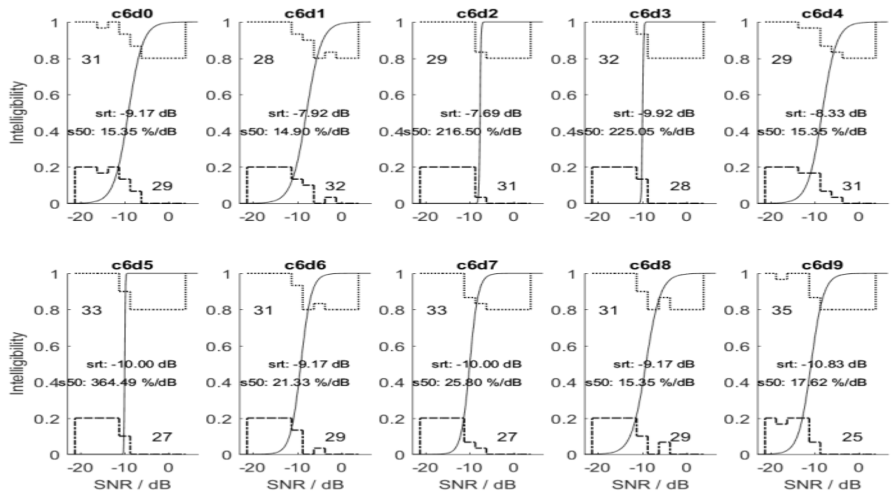


Figure E.27.: Sepuluh (ten)

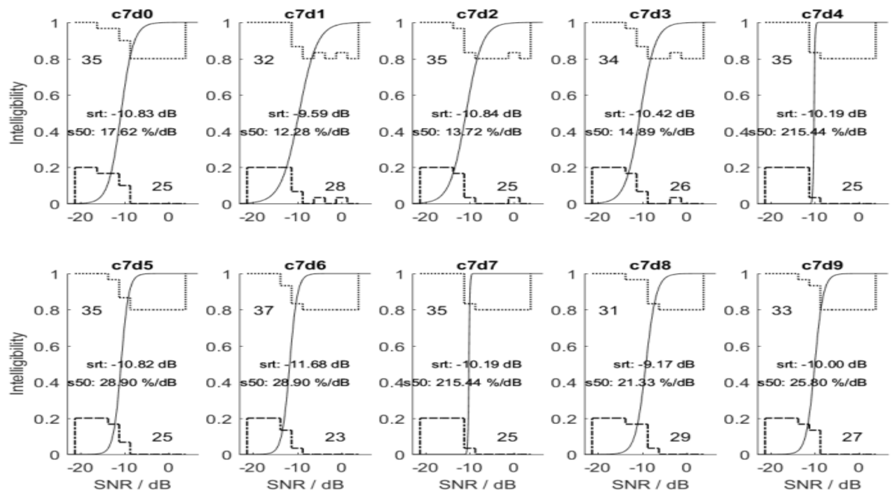


Figure E.28.: Seratus

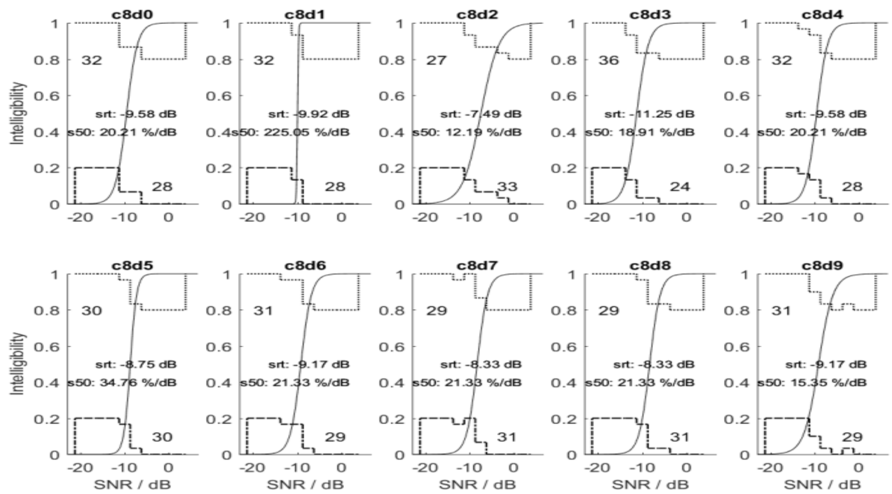


Figure E.29.: Tiga (three)

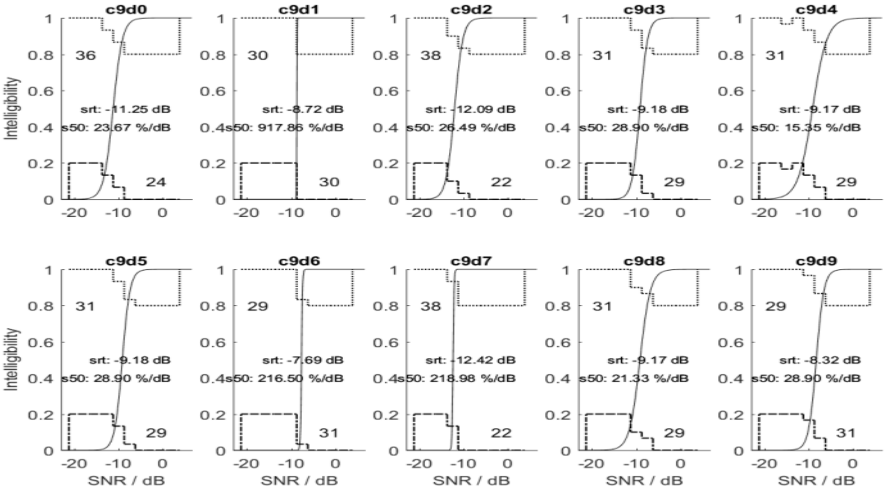


Figure E.30.: Tujuh (seven)

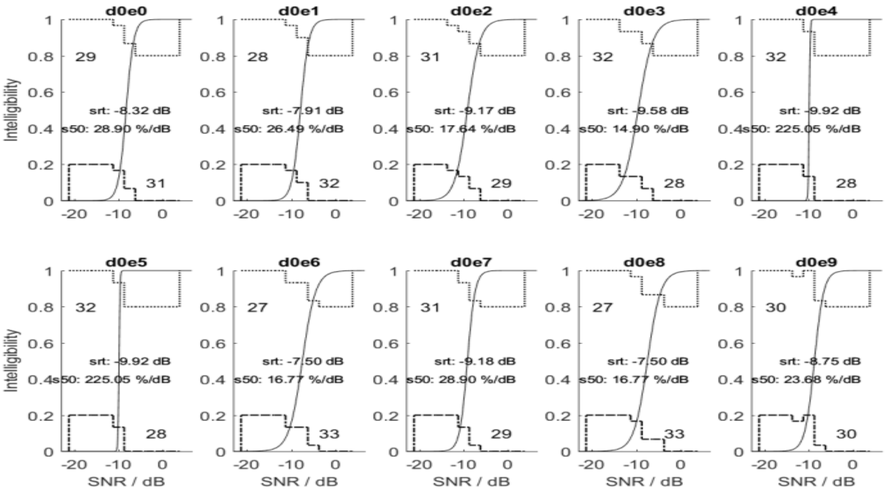


Figure E.31.: Bola (ball)

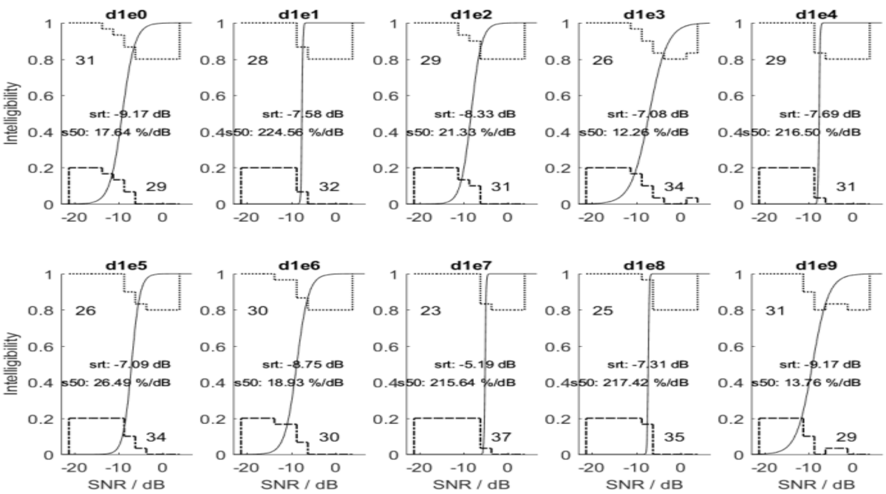


Figure E.32.: Buah (fruit)

