Development of an Afrikaans Sentence Perception Test based on the CUNY topic-related Sentences – Phase 1: Sentence Perception in Noise

by
Jolanda Scourfield

Thesis presented in partial fulfilment of the requirements for the degree Master of Audiology at the University of Stellenbosch

Supervisors: Prof. Seppo Kalervo Tuomi
Mrs Alida Maria Uys Müller
Faculty of Health Sciences
Department of Speech- Language and Hearing Therapy

March 2011

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 17 February 2011

ABSTRACT

Speech audiometry for diagnostic purposes is widely used by audiologists around the world, but its application is starting to shift more towards rehabilitative purposes for people with hearing impairment. This has created the need for the development of appropriate test materials, including speech-in-noise tests, in the first language of the person undergoing rehabilitation.

This document describes a study entailing the first phase in the development of an Afrikaans sentence perception test in noise based on the City University of New York (CUNY) topic-related sentences (Boothroyd, Hanin & Hnath, 1985). The test is called *Sinslyste in Afrikaans vir Volwassenes in Lawaai* [*Sentence lists in Afrikaans for Adults in Noise*] (SAV-L). Twenty-seven sentence lists containing 12 sentences each were compiled and evaluated by eight speakers of Afrikaans for their naturalness before they were recorded together with three-talker babble as masking noise. The recorded material was then presented to six groups of ten participants each (with the exception of nine participants in one of the rounds) and adjusted in intensity to improve inter-list reliability. Two scorers were used at intervals to determine inter-rater reliability. Test conditions were also replicated after an interval of six months or more to establish test-retest reliability. In a last round of testing, the test was administered at the intended presentation level to assess the appropriateness of the chosen level.

Through adjustment of list intensities, inter-list reliability was improved to a 13.87% variance between list scores. Test-retest reliability showed a biase with an intra-class correlation agreement of 0.859. This was thought to be due to participants' improved familiariy with the clinical environment over time, however longer term measures of test-retest reliability remains to be done. Inter-rater reliability was very high with an intra-class agreement of 0.999. The intended test level of 50dBHL with a signal-to-noise (SNR) ratio of +5dB, was found to be an acceptable fixed SNR.

The test can be used in its current form for assessment of amplification candidacy as well as monitoring of progress during rehabilitation. Further research is indicated for the establishment of the sensitivity of the test. It is concluded that just as a clinician is expected to select the most suitable diagnostic test based on the patient's history, signs and symptoms, so should the audiologist be expected to select the most appropriate rehabilitative measure based on the patient's communication needs and established treatment goals. The SAV-L is a valuable contribution to the pool of speech perception tests available as rehabilitative measures, and is specifically suitable for adults with well-developed spoken language.

OPSOMMING

Spraakoudiometrie vir diagnostiese doeleindes word algemeen gebruik deur oudioloë regoor die wêreld, maar die toepassing daarvan is besig om te skuif na rehabilitasie van mense met gehoorgestremdheid. Die behoefte vir die ontwikkeling van geskikte toetsmateriaal, insluitend spraak-in-geraastoetse, in die eerste taal van die persoon wat rehabilitasie ondergaan, het dus begin ontwikkel.

Hierdie dokument beskryf 'n studie wat die eerste fase van die ontwikkeling van a Afrikaanse spraakpersepsietoets in geraas behels. Dit is gebaseer op die City University of New York (CUNY) onderwerp-verwante sinne (Boothroyd, Hanin & Hnath, 1985). Die ontwikkelde toets is benoem *Sinslyste in Afrikaans vir Volwassenes in Lawaai* (SAV-L). Sewe-en-twintig sinslyste met twaalf sinne per lys is saamgestel en geëvalueer deur ag Afrikaans-sprekende persone om hulle natuurlikheid te bepaal voordat dit opgeneem is tesame met drie-spreker babbel as maskeringsgeraas. Die opgeneemde materiaal is aan ses groepe van tien deelnemers elk aangebied (met die uitsondering van slegs nege deelnemers in een van die rondtes) en aangepas in intensiteit om inter-lys betroubaarheid to verbeter. Twee tellers is van tyd tot tyd gebruik om inter-beoordelaar betroubaarheid te bepaal. Toetskondisies was ook ná 'n periode van ses maande gereplikeer om toets-hertoets betroubaarheid to bepaal. In 'n laaste rondte van toetsing was die toets geadministreer teen die beoogde toetsvlak om die geskiktheid van hierdie toetsvlak te bepaal.
Inter-lys betroubaarheid is verbeter na 'n 13.87% variase tussen sinslystellings deur die aanpassing van die intensiteite van die lyste. Toets-hertoets betroubaarheid het 'n voorkeur met 'n intra-klas korrelasie ooreenstemming van 0.859 getoon. Hierdie voorkeur was vermoedelik weens die deelnemers se toenemende bekendheid met die kliniese omgewing, alhoewel langer termyn metings nog gedoen moet word om hierdie tendens verder te ondersoek. Inter-beoordelaar betroubaarheid was goed met 'n intra-klas korrelasie van 0.999. Die beoogde toetsvlak van 50dBGP met 'n sein-tot-ruis ratio van +5dB is as 'n geskikte ratio bevind.

Die huidige vorm van die toets kan gebruik word vir die evaluasie van kandidaatskap van klankversterking sowel as monitering van vordering tydens rehabilitasie. Verdere navorsing is aangedui vir die bepaling van die toets se sensitiwiteit. Daar is afgelei dat net soos daar van 'n klinikus verwag word om die mees geskikte diagnostiese toets te kies gebaseer op die pasiënt se geskiedenis, tekens en simptome, net so moet daar van die die oudioloog verwag word om die mees geskikte rehabilitasiemeting te selekteer, gebaseer op die pasiënt se kommunikasiebehoeftes en vasgestelde behandelingsdoelwitte. Die SAV-L is 'n waardevolle hulpmiddel in die versameling spraakpersepsietoetse wat beskikbaar is vir rehabilitasiemetings en is spesifiek geskik vir volwassenes met goed-ontwikkelde gesproke taal.

ACKNOWLEDGMENTS

I would like to thank the following people for the crucial part they played in the development of this study:

My heavenly **Father** who has given me the abilities, wisdom and strength to have undertaken this journey.

My amazing husband, **John Scourfield**, for his sacrifices, endless support, love, patience and encouragement.

My mum, **Rina Cloete**, who has always been behind me, building my foundations, believing in me.

Prof. Seppo Tuomi, connected to the department of Speech, Language and Hearing Therapy at the University of Stellenbosch, for his commitment, exceptional insight, support and time he sacrificed as supervisor.

Mrs Lida Müller, from the Tygerberg Hospital Cochlear Implant Team, for her valuable contributions as supervisor.

Martin Kidd, statistician at the University of Stellenbosch, for his assistance with the statistical analysis of the data.

The three assistant researchers, **Desiré le Roux**, **Lize-Marié Louw** and **Estie van Niekerk**, for their time, support and willingness to help.

Gert Mostert, studio engineer at Southern Buzz recording studio, for the recording and editing of the test material.

To my employers, A van der Merwe Inc., in particular **Nico van der Merwe Jnr**, for their support for my study, the use of their facilities, as well as for encouraging an environment of continuous professional development in which I could undertake my studies.

The **seventy-two test participants** involved in different parts of the study, for their willingness to assist.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS:

AB	Arthur Boothroyd
BEST	Beautifully Efficient Speech Test
BKB	Bamford-Kowal-Bench
BKB-SIN	Bamford-Kowal-Bench Speech-in-Noise test
APD	Auditory Processing Disorders
CASPERSENT	Computer-Assisted Speech Perception and Training at the Sentence Level
CD	Compact Disc
CI	Cochlear Implant
CID	Central Institute for the Deaf
CUNY	City University of New York
dBFS	Decibel relative to Full Scale
dBHL	Decibel Hearing Level
dBSPL	Decibel Sound Pressure Level
FDA	Food and Drug Administration
FG	Foneties Gebalanseerd [Phonetically Balanced]
HINT	Hearing in Noise Test
ICC	Intra-class correlation
MHAS	Modernising Hearing Aid Services
NIST	National Institute of Standards and Technology
NVLAP	National Voluntary Laboratory, Accreditation Program
PI	Performance-intensity
POCIA	Predicting and monitoring Outcomes from Cochlear Implantation in Adults

Quick-SIN	Quick Speech-in-Noise test
RCA	Radio Corporation of America
RMS	Root Mean Square
SABS	South African Bureau for Standards
SANS	SysAdmin, Audit, Network, Security
SAV-L	Sinslyste in Afrikaans vir Volwassenes – Lawaai [Sentence lists in Afrikaans for Adults – Noise]
SAV-S	Sinslyste in Afrikaans vir Volwassenes – Stilte [Sentence lists in Afrikaans for Adults – Quiet]
SNR	Signal-to-Noise Ratio
SRT	Speech Reception Threshold
VU	Volume Units

CHAPTER OUTLINE

Chapter 1 gives an introduction to speech audiometry and its current and future uses and trends. It also identifies needs in speech audiometric tests to be developed, current tests already available and the rationale for the study undertaken. A description of terminology used completes the chapter.

Chapter 2 describes in detail the components to consider when developing a speech test and applies them to the test developed during this study. The chapter ends with the research question for the current study.

Chapter 3 describes the methodology for the current study while **Chapter 4** contains the statistical results and outcomes thereof.

Chapter 5 contains a discussion of further considerations, recommendations for further research and a conclusion to the study.