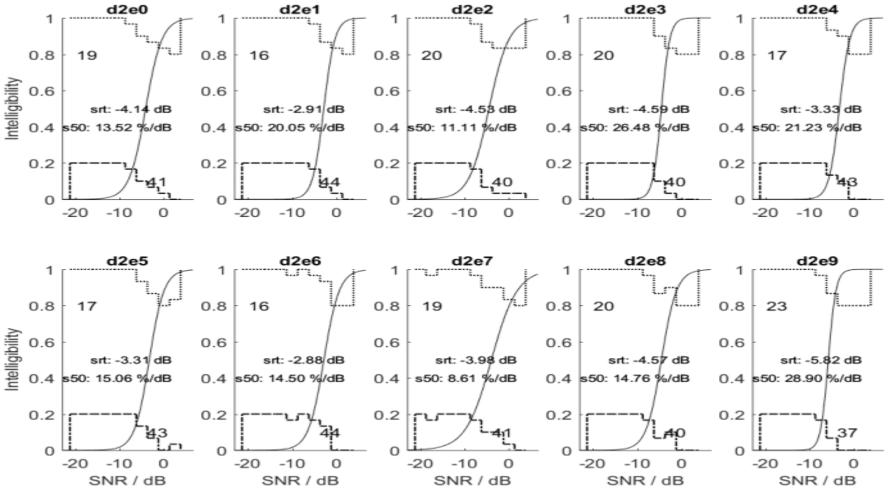


Figure E.33.: Buku (book)

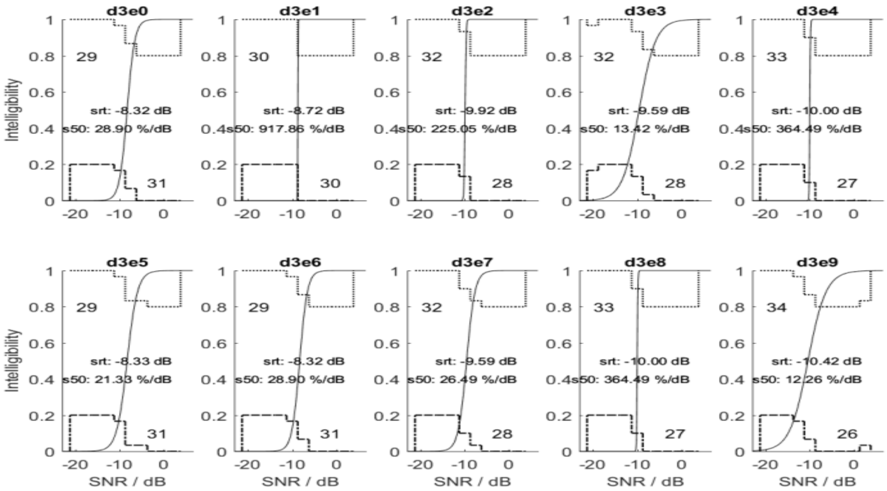


Figure E.34.: Cangkir (cup)

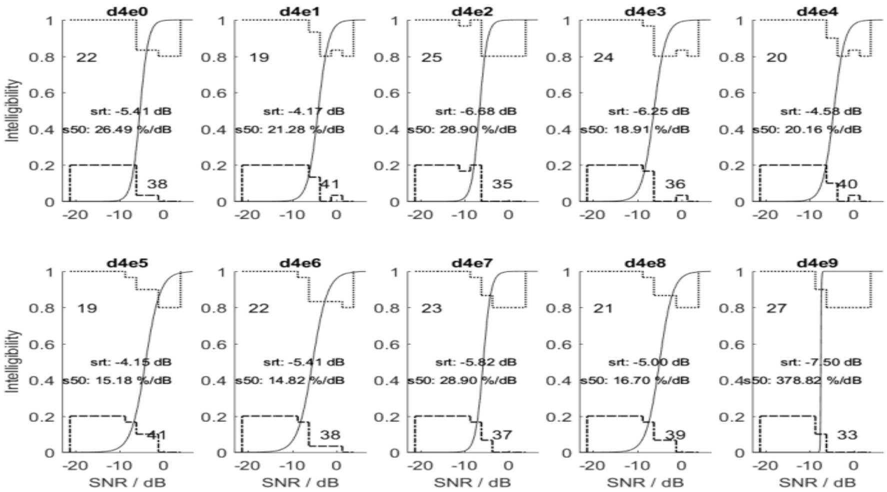


Figure E.35.: Gitar (guitar)

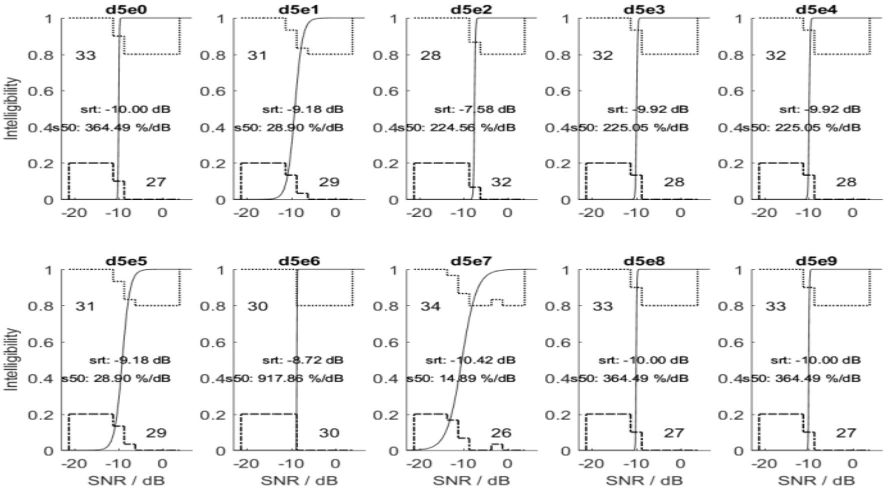


Figure E.36.: Kaleng (can)

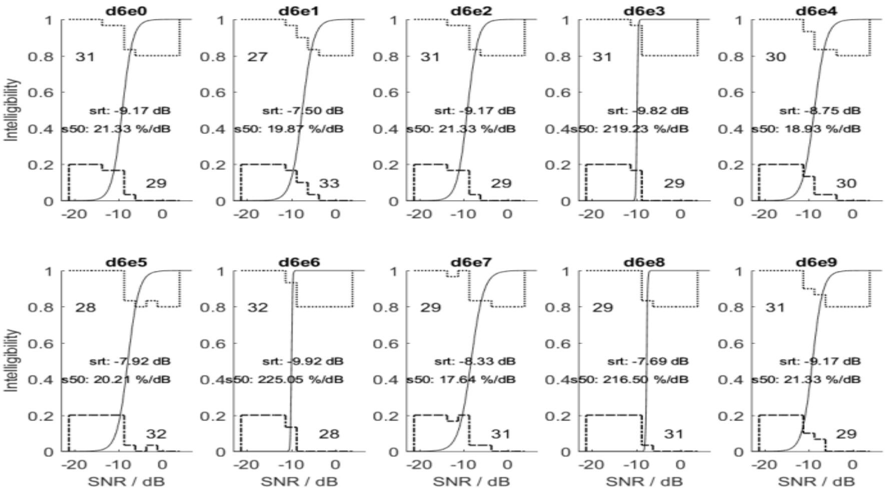


Figure E.37.: Kotak (box)

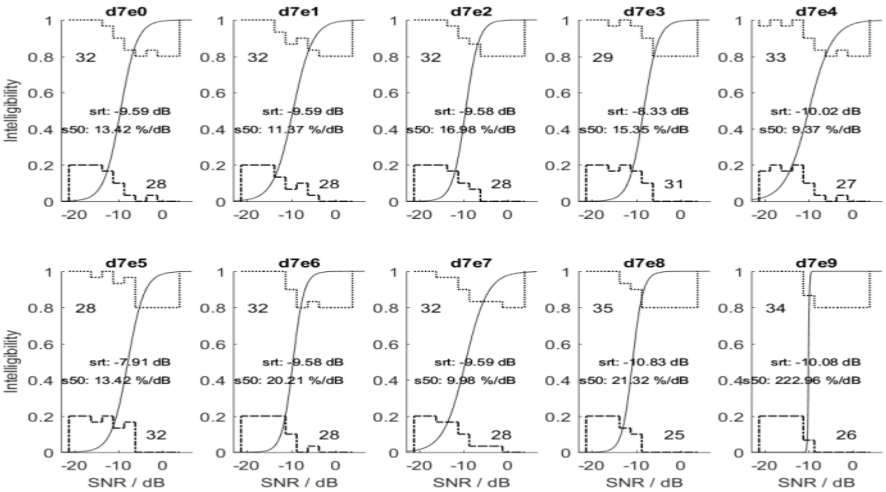


Figure E.38.: Kunci (key)

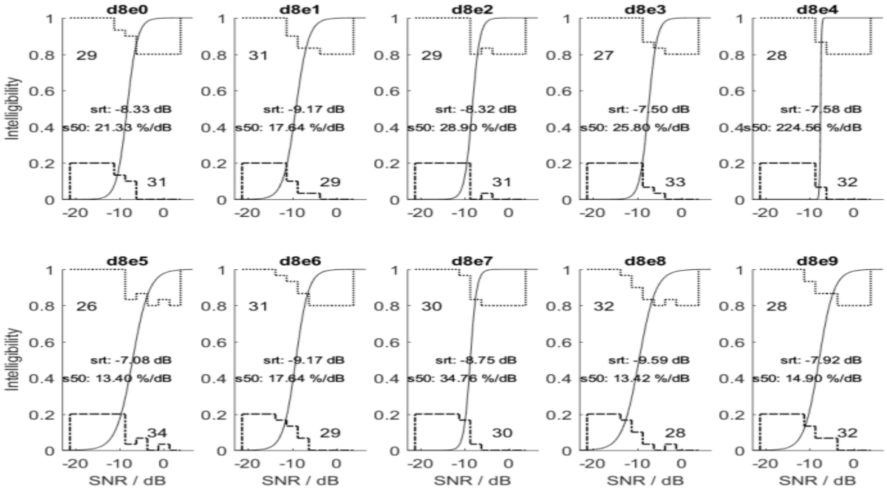


Figure E.39.: Sendok (spoon)

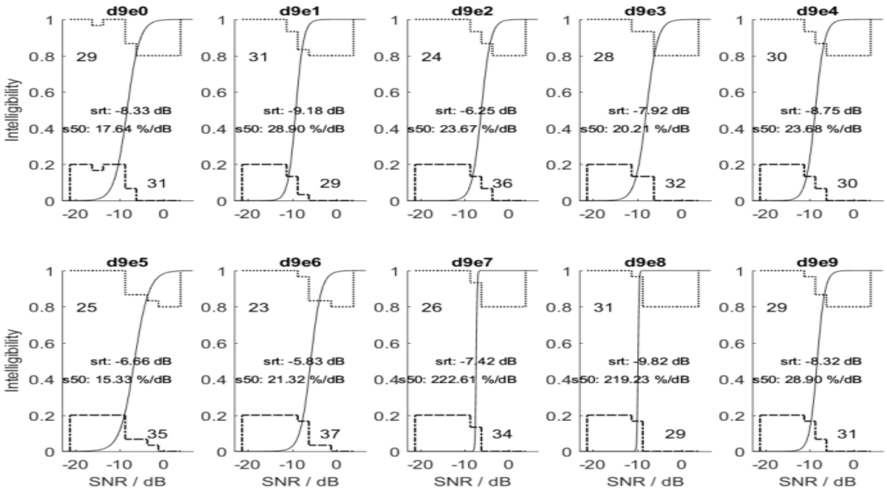


Figure E.40.: Sofa (sofa)

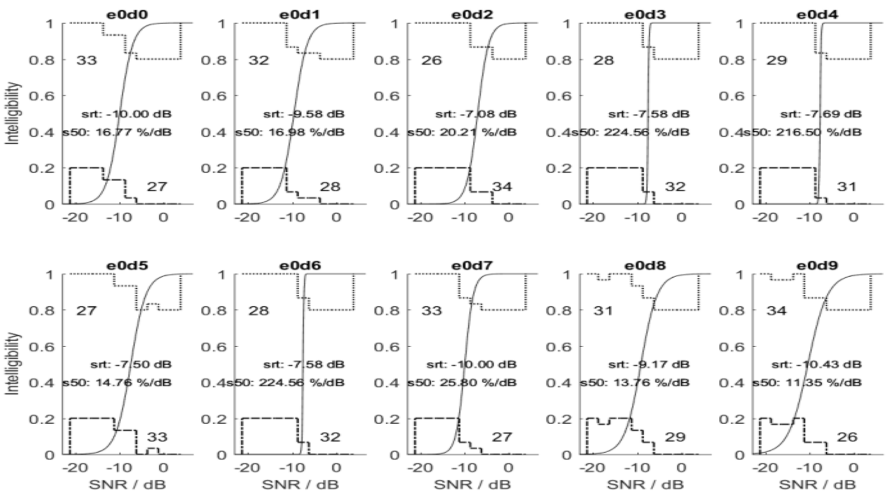


Figure E.41.: Bagus (good)

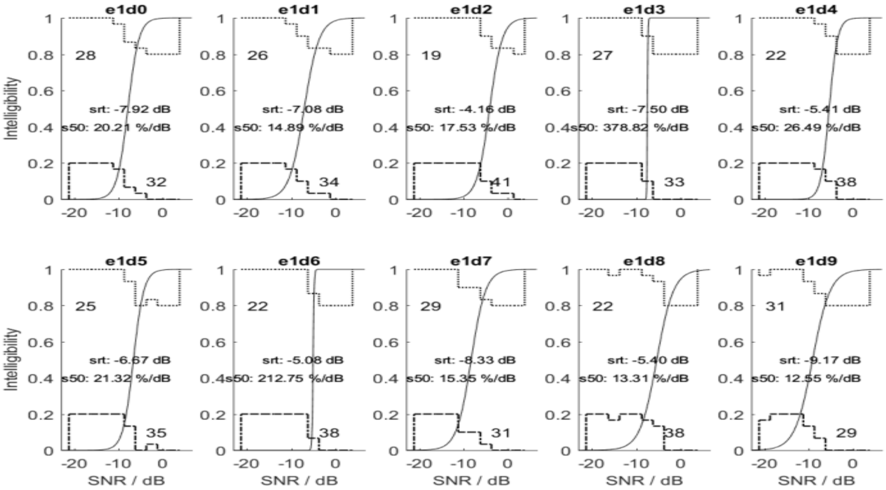


Figure E.42.: Baru (new)

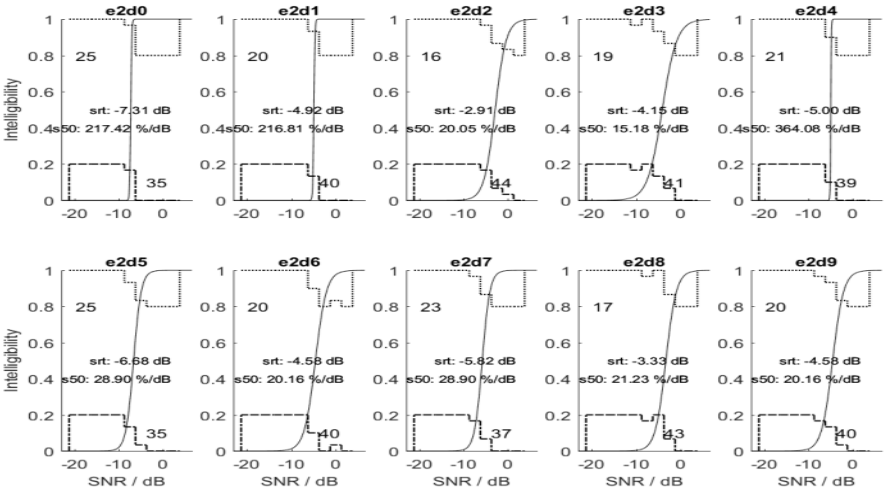


Figure E.43.: Berat (heavy)

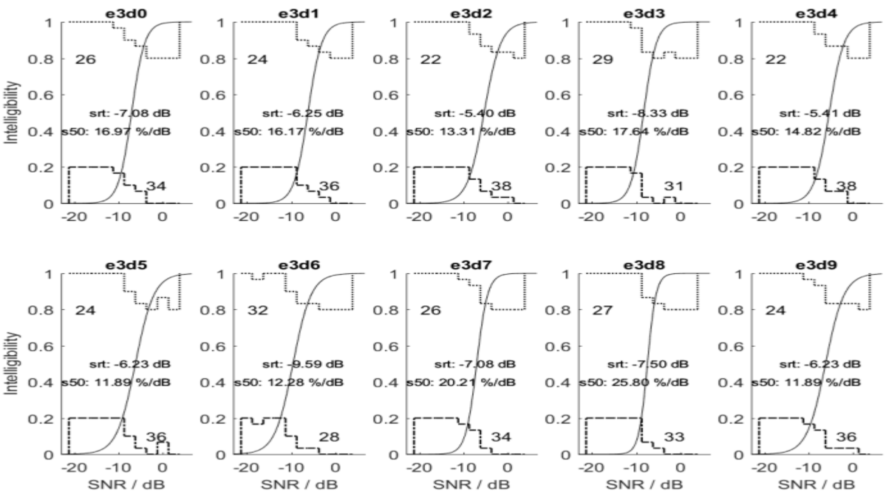


Figure E.44.: Bersih (clean)