CHAPTER 1

SPEECH AUDIOMETRY: OVERVIEW

1.1. INTRODUCTION

Speech audiometry is frequently used by audiologists worldwide. It is most widely applied for diagnostic purposes but in recent years the focus of its applications has shifted more towards use during rehabilitation of people with hearing impairment. This shift has created a new interest and urgency to develop relevant, reliable test materials that would be suitable specifically for rehabilitative use.

Speech audiometry is defined as “any method for assessing the state or ability of the auditory system of an individual, using speech sounds as the response evoking stimuli” (Lyregaard, 1997, p. 35). By definition the field covers an extremely large area. This is demonstrated by its numerous applications and the number of variables used in choosing and developing a speech perception test. Tobin (1978, as cited in Hannley, 1986, p. 149) illustrated this variability using a three-way matrix, which demonstrated combinations of speech stimuli (e.g. syllables, words and sentences) versus response tasks (e.g. detection, discrimination and recognition tasks) versus communication conditions (e.g. quiet versus noise, and auditory-only versus visual-only clues). The matrix ultimately resulted in 288 possibilities of combinations (Figure 1.1). This complexity may explain why, despite existence of standardised procedures for speech audiometry as well as requirements on the recordings used (SANS 8253-3, Standards South Africa, 1996), no attempts have been made towards international standardisation of speech perception test material itself (Martin, 1997,

p. vii). Lyregaard (1997, p. 36) correctly described speech perception tests as "notoriously difficult to specify and standardize".

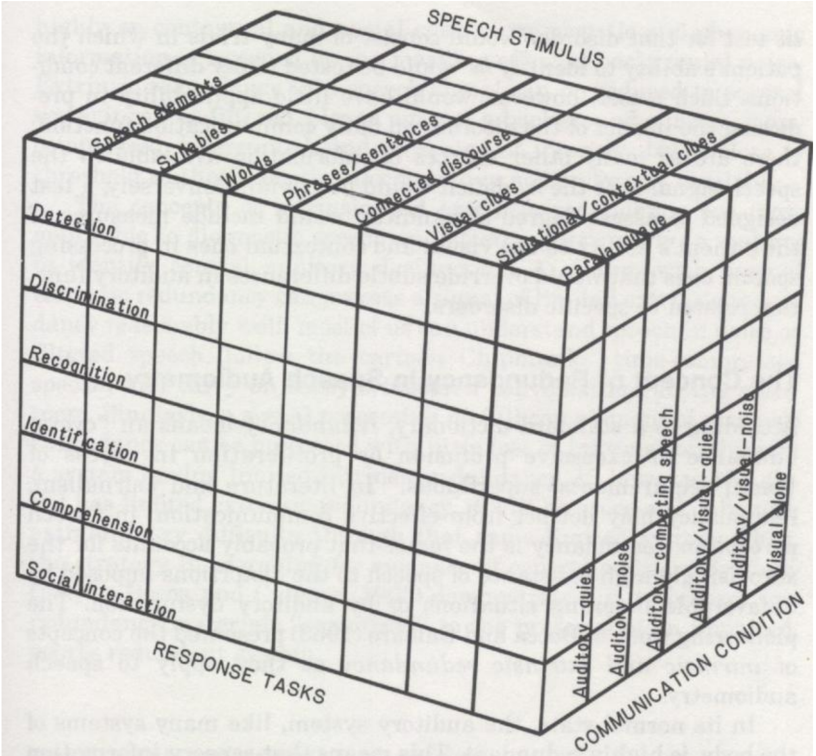


Figure 1.1. Factors influencing speech processing (Tobin, 1978, in Hannley, 1986, p. 149)

1.2. CURRENT USES OF SPEECH AUDIOMETRY IN THE FIELD OF AUDIOLOGY

In the USA 99% of audiologists reportedly use speech perception tests consisting of single words (Gelfand, 1998, p. 1088). In contrast to this high number, less than 6% use sentence materials to conduct aided speech perception testing (Mueller, 2001, p. 20).

In South Africa speech audiometry is also widely used and normally forms part of the basic test battery that audiologists use. However, as in the USA, sentence material is rarely used (Roets, 2006, p. 84). Data on the use of speech audiometry imply that audiologists in South Africa use monosyllabic and spondee words as stimuli most of the time (Roets, 2006, p. 84). Tests with words as stimuli are generally used for diagnostic purposes: spondee words are used to confirm the pure tone audiogram during speech recognition threshold testing and monosyllabic words are used to help determine the site of lesion during word identification testing. Although speech perception testing is therefore widely used in South Africa, it is mostly used for diagnostic purposes and currently used very sparingly to evaluate and monitor auditory devices, i.e. for rehabilitative purposes.

The exception to this would be in cochlear implant teams, where it is acknowledged that although the pure-tone average can provide an indication of hearing impairment, the individual's access to speech with auditory devices is seen as more important (Niparko, Lingua & Carpenter, 2009, p. 138). Aided speech perception testing is therefore done routinely to determine cochlear implant candidacy.

Only 42% of hearing aid dispensing professionals in the USA use some type of speech-in-noise task to evaluate their patients' ability to understand speech in background noise (Strom, 2006, ¶74). Data with regards to the use of speech-in-noise tests in South Africa is not available. In a recent survey regarding the use of speech audiometry in South Africa (Roets, 2006), speech-in-noise tasks were considered to be used only for the assessment of auditory processing problems and therefore excluded from the data collection (Roets, 2006, p. 98). The use of speech-in-noise tasks therefore seems very limited, despite the importance of measuring speech perception in background noise which has long been highlighted as an ecological measure of communication disability (Carhart and Tillman, 1970, p. 279). Hoen (2010, slide 3) argued that to test natural speech comprehension, speech in noise performance must be tested, since communication hardly ever takes place in complete silence.

In instances where speech-in-noise tests are administered, most audiologists will assess their patients using single words in noise. This is heading in the right direction, since difficulty hearing speech in noise is the most common complaint that hearing aid wearers and people with hearing loss report (McArdle & Hnath-Chisolm, 2009, p. 72; Vaillancourt et al., 2005, p. 358; Wilson, Burks & Weakley, 2005, p. 499; Killion, Niquette, Gudmundsen, Revit & Banarjee, 2004, p. 2395; Kramer, Kapteyn & Festen, 1998, p. 310; Plomp, 1994, p. 2). However, single words are not representative of normal conversation and have very low face validity when assessing a person's everyday communication difficulties.

1.3. EVOLUTION AND POSSIBLE FUTURE TRENDS IN SPEECH AUDIOMETRY

Speech perception testing probably started in 1910, when Campbell developed tests using nonsense syllables to assess telephone channels. In 1947 Hudgins, Hawkins, Karlin & Stevens developed the first recorded auditory test to determine thresholds for speech (Lucks Mendel & Danhauer, 1997, p. 2). Although speech perception testing was initially used to compare different hearing aids, i.e. for rehabilitation purposes, it was not believed to be a very good tool for comparative evaluation. As a result, aided speech perception testing started to disappear in the 1980s (Mueller, 2001, p. 19). The assessment of clarity of speech (McArdle & Hnath-Chisolm, 2009, p. 64), i.e. communication in daily life, therefore started to diminish.

The diagnostic aspect of speech audiometry serves to assess sensitivity for speech (McArdle & Hnath-Chisolm, 2009, p. 64), i.e. hearing thresholds. This aspect remains a part of the routine test battery.

Recently the direction of speech perception test applications has however slowly started to shift back from being a diagnostic tool to being a rehabilitation tool, aimed specifically for the assessment and monitoring of a patient's auditory disability pre- and post-fitting of hearing aids or cochlear implants (Mueller, 2001, p. 24). More than a decade ago Gatehouse and Robinson (1997, p. 80) predicted this shift. They attributed this prediction to the opinion that diagnostic behavioural testing in audiology will decline as electrophysiological techniques develop, leaving the main aim of speech perception testing to develop into that of assessing communication problems experienced by hearing impaired listeners.

The increased use of speech perception tests to evaluate the effectiveness of communication in daily life is inevitable, since it is well documented that pure tone audiometry, which correlates with speech reception thresholds using spondee words, cannot predict speech understanding (Taylor, 2003, p.42; Killion & Niquette, 2000, p. 46; Killion, 1997, p. 28). Pure-tone audiometry only measures the audibility component of a hearing loss (Wilson et al., 2005, p. 504) and cannot predict successful use of amplification (Walden & Walden, 2004, p. 344) or signal-to-noise ratio (SNR) loss. Without the knowledge of a listener's SNR loss, there is virtually no way of predicting the benefit a person would obtain from hearing aids when used in noise (Killion et al., 2004, p. 2395). SNR loss is influenced by the different levels of cochlear inner hair cell loss (Killion, 1997, p. 28) and therefore varies greatly in persons with similar pure tone hearing losses (Killion & Niquette, 2000, p. 52; Killion, 1997, p. 28). Only measures, such as speech-in-noise testing, reflect the distortion component of hearing loss (Wilson, et al., 2005, p. 504). The only way therefore to determine how well an individual hears in noise and to establish the SNR loss, is to actually measure it (Killion & Niquette, 2000, p. 46).