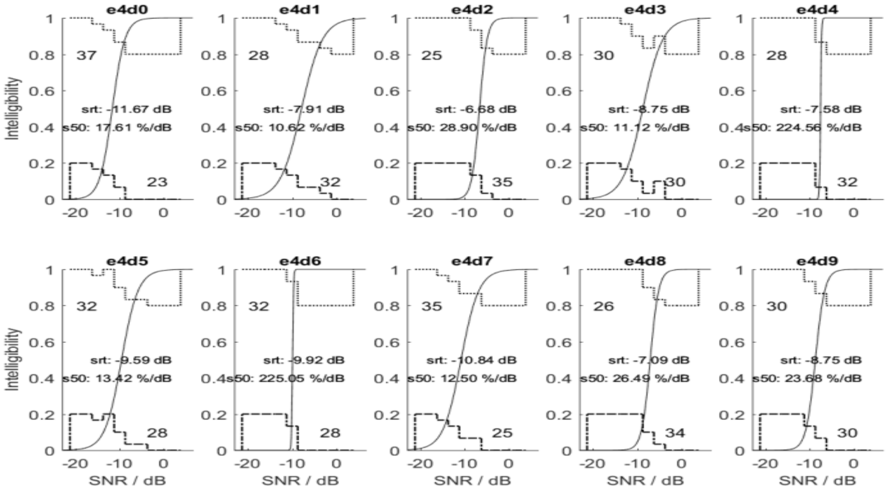


Figure E.45.: Kecil (small)

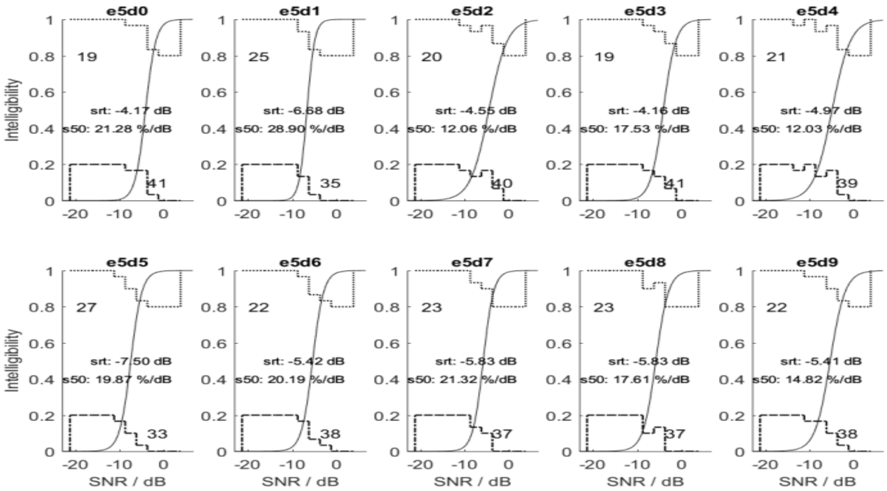


Figure E.46.: Keras (hard)

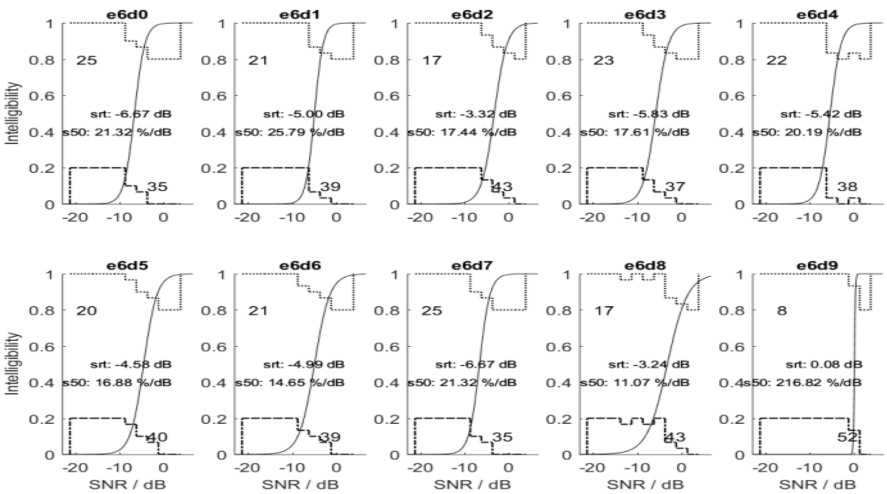


Figure E.47.: Kotor (dirty)

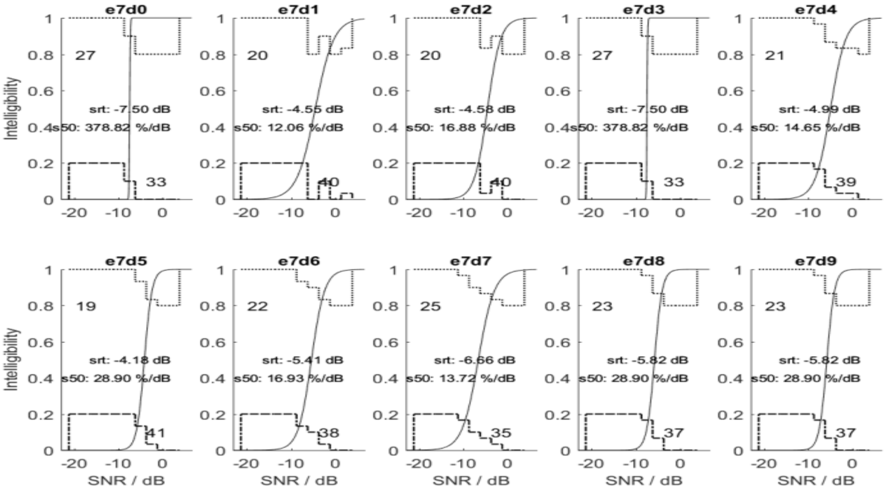


Figure E.48.: Kuning (yellow)

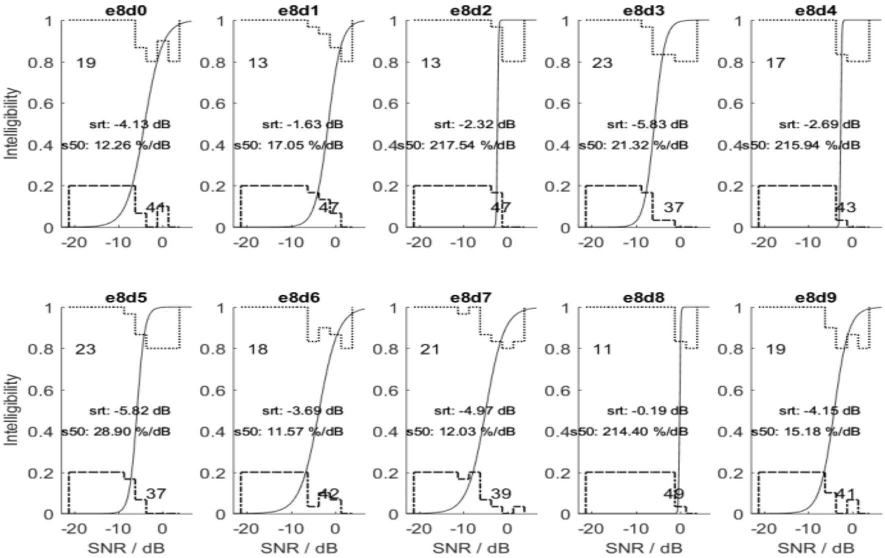


Figure E.49.: Mahal (expensive)

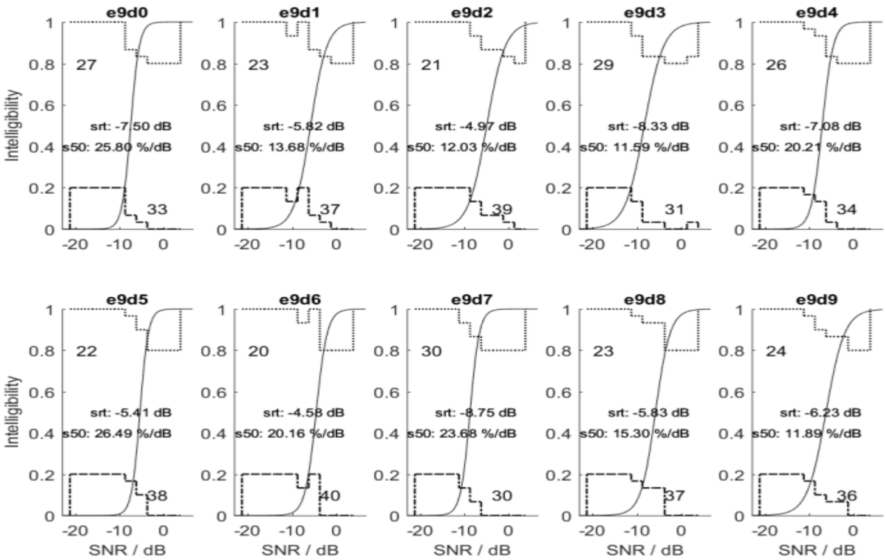


Figure E.50.: Putih (white)

F. Evaluation:Test list equivalence

Paarweise Vergleiche

Maß: MEASURE_1

(I)Testlist	(J)Testlist	Mittlere Differenz (I-J)	Standard-fehler	Sig.[b]	95% Konfidenzintervall für die Differenz[b] Untergrenze	Obergrenze
1	2	-,056	,115	1,000	-,555	,444
	3	-,149	,143	1,000	-,770	,472
	4	,191	,146	1,000	-,445	,826
	5	,060	,161	1,000	-,642	,762
	6	-,194	,135	1,000	-,779	,392
	7	-,452	,139	,797	-1,058	,154
	8	,109	,145	1,000	-,521	,739
	9	-,141	,136	1,000	-,735	,452
	10	,420	,140	1,000	-,189	1,030
	11	-,020	,152	1,000	-,680	,641
	12	,078	,154	1,000	-,593	,749
	13	,112	,117	1,000	-,398	,622
	14	-,133	,104	1,000	-,586	,320
	15	,170	,099	1,000	-,260	,601
	16	-,043	,107	1,000	-,506	,420
	17	-,030	,116	1,000	-,535	,476
	18	-,209	,107	1,000	-,676	,259
	19	-,036	,099	1,000	-,465	,394
	20	,144	,121	1,000	-,382	,670
	21	-,035	,140	1,000	-,643	,573
	22	-,077	,124	1,000	-,619	,465
2	1	,056	,115	1,000	-,444	,555
	3	-,093	,161	1,000	-,792	,605
	4	,246	,160	1,000	-,452	,945
	5	,116	,143	1,000	-,508	,739
	6	-,138	,123	1,000	-,674	,398
	7	-,396	,142	1,000	-1,015	,223
	8	,165	,135	1,000	-,422	,752
	9	-,086	,118	1,000	-,598	,427
	10	,476	,147	,790	-,162	1,114
	11	,036	,136	1,000	-,556	,628
	12	,134	,161	1,000	-,568	,836
	13	,167	,109	1,000	-,305	,640
	14	-,077	,113	1,000	-,570	,416
	15	,226	,119	1,000	-,293	,746
	16	,013	,112	1,000	-,473	,498
	17	,026	,121	1,000	-,502	,554
	18	-,153	,129	1,000	-,714	,408
	19	,020	,130	1,000	-,544	,584
	20	,200	,130	1,000	-,367	,767
	21	,021	,136	1,000	-,572	,613
	22	-,021	,149	1,000	-,672	,629
3	1	,149	,143	1,000	-,472	,770
	2	,093	,161	1,000	-,605	,792
	4	,340	,180	1,000	-,445	1,125
	5	,209	,145	1,000	-,421	,839
	6	-,044	,123	1,000	-,580	,492
	7	-,303	,105	1,000	-,761	,155
	8	,258	,133	1,000	-,321	,837
	9	,008	,165	1,000	-,711	,726
	10	,569	,157	,307	-,113	1,252
	11	,130	,143	1,000	-,494	,753
	12	,227	,130	1,000	-,340	,794
	13	,261	,133	1,000	-,319	,841
	14	,016	,128	1,000	-,542	,575
	15	,320	,125	1,000	-,223	,862
	16	,106	,164	1,000	-,609	,821
	17	,120	,115	1,000	-,381	,620
	18	-,060	,141	1,000	-,673	,554
	19	,114	,096	1,000	-,305	,532
	20	,293	,121	1,000	-,233	,819
	21	,114	,160	1,000	-,581	,809
	22	,072	,108	1,000	-,399	,543
4	1	-,191	,146	1,000	-,826	,445
	2	-,246	,160	1,000	-,945	,452
	3	-,340	,180	1,000	-1,125	,445
	5	-,131	,170	1,000	-,869	,608
	6	-,384	,163	1,000	-1,093	,325
	7	-,643	,157	,095	-1,325	,040
	8	-,081	,148	1,000	-,723	,561
	9	-,332	,140	1,000	-,943	,279
	10	,230	,135	1,000	-,356	,816
	11	-,210	,145	1,000	-,843	,423
	12	-,113	,173	1,000	-,865	,640
	13	-,079	,143	1,000	-,703	,545
	14	-,323	,151	1,000	-,979	,332
	15	-,020	,141	1,000	-,635	,595
	16	-,234	,154	1,000	-,903	,436
	17	-,220	,129	1,000	-,783	,343
	18	-,399	,125	,887	-,942	,144
	19	-,226	,140	1,000	-,837	,385
	20	-,046	,152	1,000	-,710	,617
	21	-,226	,161	1,000	-,927	,475

	22	-,268	,129	1,000	-,831	,296
5	1	-,060	,161	1,000	-,762	,642
	2	-,116	,143	1,000	-,739	,508
	3	-,209	,145	1,000	-,839	,421
	4	,131	,170	1,000	-,608	,869
	6	-,253	,128	1,000	-,809	,303
	7	-,512	,137	,230	-1,106	,082
	8	,049	,145	1,000	-,581	,679
	9	-,201	,129	1,000	-,763	,360
	10	,360	,182	1,000	-,432	1,153
	11	-,079	,157	1,000	-,763	,604
	12	,018	,160	1,000	-,680	,716
	13	,052	,158	1,000	-,634	,738
	14	-,193	,149	1,000	-,839	,454
	15	,111	,153	1,000	-,557	,778
	16	-,103	,158	1,000	-,790	,584
	17	-,090	,137	1,000	-,684	,505
	18	-,269	,151	1,000	-,928	,390
	19	-,095	,158	1,000	-,781	,590
	20	,084	,141	1,000	-,529	,697
	21	-,095	,149	1,000	-,742	,552
	22	-,137	,147	1,000	-,778	,504
6	1	,194	,135	1,000	-,392	,779
	2	,138	,123	1,000	-,398	,674
	3	,044	,123	1,000	-,492	,580
	4	,384	,163	1,000	-,325	1,093
	5	,253	,128	1,000	-,303	,809
	7	-,259	,126	1,000	-,805	,288
	8	,303	,113	1,000	-,191	,796
	9	,052	,129	1,000	-,509	,613
	10	,614	,157	,155	-,070	1,298
	11	,174	,122	1,000	-,356	,704
	12	,272	,159	1,000	-,420	,963
	13	,305	,138	1,000	-,297	,908
	14	,061	,142	1,000	-,558	,679
	15	,364	,106	,499	-,097	,825
	16	,151	,132	1,000	-,423	,725
	17	,164	,122	1,000	-,365	,693
	18	-,015	,139	1,000	-,619	,589
	19	,158	,120	1,000	-,364	,680
	20	,338	,121	1,000	-,187	,862
	21	,158	,138	1,000	-,443	,760
	22	,117	,141	1,000	-,496	,729
7	1	,452	,139	,797	-,154	1,058
	2	,396	,142	1,000	-,223	1,015
	3	,303	,105	1,000	-,155	,761
	4	,643	,157	,095	-,040	1,325
	5	,512	,137	,230	-,082	1,106
	6	,259	,126	1,000	-,288	,805
	8	,561	,157	,348	-,121	1,243
	9	,311	,162	1,000	-,394	1,015
	10	,872*	,159	,003	,181	1,564
	11	,432	,116	,243	-,073	,938
	12	,530*	,114	,024	,033	1,028
	13	,564*	,126	,035	,017	1,110
	14	,319	,119	1,000	-,197	,836
	15	,623*	,121	,007	,097	1,148
	16	,409	,135	1,000	-,179	,997
	17	,422	,103	,097	-,027	,872
	18	,243	,132	1,000	-,333	,819
	19	,417	,102	,101	-,028	,861
	20	,596*	,118	,008	,083	1,109
	21	,417	,133	1,000	-,160	,994
	22	,375	,115	,774	-,126	,876
8	1	-,109	,145	1,000	-,739	,521
	2	-,165	,135	1,000	-,752	,422
	3	-,258	,133	1,000	-,837	,321
	4	,081	,148	1,000	-,561	,723
	5	-,049	,145	1,000	-,679	,581
	6	-,303	,113	1,000	-,796	,191
	7	-,561	,157	,348	-1,243	,121
	9	-,250	,108	1,000	-,719	,218
	10	,311	,151	1,000	-,346	,968
	11	-,129	,126	1,000	-,677	,420
	12	-,031	,150	1,000	-,686	,623
	13	,003	,140	1,000	-,604	,610
	14	-,242	,156	1,000	-,919	,435
	15	,061	,131	1,000	-,509	,631
	16	-,152	,161	1,000	-,853	,549
	17	-,139	,143	1,000	-,760	,483
	18	-,318	,141	1,000	-,933	,297
	19	-,145	,129	1,000	-,705	,415
	20	,035	,135	1,000	-,552	,622
	21	-,144	,169	1,000	-,878	,590
	22	-,186	,146	1,000	-,820	,448
9	1	,141	,136	1,000	-,452	,735
	2	,086	,118	1,000	-,427	,598
	3	-,008	,165	1,000	-,726	,711