As a rehabilitation tool speech perception tests with high face validity should be used, including tests that assess the patient's ability to understand speech in background noise (Vaillancourt et al., 2005, p. 358). It is this measurement that provides useful information for selecting an amplification strategy and for counseling patients with regards to expectations during audiological rehabilitation (Wilson et al., 2005, p. 504; Wilson & McArdle, 2005, p. 82), as it provides a percentage score, which is more easily understood by patients than degree of hearing loss (Taylor, 2003, p. 40). Speech perception testing during rehabilitation can also be applied to establish appropriateness of binaural fitting, to predict aided outcomes, to demonstrate to the patients their degree of impairment in comparison to a person with normal hearing, and to demonstrate aided performance (Mueller, 2001, p. 24). Speech perception tests as the main indicator in establishing whether a person would obtain more benefit with monaural or binaural fitting, is an especially important consideration since, although it is widely accepted that binaural aiding is what is most appropriate for bilateral hearing loss, some individuals actually get more benefit from monaural aiding (Walden, 2006, p. 318). Establishment of optimal benefit conditions greatly affects the effectiveness and financial implications of audiological rehabilitation programmes. The availability of appropriate tests with high face validity can therefore potentially provide a whole new dimension in audiological rehabilitation.

As a person's ability to communicate in everyday situations is what should ultimately be assessed (Clark, 2003, p. 708), the areas of speech perception testing that will therefore most likely develop are the assessment of the degree of auditory disability and the extent to which a disability is reduced by intervention.

1.4. CURRENT AND FUTURE DEVELOPMENTS

With the shift in use of speech perception tests towards assessing auditory disability in the rehabilitation process, there has been a recent surge in the development of standardised sentence tests in quiet and in noise (Wong, Soli, Liu, Hang & Huang, 2007; Hällgren, Larsby & Arlinger, 2006; Wong & Soli, 2005; Vaillancourt et al., 2005; Killion et al., 2004; Versveld, Daalder, Festen & Houtgast, 2000; Kollmeier & Wesselkamp, 1997; Nilsson, Soli & Sullivan, 1994; MacLeod & Summerfield, 1990). These tests using sentences as stimuli have much higher face validity than tests using single words as stimuli.

Speech perception tests, that use sentences as stimuli, are already routinely used in cochlear implant rehabilitation to help assess candidacy for implantation and to quantify the benefit of the cochlear prosthesis (Clark, 2003, p. 708). The Food and Drug Administration (FDA) in the USA has set clear guidelines, based on word identification scores in a sentence context, for candidacy of unilateral and bilateral implantation (Kirk & Choi, 2009, p. 193; Herzog & Rehwinkel, 2008, p. 52). Using sentence testing during evaluation is a logical necessity since the single most significant benefit of cochlear implantation is improved speech perception (Kirk & Choi, 2009, p. 193).

Speech perception testing is anticipated to become part of hearing aid evaluation as well. Kruger and Kruger (1994, p. 303) predicted that prescriptive methods, real-ear probe measurements, speech-based comparative methods and benefit determination methods would all be integrated into future hearing aid selection, fitting and verification procedures and systems. This prediction has been realised already for all these aspects except speech-based comparative methods. This aspect is also expected to expand in the near future, since outcome-based treatment is becoming progressively more important, not only in the field of audiology, but in the medical profession as a whole. With the development of the Modernising Hearing Aid Services (MHAS) initiative in the United Kingdom for example, great emphasis was placed on the verification of hearing aid fitting, which is considered a minimum routine requirement during hearing aid fitting (Royal National Institute for the Deaf, 2002, p. 13). Objective verification is now routinely done in the form of real-ear measurements, while it is also done subjectively through informal live voice testing. More formal speech perception testing in addition is indicated as part of the outcome-based culture, since understanding running speech is the only predictor of how well a person would do outside the clinical environment.

1.5. NEEDS IN THE DEVELOPMENT OF SPEECH AUDIOMETRIC TESTS IN SOUTH AFRICA

In South Africa new tests are needed to keep up to date with what Flynn (2003, ¶1) describes as "Sailing out of the windless sea of monosyllables" towards tests for rehabilitation purposes in the appropriate language(s).

Gat and Keith (1978, as cited in Gelfand, 2009, p. 261) noted that non-native speakers of a language typically obtained lower scores on speech recognition tests than did native speakers of the language. This was demonstrated by Danhauer, Crawford & Edgerton (1984, p. 164), when speakers of Spanish scored lower than bilingual speakers on a nonsense syllable test using English phonemes. Gelfand (2009, p. 261) rightly noted that audiologists everywhere face the same problem, only with different local languages. In South Africa appropriate speech perception testing material will therefore need to be available in all the official languages. With South Africa's 11 official languages (Afrikaans, English, Ndebele, Pedi, Sotho, Swazi, Tsonga, Tswana, Venda, Xhosa and Zulu, Statistics South Africa, 2001), the development of relevant speech material is truly in its infancy.

As the task of developing suitable speech perception test material in all these languages requires a great effort, the development of speech audiometry materials in South Africa needs to be prioritised. On the one hand, widely used tests which form part of the basic audiological test battery, need to be developed in the appropriate languages. On the other hand, tests for the assessment of auditory disability, whether for the assessment of suitability of auditory device candidacy or for monitoring of ongoing improvement post-provision of auditory devices as a means to assess residual disability, also need to be developed. The immediate effort would be more wisely spent in developing tests to assess auditory device candidacy and disability, in order to accommodate current developments in the field of audiology, rather than spending time and resources on diagnostic tests, which are being replaced more and more by electrophysiological tests (Gatehouse & Robinson, 1997, p. 80). It would be logical to start this process with the most widely spoken languages, such as Zulu (home language of 23.8% of the population), Xhosa (home language of 17.6% of the population), Afrikaans (home language of 13.3% of the population and the dominant language in the Western and Northern Cape) and English (spoken by 8.3% of the population and seen as the language of business, politics and the media and generally understood across the country) (Statistics South Africa, 2001).

1.6. TESTS ALREADY AVAILABLE FOR REHABILITATION PURPOSES IN SOUTH AFRICA

A few sentence tests are already used in South Africa to assess cochlear implant candidacy and to monitor progress. The tests currently used for testing children, are recordings of the Bench-Kowal-Bamford (BKB) sentences (Bench, Kowal & Bamford, 1979) and its Afrikaans translation (Cloete, 1997). For adults, recordings of the Central Institute for the Deaf (CID) sentences (Silverman & Hirsh, 1955) are available in English and Afrikaans (Müller & De Stadler, 1989).

Because cochlear implant rehabilitation requires repeated assessment of improvement of performance, there is a need for a large number of sentence lists.

The ten test lists of the CID sentence test are not sufficient, since practice-effects increase as the number of stimuli decreases (Lucks Mendel & Danhauer, 1997, p. 23), i.e. participant scores improve if sentence lists are repeated too often. This means that sentences cannot be repeated at regular intervals in view of their high learning rate (Boothroyd, 2006, p. 3; Gatehouse & Robinson, 1997, p. 82; Lucks Mendel & Danhauer, 1997, p. 20; Olsen & Matkin, 1991, p. 73). The CID sentence test is also an old test and may, like many of the tests developed before the mid 1960's, not be based on the information that is now known about constructing and administering sensitive tests of speech perception (Lucks Mendel & Danhauer, 1997, p. 60).

In addition to the CID sentence test in English and Afrikaans, a Xhosa sentence test is also available (Olivier, 2000). This is a sentence-in-noise test based on the BKB sentence lists, consisting of a recording of Xhosa sentences read by a female talker in the presence of six-talker babble. An Afrikaans sentence test in noise was also recently compiled (Theunissen, 2008), which is loosely based on the principles of the Hearing in Noise Test (HINT) (Nilsson, Soli & Sullivan, 1994). It is an adaptive signal-to-noise test recorded by a female speaker and consists of 20 sentence lists with 10 sentences per list, presented in the presence of speech-weighted noise.

Although the recently developed Afrikaans sentence test (Theunissen, 2008) is methodologically sound, it uses sentences with a simple structure, similar to the HINT sentences which were based on the BKB sentences, originally developed to test children. The simplicity of the sentence structures allows for a reliable adaptive signal-to-noise test. It does however reduce the face validity of the test when used to assess adults, because if test material is too easy, the participant will over-perform (Madell, 2010, slide 21; Madell & Flexer, 2008, p. 101). The choice of speech-weighted noise as masking noise was also identified by Theunissen (2008, p. 222) as a possible limitation of the study. It is believed that using modulated noise, such as multi-talker babble during testing, leads to a higher correlation between self-reported disability and test results than when steady noise, like speech-weighted noise, is used during sentence-in-noise testing (Tyler & Smithe, 1983, as cited in Lutman, 1997, p. 69). This is possibly due to the fact that multi-talker babble masks speech more effectively than when a steady noise is used (Lutman, 1997, p. 69). Consequently this is the preferred masking noise for testing bimodal and bilateral speech processing (Clark, 2003, p. 718), with the best face validity (Plant, 1997, p. 228) as it represents the everyday situation most closely (Clark, 2003, p. 718).

1.7. RATIONALE FOR THE CURRENT STUDY

In view of the limited number of tests available for the assessment and rehabilitation of adult cochlear implant candidates and users, cochlear implant teams in South Africa face the dilemma of not having suitable speech perception tests in all the appropriate languages. Although there are a few Afrikaans sentence tests available in South Africa, in the Western Cape, where Afrikaans is the most widely spoken language (Statistics South Africa, 2001), a strong need remained to develop a clinically feasible and acceptable sentence test in Afrikaans. Such a test would provide a larger pool of sentence lists to be used during rehabilitation over an extensive period of time. Although the test developed in the current study was specifically developed in view of the need identified for such a test for cochlear implantees, the number of lists also makes it suitable for monitoring progress in hearing aid users, since the validity of benefit data obtained very soon after hearing aid fitting is questioned (Cox, 1993, p. 300). Gatehouse (1992, p. 1266) found that laboratory measures of hearing aid benefit improved over a period of 12 weeks, indicating that the adjustment period to hearing aid use is long, therefore the need for long-term monitoring of progress is also present in hearing aid users.