	4	,332	,140	1,000	-,279	,943
	5	,201	,129	1,000	-,360	,763
	6	-,052	,129	1,000	-,613	,509
	7	-,311	,162	1,000	-1,015	,394
	8	,250	,108	1,000	-,218	,719
	10	,562*	,123	,030	,025	1,099
	11	,122	,157	1,000	-,561	,805
	12	,219	,149	1,000	-,430	,869
	13	,253	,131	1,000	-,318	,825
	14	,009	,135	1,000	-,577	,594
	15	,312	,129	1,000	-,248	,872
	16	,098	,154	1,000	-,572	,769
	17	,112	,141	1,000	-,501	,725
	18	-,067	,129	1,000	-,629	,494
	19	,106	,126	1,000	-,443	,655
	20	,286	,106	1,000	-,177	,748
	21	,106	,152	1,000	-,557	,769
	22	,064	,150	1,000	-,590	,718
10	1	-,420	,140	1,000	-1,030	,189
	2	-,476	,147	,790	-1,114	,162
	3	-,569	,157	,307	-1,252	,113
	4	-,230	,135	1,000	-,816	,356
	5	-,360	,182	1,000	-1,153	,432
	6	-,614	,157	,155	-1,298	,070
	7	-,872*	,159	,003	-1,564	-,181
	8	-,311	,151	1,000	-,968	,346
	9	-,562*	,123	,030	-1,099	-,025
	11	-,440	,144	1,000	-1,068	,188
	12	-,342	,146	1,000	-,978	,294
	13	-,309	,116	1,000	-,815	,198
	14	-,553*	,122	,031	-1,084	-,022
	15	-,250	,128	1,000	-,805	,305
	16	-,463	,176	1,000	-1,229	,302
	17	-,450	,117	,177	-,958	,058
	18	-,629*	,118	,004	-1,144	-,114
	19	-,456	,110	,082	-,933	,021
	20	-,276	,128	1,000	-,832	,280
	21	-,456	,158	1,000	-1,141	,230
	22	-,497	,130	,192	-1,064	,069
11	1	,020	,152	1,000	-,641	,680
	2	-,036	,136	1,000	-,628	,556
	3	-,130	,143	1,000	-,753	,494
	4	,210	,145	1,000	-,423	,843
	5	,079	,157	1,000	-,604	,763
	6	-,174	,122	1,000	-,704	,356
	7	-,432	,116	,243	-,938	,073
	8	,129	,126	1,000	-,420	,677
	9	-,122	,157	1,000	-,805	,561
	10	,440	,144	1,000	-,188	1,068
	12	,098	,121	1,000	-,427	,623
	13	,131	,111	1,000	-,352	,615
	14	-,113	,124	1,000	-,651	,424
	15	,190	,123	1,000	-,347	,727
	16	-,023	,157	1,000	-,708	,661
	17	-,010	,122	1,000	-,542	,522
	18	-,189	,134	1,000	-,772	,394
	19	-,016	,115	1,000	-,518	,486
	20	,164	,143	1,000	-,458	,786
	21	-,016	,127	1,000	-,569	,538
	22	-,057	,131	1,000	-,629	,514
12	1	-,078	,154	1,000	-,749	,593
	2	-,134	,161	1,000	-,836	,568
	3	-,227	,130	1,000	-,794	,340
	4	,113	,173	1,000	-,640	,865
	5	-,018	,160	1,000	-,716	,680
	6	-,272	,159	1,000	-,963	,420
	7	-,530*	,114	,024	-1,028	-,033
	8	,031	,150	1,000	-,623	,686
	9	-,219	,149	1,000	-,869	,430
	10	,342	,146	1,000	-,294	,978
	11	-,098	,121	1,000	-,623	,427
	13	,034	,145	1,000	-,596	,663
	14	-,211	,124	1,000	-,749	,327
	15	,092	,152	1,000	-,570	,755
	16	-,121	,172	1,000	-,870	,628
	17	-,108	,138	1,000	-,706	,491
	18	-,287	,148	1,000	-,931	,357
	19	-,114	,114	1,000	-,608	,381
	20	,066	,116	1,000	-,438	,570
	21	-,113	,141	1,000	-,726	,499
	22	-,155	,147	1,000	-,796	,485
13	1	-,112	,117	1,000	-,622	,398
	2	-,167	,109	1,000	-,640	,305
	3	-,261	,133	1,000	-,841	,319
	4	,079	,143	1,000	-,545	,703
	5	-,052	,158	1,000	-,738	,634
	6	-,305	,138	1,000	-,908	,297
	7	-,564*	,126	,035	-1,110	-,017

	8	-,003	,140	1,000	-,610	,604
	9	-,253	,131	1,000	-,825	,318
	10	,309	,116	1,000	-,198	,815
	11	-,131	,111	1,000	-,615	,352
	12	-,034	,145	1,000	-,663	,596
	14	-,245	,094	1,000	-,651	,162
	15	,059	,085	1,000	-,312	,430
	16	-,155	,132	1,000	-,730	,420
	17	-,141	,115	1,000	-,641	,358
	18	-,321	,110	1,000	-,799	,158
	19	-,147	,086	1,000	-,521	,226
	20	,032	,122	1,000	-,498	,563
	21	-,147	,131	1,000	-,719	,425
	22	-,189	,117	1,000	-,697	,319
14	1	,133	,104	1,000	-,320	,586
	2	,077	,113	1,000	-,416	,570
	3	-,016	,128	1,000	-,575	,542
	4	,323	,151	1,000	-,332	,979
	5	,193	,149	1,000	-,454	,839
	6	-,061	,142	1,000	-,679	,558
	7	-,319	,119	1,000	-,836	,197
	8	,242	,156	1,000	-,435	,919
	9	-,009	,135	1,000	-,594	,577
	10	,553*	,122	,031	,022	1,084
	11	,113	,124	1,000	-,424	,651
	12	,211	,124	1,000	-,327	,749
	13	,245	,094	1,000	-,162	,651
	15	,303	,108	1,000	-,165	,772
	16	,090	,146	1,000	-,546	,726
	17	,103	,098	1,000	-,322	,529
	18	-,076	,110	1,000	-,556	,404
	19	,097	,081	1,000	-,254	,449
	20	,277	,121	1,000	-,251	,804
	21	,098	,123	1,000	-,438	,633
	22	,056	,111	1,000	-,429	,541
15	1	-,170	,099	1,000	-,601	,260
	2	-,226	,119	1,000	-,746	,293
	3	-,320	,125	1,000	-,862	,223
	4	,020	,141	1,000	-,595	,635
	5	-,111	,153	1,000	-,778	,557
	6	-,364	,106	,499	-,825	,097
	7	-,623*	,121	,007	-1,148	-,097
	8	-,061	,131	1,000	-,631	,509
	9	-,312	,129	1,000	-,872	,248
	10	,250	,128	1,000	-,305	,805
	11	-,190	,123	1,000	-,727	,347
	12	-,092	,152	1,000	-,755	,570
	13	-,059	,085	1,000	-,430	,312
	14	-,303	,108	1,000	-,772	,165
	16	-,213	,125	1,000	-,758	,331
	17	-,200	,097	1,000	-,622	,222
	18	-,379	,108	,420	-,850	,091
	19	-,206	,080	1,000	-,552	,140
	20	-,026	,105	1,000	-,482	,429
	21	-,206	,142	1,000	-,824	,413
	22	-,247	,107	1,000	-,711	,216
16	1	,043	,107	1,000	-,420	,506
	2	-,013	,112	1,000	-,498	,473
	3	-,106	,164	1,000	-,821	,609
	4	,234	,154	1,000	-,436	,903
	5	,103	,158	1,000	-,584	,790
	6	-,151	,132	1,000	-,725	,423
	7	-,409	,135	1,000	-,997	,179
	8	,152	,161	1,000	-,549	,853
	9	-,098	,154	1,000	-,769	,572
	10	,463	,176	1,000	-,302	1,229
	11	,023	,157	1,000	-,661	,708
	12	,121	,172	1,000	-,628	,870
	13	,155	,132	1,000	-,420	,730
	14	-,090	,146	1,000	-,726	,546
	15	,213	,125	1,000	-,331	,758
	17	,013	,147	1,000	-,626	,653
	18	-,166	,144	1,000	-,791	,460
	19	,007	,139	1,000	-,599	,614
	20	,187	,154	1,000	-,482	,856
	21	,008	,142	1,000	-,610	,626
	22	-,034	,164	1,000	-,746	,678
17	1	,030	,116	1,000	-,476	,535
	2	-,026	,121	1,000	-,554	,502
	3	-,120	,115	1,000	-,620	,381
	4	,220	,129	1,000	-,343	,783
	5	,090	,137	1,000	-,505	,684
	6	-,164	,122	1,000	-,693	,365
	7	-,422	,103	,097	-,872	,027
	8	,139	,143	1,000	-,483	,760
	9	-,112	,141	1,000	-,725	,501
	10	,450	,117	,177	-,058	,958
	11	,010	,122	1,000	-,522	,542

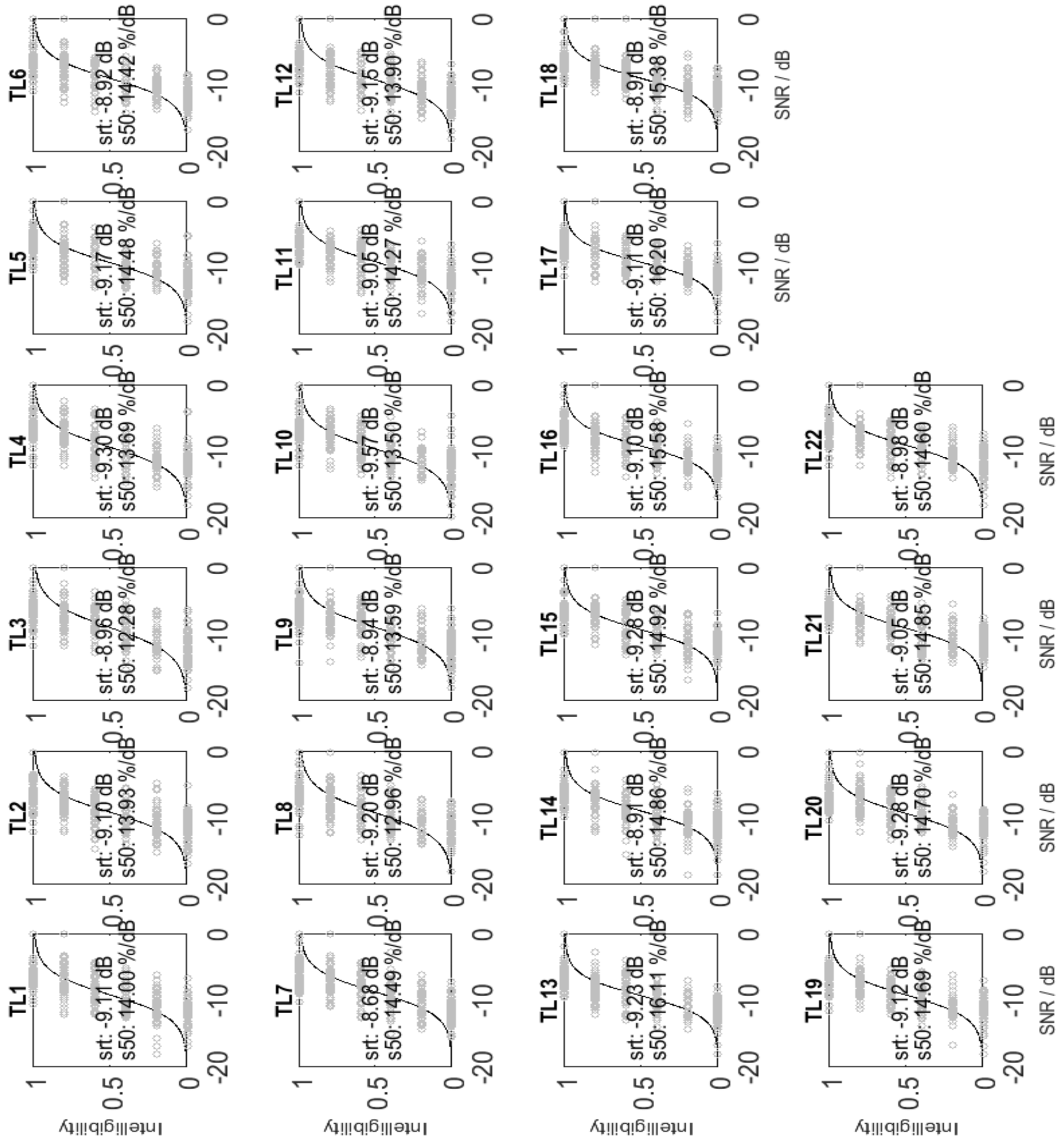
	12	,108	,138	1,000	-,491	,706
	13	,141	,115	1,000	-,358	,641
	14	-,103	,098	1,000	-,529	,322
	15	,200	,097	1,000	-,222	,622
	16	-,013	,147	1,000	-,653	,626
	18	-,179	,084	1,000	-,546	,188
	19	-,006	,089	1,000	-,393	,381
	20	,174	,103	1,000	-,273	,620
	21	-,006	,145	1,000	-,636	,624
	22	-,047	,098	1,000	-,472	,377
18	1	,209	,107	1,000	-,259	,676
	2	,153	,129	1,000	-,408	,714
	3	,060	,141	1,000	-,554	,673
	4	,399	,125	,887	-,144	,942
	5	,269	,151	1,000	-,390	,928
	6	,015	,139	1,000	-,589	,619
	7	-,243	,132	1,000	-,819	,333
	8	,318	,141	1,000	-,297	,933
	9	,067	,129	1,000	-,494	,629
	10	,629*	,118	,004	,114	1,144
	11	,189	,134	1,000	-,394	,772
	12	,287	,148	1,000	-,357	,931
	13	,321	,110	1,000	-,158	,799
	14	,076	,110	1,000	-,404	,556
	15	,379	,108	,420	-,091	,850
	16	,166	,144	1,000	-,460	,791
	17	,179	,084	1,000	-,188	,546
	19	,173	,091	1,000	-,223	,570
	20	,353	,108	,771	-,118	,824
	21	,174	,160	1,000	-,521	,868
	22	,132	,122	1,000	-,398	,661
19	1	,036	,099	1,000	-,394	,465
	2	-,020	,130	1,000	-,584	,544
	3	-,114	,096	1,000	-,532	,305
	4	,226	,140	1,000	-,385	,837
	5	,095	,158	1,000	-,590	,781
	6	-,158	,120	1,000	-,680	,364
	7	-,417	,102	,101	-,861	,028
	8	,145	,129	1,000	-,415	,705
	9	-,106	,126	1,000	-,655	,443
	10	,456	,110	,082	-,021	,933
	11	,016	,115	1,000	-,486	,518
	12	,114	,114	1,000	-,381	,608
	13	,147	,086	1,000	-,226	,521
	14	-,097	,081	1,000	-,449	,254
	15	,206	,080	1,000	-,140	,552
	16	-,007	,139	1,000	-,614	,599
	17	,006	,089	1,000	-,381	,393
	18	-,173	,091	1,000	-,570	,223
	20	,180	,088	1,000	-,204	,563
	21	,000	,138	1,000	-,599	,599
	22	-,042	,108	1,000	-,514	,431
20	1	-,144	,121	1,000	-,670	,382
	2	-,200	,130	1,000	-,767	,367
	3	-,293	,121	1,000	-,819	,233
	4	,046	,152	1,000	-,617	,710
	5	-,084	,141	1,000	-,697	,529
	6	-,338	,121	1,000	-,862	,187
	7	-,596*	,118	,008	-1,109	-,083
	8	-,035	,135	1,000	-,622	,552
	9	-,286	,106	1,000	-,748	,177
	10	,276	,128	1,000	-,280	,832
	11	-,164	,143	1,000	-,786	,458
	12	-,066	,116	1,000	-,570	,438
	13	-,032	,122	1,000	-,563	,498
	14	-,277	,121	1,000	-,804	,251
	15	,026	,105	1,000	-,429	,482
	16	-,187	,154	1,000	-,856	,482
	17	-,174	,103	1,000	-,620	,273
	18	-,353	,108	,771	-,824	,118
	19	-,180	,088	1,000	-,563	,204
	21	-,179	,148	1,000	-,823	,465
	22	-,221	,124	1,000	-,761	,318
21	1	,035	,140	1,000	-,573	,643
	2	-,021	,136	1,000	-,613	,572
	3	-,114	,160	1,000	-,809	,581
	4	,226	,161	1,000	-,475	,927
	5	,095	,149	1,000	-,552	,742
	6	-,158	,138	1,000	-,760	,443
	7	-,417	,133	1,000	-,994	,160
	8	,144	,169	1,000	-,590	,878
	9	-,106	,152	1,000	-,769	,557
	10	,456	,158	1,000	-,230	1,141
	11	,016	,127	1,000	-,538	,569
	12	,113	,141	1,000	-,499	,726
	13	,147	,131	1,000	-,425	,719
	14	-,098	,123	1,000	-,633	,438
	15	,206	,142	1,000	-,413	,824
	16	-,008	,142	1,000	-,626	,610
	17	,006	,145	1,000	-,624	,636
	18	-,174	,160	1,000	-,868	,521
	19	,000	,138	1,000	-,599	,599
	20	,179	,148	1,000	-,465	,823
	22	-,042	,134	1,000	-,627	,543
22	1	,077	,124	1,000	-,465	,619
	2	,021	,149	1,000	-,629	,672
	3	-,072	,108	1,000	-,543	,399
	4	,268	,129	1,000	-,296	,831
	5	,137	,147	1,000	-,504	,778
	6	-,117	,141	1,000	-,729	,496
	7	-,375	,115	,774	-,876	,126
	8	,186	,146	1,000	-,448	,820
	9	-,064	,150	1,000	-,718	,590
	10	,497	,130	,192	-,069	1,064
	11	,057	,131	1,000	-,514	,629
	12	,155	,147	1,000	-,485	,796
	13	,189	,117	1,000	-,319	,697
	14	-,056	,111	1,000	-,541	,429
	15	,247	,107	1,000	-,216	,711
	16	,034	,164	1,000	-,678	,746
	17	,047	,098	1,000	-,377	,472
	18	-,132	,122	1,000	-,661	,398
	19	,042	,108	1,000	-,431	,514
	20	,221	,124	1,000	-,318	,761
	21	,042	,134	1,000	-,543	,627

Basiert auf den geschätzten Randmitteln

*. Die mittlere Differenz ist auf dem ,05-Niveau signifikant.

b. Anpassung für Mehrfachvergleiche: Bonferroni.