For the current study, the City University of New York (CUNY) topic-related sentences (Boothroyd, Hanin & Hnath, 1985) were chosen as the model to develop an Afrikaans sentence test. The CUNY topic-related sentences are already frequently used in cochlear implant rehabilitation abroad (Zwolan, 2009, p. 922) as well as in the *Predicting and monitoring Outcomes from Cochlear Implantation in Adults* (POCIA) system, a software and hardware system used by cochlear implant teams in the United Kingdom and Ireland for testing profoundly deaf adults (Medical Research Council, 2008). The test was particularly attractive, because of its apparent high face validity for testing adults. It can therefore potentially give a good estimation of the level of difficulty that adults with hearing impairment may have in daily communication, which is an important aim of speech perception testing (Flynn, 2003, ¶4). Several characteristics give it its high face validity. Firstly it consists of sentences of varying lengths and types (statements, commands and questions). This feature allows for a quantitative as well as qualitative assessment of errors, resulting in the examination of independent variables. Each list also has different topics, which allow closer simulation of conversation, where the topic can be made known to the listener prior to the presentation of a sentence. This can also be a potential disadvantage of the test, if topics used are unfamiliar to the participant (R.S. Tyler, personal communication, October 6, 2009). However, the topics were carefully selected to be familiar to the average adult in the relevant culture. For ease of reference, the developed test was named the *Sinslyste in Afrikaans vir Volwassenes in Lawaai* [Sentence lists in Afrikaans for Adults in Noise] (SAV-L).

1.8. DEFINITIONS OF TERMS

Terminologies used for the purpose of this study are defined as follows:

- ***Aided and unaided testing:*** aided audiological testing is the assessment of a participant's hearing whilst wearing auditory devices, whereas unaided testing is assessment without auditory devices. The purpose of unaided testing is to evaluate an individual's hearing/communication loss, whilst aided testing assesses residual hearing/communication loss following supply of auditory devices.

- ***Audiological rehabilitation:*** for the purpose of this document, the term *audiological rehabilitation* refers to the process of assessment of auditory device candidacy and the monitoring of ongoing improvement. It does not include auditory training. In fact, the use of an assessment tool such as the SAV-L is discouraged for use during auditory training in order to avoid learning the material, which would boost performance scores during ongoing assessment.

- ***Auditory device:*** for the purpose of this document the term *auditory device* is used as a general word for a hearing aid and/or cochlear implant.

- ***Bottom-up processing:*** the coding of the acoustical signal in the peripheral auditory system, and the transfer and processing through the central auditory system into the auditory cortex (Kramer, Zekveld & Houtgast, 2009, p. 507).

- ***Cocktail party problem:*** the challenge of segregating a single target voice from a competing milieu (Cherry, 1953, p. 976).

- ***Cold running speech:*** a selection of prose material with relatively uninteresting contents (Gelfand, 2009, p. 247).

- ***Decibel relative to full scale (dBFS):*** decibel amplitude levels in digital systems which have a maximum available peak level. The maximum peak level is 0dBFS with softer levels being of a negative value ("Understanding dB"). Levels above 0dBFS will cause distortion (Katz, 2002, p. 62).

- ***Disability:*** an umbrella term, covering impairments, activity limitations, and participation restrictions (World Health Organisation, n.d.).

- ***Energetic masking:*** the uncertainty or complete loss of information due to the masker being sufficiently intense (Brungart, Simpson, Ericson & Scott, 2001, p. 2528; Garcia Lecumberri & Cooke, 2006, p. 2446).

- ***Handicap:*** restrictions in participation in usual roles (Mathers, Sadana, Salomon, Murray & Lopez, 2000, p. 1),

- ***Informational masking:*** the uncertainty created for the listener as to whether he/she is hearing the target or the masker, when enough talkers in the babble noise are used to maintain linguistic content (McArdle & Hnath-Chisolm, 2009, p. 75; Hawley, Litovsky & Culling, 2004, p. 834).

- ***Masking:*** masking occurs if one sound (the masker) makes an acoustical signal inaudible (Roeser & Clark, 2007, p. 262).

- ***Multi-talker babble:*** the summed waveform of several simultaneous talkers (Simpson & Cooke, 2005, p. 2775).

- ***Performance-Intensity (PI) function:*** A graph of the change in percent correct score (the dependent variable, plotted on the y-axis) based on changes of signal-to-noise ratio or intensity (the independent variable, plotted on the x-axis) (McArdle & Hnath-Chisolm, 2009, p. 66). It is usually S-shaped, with the steepest rise being the most sensitive area.

- ***Proxemics:*** the study of how man unconsciously structures microspace, i.e. the physical distance between persons in the conduct of daily transactions (Lange-Seidl, 1977, p. 17).

- ***Root mean square (RMS):*** a method of averaging levels of a recording which computes the equivalent power of the material. The RMS-responding meter will read several dB below the actual peak level at any moment in time for all naturally-occurring music (Katz, 2002, p. 309). The same is true for recordings of speech with its natural level variations.

- **Signal-*to-noise ratio (SNR):*** The difference in dB between a sound of interest and background noise (Stach, 2010, p. 361).

- ***Signal-to-noise ratio (SNR) loss:*** The increase in signal-to-noise ratio required for a listener to obtain 50% correct words, sentences, or words in sentences, compared to normal performance (Killion et al., 2004, p. 2395). This is therefore a comparison between how well a person with hearing impairment versus a person with normal hearing would understand speech in noise (Taylor, 2003, p. 41).

- ***Speech identification:*** identifying speech where the alternatives are unspecified, i.e. open-ended (Hannley, 1986, p. 152).

- ***Speech perception:*** the decoding of the acoustic dimensions of a speech signal, which is thought to depend on the use of language processes (Escudero, 2009, p. 154). The ultimate goal of speech perception is to determine the meaning behind the spoken message (Boothroyd, 1998, p. 103).

- ***Speech perception test:*** for the scope of this document, the term 'speech perception test' is a generic term for any audiological test that uses speech as its stimulus.

- ***Speech reception threshold (SRT):*** spondee words are presented to determine the lowest level at which participants can correctly repeat or understand 50% of the stimuli (i.e. the threshold of speech) (Lucks Mendel & Danhauer, 1997, p. 64).

- ***Speech recognition:*** selecting a word from a list of alternatives (Hannley, 1986, p. 151), i.e. from a closed set of options.

- ***Speech-weighted (speech spectrum) noise:*** white noise filtered to a low- and middle-frequency band, simulating the long-term average speech spectrum of conversational speech (Sanders, 1991, p. 173).

- ***Standard variety of a language:*** a form of the language that becomes the 'model' to which language use is 'measured'. It is established through three criteria: a) the *status* of the involved language community (usually that of leading members of the community); b) the *form* of the variety, i.e. whether the form has been adjusted to meet the needs of the language community and c) the *use* of the variety for standard language functions e.g. teaching, church, media, sports and culture (Carstens, 2003, p. 283).

- ***Standardised* test:** a test which has a high degree of reliability (Bench & Bamford, 1979, p. 475).

- ***Top-down processing:*** the influence of information at higher neurocognitive processing levels on lower levels (Carroll, 2008, p. 56). This information includes language, attention, memory, cognition and auditory processing (Bellis, 2003, pp. 53 & 54).

- ***Truncation (floor and ceiling) effects:*** where scores cluster around the top or bottom end of a scale, change can only be detected in one direction, which can introduce a bias in the nature of the observed change (De Vaus, 2005, p. 110). An example of a ceiling effect would be when two participants score 100% in a speech-in-noise test, one might have had to put more effort into obtaining this score than the other, although the effort is not reflected in the score. In the same manner a floor effect is where two participants for instance scored 0% in a speech-in-noise task, again, not reflecting whether their abilities were indeed similar.

- ***Working memory:*** a system for the temporary storage and manipulation of information (Kramer et.al., 2009, p. 507)

1.9. SUMMARY

Chapter 1 introduced the concept of speech audiometry in audiology. It also focused on its current uses and future trends. The chapter was completed with a discussion of speech perception tests that are already in use in South Africa, the rationale for the study undertaken as well as a definition of terminology used.

CHAPTER 2

COMPONENTS OF SPEECH PERCEPTION TESTS CONSIDERED IN THE CURRENT STUDY

2.1. INTRODUCTION

This chapter provides an overview of the aspects that need to be considered when developing or evaluating a test of speech perception for audiometric use. It also explains arguments that led to the choice of each variable used in the development of the SAV-L. It concludes with the research question.

2.2. MAJOR COMPONENTS OF A SPEECH PERCEPTION TEST

When a speech perception test is developed, or when different speech perception tests are compared, a vast array of relevant variables needs to be considered. To allow for a systematic overview of the components considered during the development of the SAV-L, the major components or variables of speech perception tests as suggested by Lyregaard et al. (1976, as cited in Sherwood & Fuller, 1997, p. 90) were considered. These variables can be grouped into stimulus-, patient- and results-related variables as set out in Figure 2.1. The variables will consequently be discussed together with reasons for the choice of each variable made for the SAV-L.

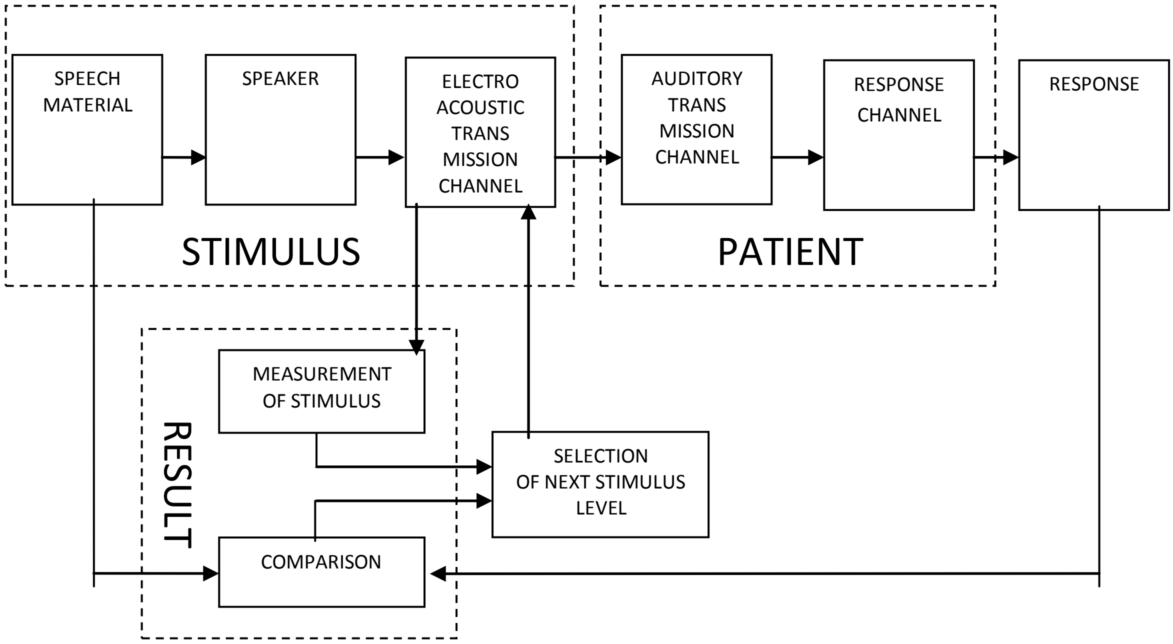


Figure 2.1. Components of speech perception tests (Lyregaard et al., 1976, as cited in Sherwood & Fuller, 1997, p. 90)

2.2.1. Stimulus components

Stimulus components refer to the type of speech material used and its characteristics, the speaker variables as well as the delivery mode, (i.e. the electro-acoustic transmission channel). In the following discussion these three aspects will be used as the frame for discussion.

2.2.1.1. Speech material

The actual speech material constitutes the most extensive component of a speech perception test. It will be discussed using the following 11 categories: psychometric considerations, stimulus type, phonemic balancing, word familiarity, number of test items, test technique, sentence list equivalence, establishing normative values, masking type, open versus closed set responses and use of a carrier phrase.

2.2.1.1.1. Psychometric considerations

Mackersie (2002, p. 392) noted that "an ideal speech perception test will be *reliable*, will be highly *sensitive* to differences between test conditions, and will correlate well with speech perception abilities in the real world", i.e. will have high *validity* (italics added). These three concepts (sensitivity, reliability and validity), which should be a significant consideration during the selection of appropriate speech material, are discussed below.

Sensitivity

Sensitivity of a speech perception test refers to "how objectively and accurately a test measures aspects of listeners' speech perception abilities as a reflection of their performance in realistic listening situations" (Lucks Mendel & Danhauer, 1997, p. 7). It is especially relevant in the assessment and monitoring of auditory device candidacy, since in the first instance, this will reflect a person's disability, which will allow an informed decision with regards to the need for intervention. High sensitivity is present if a small change in the experimental conditions causes a large and repeatable change in test scores (Dillon, 1983, p. 336).

Sensitivity in speech perception test material is maximised if all the items in every list are of equal difficulty (Dillon, 1983, p. 343), the lists themselves are of equal difficulty (Bamford & Wilson, 1979, p. 149), the long-term average speech spectrum matches that of the noise and the stimuli have highly redundant information (Dillon & Ching, 1995, p. 321). These sensitivity measures need to be balanced with the test's validity measures, especially its ecological validity, which depicts how close the test is to natural listening environments (Mackersie, 2002, p. 395). Some sentence tests, e.g. the HINT (Nilsson et al., 1994), have sentences of equal length, a requirement for a highly sensitive test. However, this feature makes it less representative of everyday conversation in which sentence lengths vary considerably. In the development of the SAV-L, the sentence lengths were not kept constant, in order to maintain high ecological validity. This meant that all items in every list would not be of equal difficulty. Other measures aimed at maintaining the sensitivity of the test, like keeping the lists equally difficult, therefore had to carry more weight.

To ensure high sensitivity, the masking noise was also carefully considered, in an effort to match the long-term average speech spectrum (Dillon & Ching, 1995, p. 321). Many tests of speech perception, e.g. the HINT (Nilsson et al., 1994) and the Afrikaans speech recognition test (Theunissen, 2008), use speech spectrum noise, which contributes to increased sensitivity, since speech spectrum noise is more efficient than white noise with a masking advantage of about 8dB (Sanders, 1991, p. 173). For the SAV-L, however, it was decided to use multi-talker babble in order to keep sensitivity as well as the ecological validity high, since multi-talker babble is the most 'natural' masking noise.

In addition to the attributes of speech perception tests that ensure sensitivity as described above, Bilger (1984, pp. 2-7) indicated that a sensitive speech perception test also needs to have the following aspects clearly established: its purpose, the designation of the population of individuals for whom the test is designed, its validity (predictive, content, construct and face), reliability, including measurement of the standard error and typical participant variance for the population of interest, reliable and equivalent alternative test forms, and lastly a clear procedure for test administration, scoring and interpretation. Of these, reliability and validity measures will subsequently be discussed.

Reliability

The reliability of a speech perception test refers to "the degree to which repeated application of the speech test under identical conditions results in identical scores" (Dillon & Ching, 1995, p. 319). This implies that similar conditions need to be specified and maintained during testing and retesting, on the basis of which the consistency of measurements can then be established. In practical terms, however, identical conditions cannot easily be created, since even if the test material is kept identical, i.e. a particular speech list is repeated, the participant will be more familiar with the test items the second time around (Dillon & Ching, 1995, p. 319). Random errors of measurement are therefore never completely eliminated and the degree of error in the test consequently needs to be known. A measurement can be labeled as reliable, when the amount of random error, i.e. the standard deviation, is slight (Lucks Mendel & Danhauer, 1997, p. 11).

Because of potentially reduced sensitivity in the SAV-L due to the sentences per list not being of equal length, this would have had to be compensated for by keeping the reliability of the test high. High reliability means that random fluctuations in test results are very small, and the sensitivity can consequently be lower, since most differences in scores will be significant (Dillon, 1983, p. 336).

Different kinds of reliability exist and need to be considered. Firstly *inter-judge (inter-rater) reliability* measures how consistently two or more judges/raters, researchers or clinicians independently measure the same event, and is usually given as a percentage of agreement (Bamford & Wilson, 1979, p. 162). The establishment of inter-rater reliability is important in speech perception test development, since the test will in practice be administered by different clinicians.

Secondly, *test-retest reliability* measures the consistency of testing over time. This is also a very important measure, since the developed speech perception test will be used to monitor a participant's progress over time during rehabilitation. During the development of the SAV-L, the test-retest reliability was established by testing the same participants with the same test material with at least a six month break between tests to minimise the potential effects of memorising the test materials on test scores.

Test-retest reliability is a particularly useful measure, giving information on the variability of scores, which is attributable to 'random error' in the test or in the listeners, and which is not attributable to any known or measured factors, such as fluctuations in listener motivation and attention (Bamford & Wilson, 1979, p. 148).

The test and retest scores cannot be expected to be identical due to variability attributed to unknown measured variables (Bamford & Wilson, 1979, p. 162), however, ideally they need to be very low.

Adequate reliability needs to be maintained, but testing still needs to be done in a clinically feasible amount of time to achieve an efficient test (Boothroyd, 1995, p. 354). Inter-test and inter-judge reliability were therefore established during the development of the SAV-L, because these aspects will rarely be reconfirmed in a clinical situation.

Validity

A high degree of reliability or sensitivity does not mean that a test necessarily has high validity (Lucks Mendel & Danhauer, 1997, p. 11). Validity of a test is high only when the test measures what it is supposed to measure (Nunnally & Bernstein, 1994, p. 83; Walden, 1984, p. 16). Some variables that increase sensitivity can actually compromise validity. This is especially true when developing or using a speech perception test to estimate a person's ability to hear in everyday life. The HINT (Nilsson et al., 1994) for example is a sensitive test for establishing a person's SNR loss, but the fact that its sentences are all of equal length and its choice of masking is speech noise (as opposed to multi-talker babble), makes the test less representative of everyday life.

At least four types of validity need to be considered as specified by Walden (1984, p. 16):

Face validity is the extent to which a test instrument appears to measure what it is supposed to measure (Nunnally & Bernstein, 1994, p. 110; Walden, 1984, p. 17). In a speech perception test this means how well the listener understands speech by sound alone in everyday life (Skinner, Holden, Holden, Demorest and Fourakis, 1997, p. 3781). The judgment of face validity is determined in the initial stages of test construction to offer a global decision of whether the plan of content validity appears to be appropriate (Walden, 1984, p. 17). This was the driving force behind the choice of the stimulus and masking noise in the SAV-L's development.