G. Evaluation: Test list-specific functions



Bibliography

Abiratno, S. F. (2016a). *Audiometri Tutur dengan Kalimat Bahasa Indonesia: Penyusunan, Pembakuan, dan Penerapannya dalam Tes Audiometri Tutur di Tempat Ramai: IndoHINT, The Indonesian Version of Hearing in Noise Test (Speech Audiometry by means of Sentences in Indonesian:Construction, Validation and Application for Hearing in Background Noise: IndoHINT, The Indonesian Version of Hearing in Noise Test).* PhD thesis.

Abiratno, S. F. (2016b). Matrix test versi bahasa indonesia & koneksi ke indonesia. Email. Available Email: sfaisa@kasoem.co.id.

Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., Gagné, J.-P., Lutman, M., Wouters, J., Wong, L., et al. (2015). International collegium of rehabilitative audiology (icra) recommendations for the construction of multilingual speech tests: Icra working group on multilingual speech tests. *International journal of audiology*, 54(sup2):17–22.

Armandhani, H. (2012). Bahasa melayu akar dari bahasa indonesia(malay language, root of indonesian language). Website. Available online on `http://www.kompasiana.com/armandhani/bahasa-melayu-akar-dari-bahasa-indonesia_5518f21f81331193729de0be`; accessed on July 28th, 2016.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *Journal of Neuroscience*, 35(5):2161–2172.

Boenninghaus, H.-G. and Lenarz, T. (2006). *Hals-Nasen-Ohren-Heilkunde.* Springer-Verlag.

Brand, T. (1998). Personal information to K.Wagener.

Brand, T. and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6):2801–2810.

Brand, T., Wittkop, T., Wagener, K., and Kollmeier, B. (2004). Vergleich von oldenburger satztest und freiburger wörtertest als geschlossene versionen. 7. In *Comparison of the Oldenburg sentence test and the Freiburg word test as closed versions), Proceedings of the 7 th annual meeting of the Deutsche Gesellschaft für Audiologie (DGA), Leipzig.*

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., et al. (1994). An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, 96(4):2108–2120.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (2001). Icra noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment: Ruidos icra: Señates de ruido artificial con espectro similar al habla y propiedades temporales para pruebas de instrumentos auditivos. *Audiology*, 40(3):148–157.

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian audiology*, 11(2):79–87.

Hana (2010). Top 1000 nama populer indonesia. Website. Available online on `https://namafb.com/2010/08/12/top-1000-nama-populer-indonesia/`; accessed on May 11th, 2016.

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., and Kollmeier, B. (2012). A spanish matrix sentence test for assessing speech reception thresholds in noise. *International journal of audiology*, 51(7):536–544.

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vormès, E., Frachet, B., et al. (2012). Comparison of three types of french speech-in-noise tests: A multi-center study. *International journal of audiology*, 51(3):164–173.

Kartika, H. (2007). Tes pendengaran. Website. Available online on `https://hennykartika.wordpress.com/2007/12/29/tes-pendengaran/`.

Kollmeier, B. (1990). *Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache (Measurement methods, models, and improvement of the intelligibility of speech).* Faculty of Physics, University of Göttingen. Habilitation treatise.

Kollmeier, B., van de Par, S., Verhulst, S., Ewert, S., and Brand, T. (2014). *Module 1: Psychophysics and Audiology.* Institute of Medical Physics.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., and Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International journal of audiology*, 54(sup2):3–16.

Moeliono, A. M. and Dardjowidjojo, S. (1988). *Tata bahasa baku bahasa Indonesia (Standardized Grammar of Indonesian Language).* Departemen Pendidikan dan Kebudayaan, Republik Indonesia.

Muljana, S. (2008). *Kesadaran nasional: dari kolonialisme sampai kemerdekaan*, volume 1. PT LKiS Pelangi Aksara.

Muslich, M. (2008). *Fonologi bahasa Indonesia: tinjauan deskriptif sistem bunyi bahasa Indonesia (Indonesian Language Phonology: Descriptive Observation on Indonesian Language Sounding System).* Bumi Aksara.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, 95(2):1085–1099.

Ozimek, E., Warzybok, A., and Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International journal of audiology*, 49(6):444–454.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, 63(2):533–549.

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

Puglisi, G. E., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., and Kollmeier, B. (2015). An italian matrix sentence test for the evaluation of speech intelligibility in noise. *International journal of audiology*, 54(sup2):44–50.

Quasthoff, U., Fiedler, S., Hallsteinsdóttir, Kwary, D. A., and Goldhahn, D. (2015). *Frequency Dictionary Indonesian - Kamus Frekuensi Bahasa Indonesia.* Leipziger Universitätsverlag.

Radhi, F. (2012a). Tes pendengaran. Website. Available online on `http://publichealthnote.blogspot.de/2012/05/tes-pendengaran.html`; accessed on September 22th, 2016.

Radhi, F. (2012b). Tes pendengaran. Website. Available online on `http://publichealthnote.blogspot.de/2012/05/tes-pendengaran.html`.

Soewito (1985). *Audiometri Tutur Bahasa Indonesia: Penyusunan, Pembakuan, dan Penerapan Klinis Daftar Kata sebagai Alat Uji Pendengaran (Construction, Validation and Clinical Application of Word Lists as Hearing Test).* PhD thesis.

Soli, S. D. and Wong, L. L. (2008). Assessment of speech intelligibility in noise with the hearing in noise test. *International Journal of Audiology*, 47(6):356–361.

Statistik, B. P. (2011). Kewarganegaraan, suku bangsa, agama, dan bahasa sehari-hari penduduk indonesia: Hasil sensus penduduk 2010 (citizenship, ethnic group, religion, and daily language of indonesian population: Result of indonesian population census 2010). *Jakarta: BPS*.

Statistik, B. P. (2012). Penduduk indonesia/population of indonesia. *Hasil Sensus Penduduk 2000/Result of Indonesian Population Census 2010*.

Wagener, K. (2004). *Factors influencing sentence intelligibility in noise.* BIS Verlag.

Wagener, K., Brand, T., and Kollmeier, B. (1999a). Development and evaluation of a german sentence test part ii: Optimization of the oldenburg sentence test. *Zeitschrift Fur Audiologie*, 38:44–56.

Wagener, K., Brand, T., and Kollmeier, B. (1999b). Development and evaluation of a german sentence test part iii: Evaluation of the oldenburg sentence test. *Zeitschrift Fur Audiologie*, 38:86–95.

Wagener, K., Josvassen, J. L., and Ardenkjær, R. (2003). Design, optimization and evaluation of a danish sentence test in noise: Diseño, optimización y evaluación de la prueba danesa de frases en ruido. *International journal of audiology*, 42(1):10–17.

Wagener, K., Kühnel, V., and Kollmeier, B. (1999c). Development and evaluation of a german sentence test i: Design of the oldenburg sentence test. *Zeitschrift Fur Audiologie*, 38:4–15.

Wagener, K. C., Brand, T., and Kollmeier, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners: El papel de los intervalos de silencio para la inteligibilidad de frases en medio de ruido fluctuante en sujetos hipoacùsicos. *International Journal of Audiology*, 45(1):26–33.

Wardenga, N., Batsoulis, C., Wagener, K. C., Brand, T., Lenarz, T., and Maier, H. (2015). Do you hear the noise? the german matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment. *International journal of audiology*, 54(sup2):71–79.

Warzybok, A., Brand, T., Wagener, K. C., and Kollmeier, B. (2015a). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *International journal of audiology*, 54(sup2):88–99.

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., and Kollmeier, B. (2015b). Development of the russian matrix sentence test. *International journal of audiology*, 54(sup2):35–43.

Wesselkamp, M. (1994). *Messung und Modellierung der Verständlichkeit von Sprache*. na.

WHO (2017). Deafness and hearing loss (fact sheet). Website. Available online on `http://www.who.int/mediacentre/factsheets/fs300/en/`; accessed on March 12th, 2017.

WHO et al. (2007). Situation review and update on deafness, hearing loss and intervention programmes: proposed plans of actions for prevention and alleviation of hearing impairment in countries of the south-east asia region. *SEA-Deafness-10*.

Zokoll, M. A., Fidan, D., Türkyılmaz, D., Hochmuth, S., Ergenç, İ., Sennaroğlu, G., and Kollmeier, B. (2015). Development and evaluation of the turkish matrix sentence test. *International journal of audiology*, 54(sup2):51–61.

Statutory Declaration

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmitteln benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

Oldenburg, May 8, 2017

Felicia Primadita