Content validity refers to the relationship between the behaviour studied and the content of the test. With regards to speech perception test development, this can be achieved by using common vocabulary and language structures. During speech perception test development it is necessary to get speakers of the language in which the test is developed, to assess whether the language structures of the test reflect everyday language, as was done during the development of the HINT (Nilsson et al., 1994) and also with the SAV-L.

Criterion validity is the ability of a test instrument to estimate a particular type or level of behaviour. This can be in the form of *predictive validity*, i.e., how well the test correlates with some outside validating criterion. The predictive validity of commonly used speech perception tests is relatively unknown (Walden, 1984, p. 16). Therefore *concurrent validity* is probably a more suited means of establishing criterion validity. Concurrent validity is established by comparing the speech perception test being developed with other measures of similar behavior, for instance by comparing scores of newly developed tests with those of existing tests. This aspect was not addressed in the current study and remains to be completed for the developed SAV-L.

Construct validity of a speech perception test refers to the actual ability of the test to measure the abstract construct of speech perception. The construct of the test material should therefore reflect the speech it is being compared to. An example of poor construct validity would for instance be, if nonsense syllables are used to draw conclusions about speech perception in daily living (Lucks Mendel & Danhauer, 1997, p 8). The choice of stimuli for a speech perception test for rehabilitation purposes therefore needs to be limited to using discourse or sentences, if the construct validity is to be kept high. To balance controlling variables whilst maintaining high validity, the choice of sentences as stimuli rather than discourse is widely accepted. The SAV-L uses sentences of differing lengths to approximate everyday speech as closely as possible.

A fifth type of validity, *ecological validity*, can be added to Walden's (1984) proposed four types. Ecological validity depicts how closely the test resembles natural listening environments (Mackersie, 2002, p. 395). This is a particularly important consideration in the development of speech perception tests to determine (residual) disability following auditory device provision. To maintain high ecological validity, together with construct and face validity, the choice of stimulus and masking material in the development of the SAV-L was carefully considered. In a clinical environment, such as a sound-treated room, everyday speech can, however, only be simulated to a certain degree, because independent variables, such as type and level of background noise as well as speaker intensity and visual cues still need to be controlled in order to maintain high reliability.

2.2.1.1.2. Stimulus type

Speech perception is mostly assessed using nonsense syllables, spondee words, monosyllabic words, sentences or discourse.

Nonsense syllables are the most difficult to recognise (McArdle & Hnath-Chisolm, 2009, p. 69). They have minimal semantic content, thus effectively eliminating differences between participants in education and vocabulary (Lutman, 1997, p. 68; Lucks Mendel & Danhauer, 1997, p. 70). They can be constructed to measure a specific feature, which can make them very powerful in determining an individual's specific auditory abilities (Gatehouse & Robinson, 1997, p. 81). However, nonsense syllables give very little information about the auditory disability and handicap in everyday life that an individual experiences (Gatehouse & Robinson, 1997, p. 82; Lucks Mendel & Danhauer, 1997, p. 72). Additionally, participants require considerable training to achieve stable performance when nonsense syllables are used as test stimuli, which does not make them really suitable in a clinical context (Lutman, 1997, p. 68).

Monosyllabic words are used to distinguish between different types of hearing losses (Lucks Mendel & Danhauer, 1997, p. 61) by obtaining a performance intensity function and assessing the shape of the function, which correlates to the site of peripheral auditory lesion (Jerger & Hayes, 1977, p. 216). Like nonsense syllables, using monosyllabic words as stimuli can provide very useful information about a person's speech perception skills, especially if a detailed analysis is made of a participant's error responses (Stevenson, 1975, and Robertson and Plant, 1983, as cited in Plant, 1997, p. 208). Monosyllabic words as stimuli, however, cannot adequately assess an individual participant's everyday communication skills (Plant, 1997, p. 208; Berger, 1971, p. 212; Speaks, Jerger & Trammell, 1970, p. 766). They also lack face validity for predicting social adequacy (Lucks Mendel & Danhauer, 1997, p. 60). They furthermore cannot be used to differentiate amongst hearing aids (Shore, Bilger & Hirsh, 1960, p. 167), partly because the stimuli are too short to allow the compression in hearing aids to activate (Cox, 1993, ¶11).

Spondee words were specifically designed to test speech reception thresholds (SRTs) and/or speech detection thresholds (SDTs) (Lucks Mendel & Danhauer, 1997, p. 64) and have a very specific place in the audiological test battery, as they are used routinely to confirm pure tone thresholds. Spondees are, however, less representative of 'natural' language communication than sentences, since spondees spoken as isolated utterances or in carrier phrases may not represent the normal spectral weighting, level fluctuations, intonations, pauses, etc., associated with conversational speech. The limited number of spondees in any particular language further heightens learning effects associated with randomisation. This prevents measurement and comparison over time in experimental or clinical conditions. Spondees as stimuli may also be too short to sufficiently engage dynamic processing characteristics, such as compressors found in many contemporary auditory devices

(Nilsson et al., 1994, p. 1086) and are therefore, like nonsense syllables and monosyllabic words, also not stimuli with high face validity.

Sentences are relevant to everyday communication and therefore have greater face validity than syllables or words (Yotis, 2005, p. 13). They provide additional information on the ability of participants to understand speech in daily life and have proved to be a useful tool, especially for the selection of a suitable hearing aid (Brinkmann & Richter, 1997, p. 108). Recorded sentence tests are routinely used to do aided testing during cochlear implant evaluations (Niparko et al., 2009, p. 138). Some sentence tests such as the HINT (Nilsson et al., 1994) contain sentences of similar lengths and sentence structure to ensure fairly equal difficulty, a necessary feature in an adaptive SNR test to ensure high sensitivity (Dillon, 1983, p. 343). Other tests such as the CID sentences (Silverman & Hirsh, 1955) use sentences of varying lengths, which makes them more representative of every use of language.

Many sources of information, such as acoustic and contextual clues, are used to understand sentences and they have many extrinsic redundancies (McArdle & Hnath-Chisolm, 2009, p. 71). These characteristics make it harder to predict which specific information is being used by the listener, as so many sources of information are involved (Hazan, 1997, p. 200). Sentence tests therefore measure a hearing deficiency only partly at the peripheral level, and can be considered as a measure of a combination of linguistic competence and central processes (Lehmann, 1962, as cited in Lyregaard, 1997, p. 40).

Sentences are more problematic to standardise for equal difficulty (Gatehouse & Robinson, 1997, p. 82). However, they do have high face validity and an individual's performance on a sentence test can therefore be more easily related to speech hearing performance than would be the case with test results based on syllables or phonemes (Lyregaard, 1997, p. 41). Sentences tend to have a steep psychometric function (i.e. the curve relating a participant's percentage score to intensity or SNR), which makes them more suitable for a speech-in-noise task, where the SNR can be pre-set to a level that achieves the desired level of performance and avoids 'ceiling' and 'floor' effects (Lutman, 1997, p. 68), also known as truncation effects (De Vaus, 2005, p. 110).

Discourse contains all the semantic, contextual and prosodic features of everyday speech and therefore has great face validity as a material for disability measurement. However, it takes some training for the participant to learn to repeat running speech and it is difficult for the tester to score responses accurately and consistently (Lutman, 1997, p. 68). Circumstances in which the speech material attempts to mimic natural conversation, are also difficult to control and standardise, which in turn makes discourse hard to analyse (Gatehouse & Robinson, 1997, p 81). Although speech tracking is an effective aural rehabilitation strategy, it is inappropriate as a test procedure since there are a number of extraneous variables, such as the speaker, receiver and test, that are very difficult to control (Tye-Murray and Tyler, 1988, p. 226; Lucks Mendel & Danhauer, 1997, p. 76).

Considering the different types of speech stimuli, it was decided to use sentences as stimuli in the development of the SAV-L, in order to maintain high face validity, whilst still being able to control variables e.g. task learning, memory and presentation level, therefore maintaining high reliability.

2.2.1.1.3. Phonemic balancing

Phonemic balancing is another consideration in the development of a speech perception test (Dillon & Ching, 1995, p. 308). A distinction between phonemic and phonetic balancing of test material containing speech can be made. Phonetically balanced test material is material that contains a reasonably proportional representation of the sounds that occur in everyday speech (Dillon & Ching, 1995, p. 308; Lehiste & Peterson, 1959, p. 281). However, equivalent lists like these are very difficult to compile.

An alternative way to obtain phonetically balanced scores is to use lists, which contain the same proportion of phonemes in each list, i.e. phonemically balancing the lists (Lehiste & Peterson, 1959, p. 282). Phonemic balancing is widely used in word lists such as the Arthur Boothroyd (AB) word lists (Boothroyd, 1968). It is seen as more important in word lists than in sentence lists, since the items in word lists are limited mostly at the phonemic level without semantic or syntactic cues (Bamford & Wilson, 1979, p. 149). However, the relevance of precise fulfillment of phonemically balanced speech perception test material for predicting communicative difficulties in everyday life due to hearing loss, is questionable (Dillon & Ching, 1995, p. 308; Tobias, 1964, p. 99). This is partly due to the effects of co-articulation

(Yotis, 2005, p. 11) but also to the sequential and semantic constraints in normal words or sentences (Bench & Bamford, 1979, p. 13). The current study therefore did not focus on ensuring phonemic balancing, since it is also suggested that phonemic balancing is irrelevant for diagnostic purposes (Lyregaard, 1997, p. 47) or at least there is little agreement over its importance (Olsen & Matkin, 1991, p. 83). Also, it can safely be presumed that by having chosen sentences from everyday conversation for the current test, the average speech spectrum would have been unaffected. This is normally found to be the case unless the material is highly constrained, such as with the use of nonsense syllables with only a few vowels, or words with an abnormally high proportion of /s/, /a/, /sh/ or /zh/ (Dillon & Ching, 1995, p. 316).

2.2.1.1.4. Word familiarity

Word familiarity is an important consideration in compiling sentence lists, since words that are encountered more frequently in real life, tend to be recognised better in speech perception tests than words that are not familiar (Dillon & Ching, 1995, p. 309; Owens, 1961, p. 129; Gernsbacher, 1984, p. 275). This was demonstrated when children performed better on speech perception tests following semantic and syntactic training (Sarant, Blamey, Cowan & Clark, 1997, p. S136). Because of this influence of vocabulary on the test scores of words-in-sentences as well as single word test scores, the SAV-L was designed for adults only, as its grammar, topics and vocabulary used might be too difficult for children and consequently affect their performance (Clark, 2003, p. 714).

2.2.1.1.5. Number of test items

The more items are scored, the more reliable a test is (Dillon & Ching, 1995, p. 314; Dillon, 1983, p. 336). Increasing the number of items is actually the only way to improve reliability of test scores without sacrificing sensitivity (Green, 1997, p. 167). Gelfand (1998, p. 1088) suggested that to minimise variance, a speech recognition test should contain 450 items. However, this amount of items would make the test too long to be clinically feasible.

In sentence testing, the scoring units generally considered are whole sentences or words, but the use of syllables as scoring units is also a consideration.
Although whole sentence scoring is easy and quick, this scoring method limits the amount of test items. It is also likely that, especially in noise, some individuals with hearing impairment never hear the complete sentences, even when the signal level is increased (Hällgren et al., 2006, p. 233), which increases floor effects.

Using words or syllables as scoring units, allows an increase in test items in sentence testing, which increases reliability and is therefore a more appropriate consideration. To maintain a balance between test reliability and a clinically acceptable length of time needed to administer the test, the number of items in the development of the SAV-L was kept as high as possible by choosing to score the items at syllable level rather than at word or sentence level.

Scoring at morpheme or phoneme level would have increased the test items even more. However, making scoring potentially too complicated would have compromised inter-test and inter-rater reliability. Scoring at phoneme level can be affected by faulty speech production encountered in hearing-impaired patients (Kowal, 1979, p. 115), putting a hearing impaired person at a great disadvantage and increasing the chance of baseline effects.

By using scoring at syllable level, it was felt that a trade-off between precision and test duration was reached, preventing the test from being too long to fatigue, yet keeping the number of test items high to increase reliability (Bamford & Wilson, 1979, p. 148). The lists were compiled to contain 100 words each, which left the syllable score to be between 134 and 142.

2.2.1.1.6. Test technique

When administering speech-in-noise tests, two test methods need to be considered, namely adaptive SNR testing and fixed SNR testing (Taylor, 2003, p. 40).

Adaptive SNR testing is an up-down procedure, where either the intensity of the speech stimuli or the noise intensity level is adjusted until a preset (usually 50%) score is obtained (Nilsson et al., 1994, p. 1085; Middelweerd and Plomp, 1987, p. 2145). During this method of testing, truncation effects can easily be avoided (Lutman, 1997, p. 68). This method is often used for diagnostic purposes, e.g. diagnosing SNR loss (Taylor, 2003, p. 42).

During fixed SNR testing, a percent correct score is obtained at a predetermined speech and noise level (Taylor, 2003, p. 40). Although fixed SNR testing can be limited by truncation effects (Nilsson et al., 1994, p. 1085) and does not necessarily concentrate on the most sensitive region of the PI function, the advantage of this method is that it gives a percent-correct score, which is easily understood by patients (Taylor, 2003, p. 40). This form of testing can be used to compare aided versus unaided scores (Taylor, 2003, p. 41). Fixed SNR testing is also used in cochlear implant clinics as part of the test battery to determine a person's cochlear implant candidacy, with a percentage correct score being used determine cochlear implant candidacy (Zwolan, 2009, p. 923). The SAV-L was specifically designed for this purpose and it was therefore decided to use the fixed SNR technique.

2.2.1.1.7. Sentence list equivalence

The equivalence of sentences lists in a speech perception test is an important consideration in ensuring that the test can be repeated over time with results being comparable (i.e. increasing inter-list reliability). Sentence lists can never be fully equivalent, but the variability in scores when the test is presented to the same participant under identical conditions, has to be kept as low as possible. Apart from excluding difficult vocabulary, list equivalence can be further improved by slightly adjusting the intensity of the sentences that deviate from the average score, in order to make them slightly easier or more difficult (Wong et al., 2007, p. 71S; Hällgren et al., 2006, p. 229; Wong & Soli, 2005, p. 279; Kollmeier & Wesselkamp, 1997, p. 2414; Nilsson et al., 1994, p. 1088). This is typically achieved by presenting the sentence material to participants with normal hearing at a pre-determined SNR level. A mean percent intelligibility score is determined for each sentence, based on the number of words repeated correctly (Vaillancourt et al., 2005, p. 361), and adjustments in intensity are then made according to the scores of the individual sentences. The resulting differences in intensities do not pose a problem, since no particular advantages arise from having test items match exactly in signal level and signal-to-noise ratio, if this causes the items to be mismatched in intelligibility (Dillon, 1983, p. 344).

The use of normal hearing participants for this process allows the researcher to find a group of individuals that is as homogeneous as possible, since it is difficult to attain a group of hearing impaired participants that is homogeneous (Bamford & Wilson, 1979, p. 162).

In the development of the HINT (Nilsson et al., 1994), a mean percentage score of all the sentences was determined and sentences that did not fall within a determined range of this mean were eliminated. Elimination of sentences is easily done when sentences of equivalent length and a simple construction is used, because these two aspects allow for the compilation of a large number of sentences to form lists that are fairly equivalent. However, the sentences of the SAV-L have varying sentence lengths, specific topics and sentence types (statements versus questions versus commands). These restrictive characteristics did not allow a large collection of sentences to be developed, from which the unsuitable ones could be eliminated. The focus therefore had to be on the adjustment of sentence intensity to approximate equivalence rather than the elimination of unsuitable sentences.

2.2.1.1.8. Establishing normative values

When sentence tests are used in rehabilitation, two questions arise with regards to the participant's performance. Firstly, how does the participant compare to his/her peers, and secondly, does the participant improve over time in comparison to him-/herself (Lyregaard, 1997, p. 42)? To answer the first question, 'normal' values need to be established. Because one of the main purposes of the SAV-L was specifically to compare the performance of persons with hearing impairment with that of persons with normal hearing in order to establish cochlear implant candidacy, the sample used to establish tentative norms, consisted of persons with normal hearing. Because this is such a large population, the use of sampling was inevitable.

2.2.1.1.9. Masking type

While some types of masking noise, e.g. white noise, have a highly deterministic structure and can be analytically controlled, others such as speech as background noise are more representative of everyday listening environments. Ultimately the choice of masking noise has to be based on the objectives of the test (Gatehouse & Robinson, 1997, p. 84).

The masking noise used in speech perception testing should have significant energy at all frequencies present in the speech signal. One way to achieve this is by generating a noise that has a frequency spectrum that approximates the long-term spectrum of speech, i.e. using speech-shaped noise. Alternatively actual speakers, i.e. multi-talker babble (Lutman, 1997, p. 69), can be recorded as background noise.

According to Green (1997, p. 152) “if measures of speech recognition are to be useful in rehabilitation, then they must tell us something about the patient’s ability to cope in the world outside the audiology clinic”. Listeners frequently face the ‘cocktail party problem’ (Cherry, 1953, p. 976), i.e. the segregation of a single target voice from a competing milieu. Due to this commonly encountered problem, multi-talker babble was decided upon, since it appears to be the noise signal with the best face validity (Plant, 1997, p. 228), being the most common background noise experienced by listeners (Wilson et al., 2005, p. 500). Because multi-talker babble closely represents everyday listening situations, it is generally the preferred masking type for testing bimodal and bilateral speech processing (Clark, 2003, p. 718).

The advantage of multi-talker babble as opposed to speech-shaped steady noise is that, although both types of masking noises match the frequency spectrum of speech, multi-talker babble additionally resembles the temporal aspects of speech (Hawley, et al., 2004, p. 833) and is therefore more effective than steady noise to mask speech (Lutman, 1997, p. 68). This explains why a higher correlation between self-reported disability has been found when modulated noise was used during sentence-in-noise testing in comparison to when a steady noise was used (Tyler and Smith, 1983, as cited in Lutman, 1997, p. 69). This supports the observation of Bronkhorst and Plomp (1992, p. 3137) that the true handicap of the hearing impaired may be underestimated when only steady-state maskers are used in speech intelligibility testing.

The next task was to decide how many talkers to use during the recording of the multi-talker babble. Different numbers of speakers lead to different mechanisms of masking. Informational masking, also called higher-level, central cognitive, linguistic (Van Engen & Bradlow, 2007, p. 519) or nonenergetic masking (McArdle & Hnath-Chisolm, 2009, p. 75), is the uncertainty created for the listener as to whether he/she is hearing the target or the masker when few enough talkers in the babble noise are used to maintain linguistic content (McArdle & Hnath-Chisolm, 2009, p. 75; Hawley et al., 2004, p. 834). During informational masking, the target and masker signals are therefore clearly audible, but the listener is unable to segregate the elements of the target signal from the elements of the similar-sounding distracters (Brungart, et al., 2001, p. 2528). This is thought to possibly play an important role in the cocktail party problem (Brungart et al., 2001, p. 2531) and is maximised for sentence material for small values of N (e.g. Simpson & Cooke, 2005, p. 2777; Freyman, Balakrishnan & Helfer, 2004, p. 2246), where N is the amount of talkers in the babble. The linguistic uncertainty component of informational masking effects is most potent for N = 2 (Freyman et al., 2004, p. 2254) or N = 3 (Carhart, Johnson & Goodman, 1975, p. S35) and is almost absent by N = 10 (Simpson & Cooke, 2005, p. 2777).

Energetic masking, also called lower-level or peripheral masking (Van Engen & Bradlow, 2007, p. 519), on the other hand, is the uncertainty or complete loss of information due to the masker being sufficiently intense (Brungart et al., 2001, p. 2528; Garcia Lecumberri & Cooke, 2006, p. 2446). It occurs when a target speech signal and a concurrent noise overlap in time and/or frequency (Hoen, 2010, slide 6).

Energetic masking increases as speaker number increases beyond two (Simpson & Cooke, 2005, p. 2775; Van Engen & Bradlow, 2007, p. 524). However, when the number of maskers is four or more, the masking effect approaches that of steady-state noise (Bronkhorst & Plomp, 1992, p. 3136), i.e. one starts to lose the temporal characteristics. Hoen (2010, slide 15) found that more than six to eight talkers results in purely energetic masking. As the number of speech interferers is increased beyond two or more, the difference between speech interferers and noise interferers therefore decreases (Johnstone & Litovsky, 2006, p. 2177).

To keep the informational masking component, but still have sufficient energetic masking, it appears that the optimal number of speakers is two to four. Most commercially available speech-in-noise tests involve multi-talker babble with several speakers, which decreases the effects of informational masking, but increases energetic masking. However, future studies in this area are warranted (McArdle & Hnath-Chisolm, 2009, p. 75). For the current study it was decided to use three-talker babble, because three talkers represent a realistic simulation of a social gathering that requires “selective listening” i.e. tuning out one or more nearby talkers. If only two speakers had been used, the masking noise would possibly have had increased quiet periods of natural pauses during speech, i.e. decreased the energetic masking, whereas four was thought of as being too many (R.S. Tyler, personal communication, October 6, 2009).

The language in which the babble was recorded was Afrikaans, since the language of interfering babble can affect the intelligibility of the target speech (Hoen, 2010, slide

16; Van Engen & Bradlow, 2007, p. 524; Garcia Lecumberri and Cooke, 2006 p. 2445). Garcia Lecumberri & Cooke (2006, p. 2449), for example, found that English listeners were better able to tune out an unknown language. Using the same language babble as the test stimuli therefore seems to maximise the informational masking.

During multi-talker babble, maskers of different gender degrade performance less than same-gender maskers (Rhebergen, Versfeld & Dreschler, 2005, p. 1274). This is especially true if the SNR is positive. When the SNR is negative, this is also the case, however, odd-sex distraction can also occur, i.e the listener can be severely distracted if one of the masking voices is of different gender than the rest (Brungart et al., 2001, p. 2537). This is attributed to the characteristics of the speakers, such as vocal tract size and fundamental frequency (Van Engen & Bradlow, 2007, p. 520).
To keep the frequency distribution from both sexes, it was decided to use two male and one female speaker. The female speaker might pose a risk of odd-sex distraction at negative SNRs, but the use of male-only voices was felt to be less representable of everyday speaking conditions than mixed-sex voices.

2.2.1.1.10. Open versus closed set response

During closed set responses, a set number of options are given to the test participant to choose from. This is easier to score and can even be applied so that participants record their own responses and make the test self-paced by using appropriate computer software.

During open set responses, in contrast, no set options of answers are provided. A virtually unlimited number of alternatives are therefore available, meaning that individuals need to use linguistic as well as auditory factors to help them perceive the test items. During the use of open set responses, no guessing floor is present, which might otherwise boost scores (Lucks Mendel & Danhauer, 1997, p. 16). Open-set testing more closely predicts the actual ability of a person to communicate in everyday situations than closed set testing (Clark, 2003, p. 713) and was therefore the choice for the current test.

2.2.1.1.11. Use of a carrier phrase

Carrier phrases are often found in recorded word lists, e.g. the AB word lists (Boothroyd, 1968) and 'Foneties Gebalanseerde' [Phonetically Balanced] (FG) woorde [words] (Laubscher & Tesner, 1969). In sentence tests a carrier phrase like 'ready' is a consideration. In auditory-only recorded test material a carrier phrase could be drawn in parallel to visual clues of initiation of conversation. However, the issue of whether to use a carrier phrase or not is presently unresolved (Lucks Mendel & Danhauer, 1997, p. 24). In the absence of agreed best practice, in the current test, it was decided not to use a carrier phrase. The lack of a carrier phrase also helps to keep the test as compact as possible with regards to time, which would make it more practical for a clinical test battery where time is always restricted.

2.2.1.2. Speaker

Speaker characteristics can affect performance significantly and are discussed in the next section.

2.2.1.2.1. Live voice versus recordings

Live voice tests offer greater flexibility over the test environment than recorded voice (Creston, Gillespie & Krohn, 1966, p. 17). They also require shorter time for test administration, because the tester can control the rate of stimulus presentations, as well as inter-stimulus intervals (Resnick, 1962, as cited in Lucks Mendel & Danhauer, 1997, p. 24). They furthermore do not require any special equipment, simulate a more real-life situation (Plant, 1997, p. 225) and are perhaps more suitable for young children or children who are difficult to test (Bench & Bamford, 1979, p. 14). However, with live voice, scores obtained during speech perception tests by the same listeners, lists and speaker but recorded on different occasions, will yield differences of up to 10% (Brandy, 1966, as cited in Lyregaard, 1997, p. 49), which will decrease test reliability. This happens quite naturally, if the clinician consciously or unconsciously varies his/her clarity of presentation to help achieve a desired result (Dillon & Ching, 1995, p. 318).

In addition to varying clarity, intensity levels are also less well controlled during presentation of live voice, especially when testing in background noise, since speakers tend to naturally raise their voices in background noise. During speech audiometry, speech stimuli can be presented through an audiometer’s earphones or loudspeakers, or in the soundfield using live voice. If live voice in the soundfield is used, a sound pressure level meter should generally be used to monitor stimulus levels, but when recorded tests are used, the test is always presented through an audiometer. The volume units (VU) meter on an audiometer controls the signal level and serves as a better control for the stimulus level than when live voice is monitored by a sound level meter in the soundfield.

The control of these variables by using recorded test materials improves not only intra-subject precision, but also inter-subject precision (McArdle and Hnath-Chisolm, 2009, p. 66). Clark (2003, p. 709) felt strongly that pre-recorded materials should be used, pointing out that it was essential for the comparison of results either within or across centres. Pre-recording of speech perception test material therefore seems more favourable than live voice tests (Plant, 1997, p. 225). American Speech-Language-Hearing Association (ASHA) guidelines (1988, ¶21) also recommend the use of recorded stimuli.

Due to uncontrolled variables during live voice testing, the SAV-L was recorded rather than presented live, in order to better maintain consistency in presentation (Kreul, Bell & Nixon, 1969, p. 286).

2.2.1.2.2. Gender of speaker

Because male and female voices are sufficiently different to cause differences in intelligibility scores for the same material and listener, the same speaker should be used throughout a test (Lyregaard, 1997, p. 48). The basis for the choice of gender for the developed test was less clear, since both genders are encountered in everyday life. Since most existing sentence tests for adults recorded previously in South Africa utilised male voices, e.g. the re-recording of the CID sentence lists in Afrikaans (Müller & De Stadler, 1989), it was decided to also use a male speaker.

This would allow for closer comparison between test results during concurrent validity measures.

2.2.1.2.3. Dialectical appropriateness

In order for a test to be appropriate for assessment of the wider population, the vocabulary and colloquial expressions need to be considered, so that some participants are not penalised for being unfamiliar with the test materials (Plant, 1997, p. 226). By making sure that difficult items are excluded, more equivalent sentence lists are also ensured (Vaillancourt et al., 2005, p. 361; Versveld et al., 2000, p. 1673). In the present study this was addressed by getting a group of speakers with Afrikaans as first language to read through the sentence lists and rate whether the sentences were sentences they would encounter in normal everyday conversations.

To further ensure the dialectal appropriateness, a speaker, who is a professional radio commentator fulfilling the definition of someone with a standard Afrikaans accent, was chosen to record the test material. Excluding the use of a strong regional accent allowed the test material to be suitable for national use. Lyregaard et al. (1976, in Kowal, 1979, p. 140) pointed out that “the dialect of the listener may not be particularly important, if the material is recorded in the manner typical of broadcasting speakers, as listeners will generally be familiar with this kind of speech”. The choice of a national broadcaster as speaker was therefore seen as ideal.

2.2.1.3. Electro-acoustic transmission channel

The electro-acoustic transmission channel refers to the type of transducer used and involves either headphones or a loudspeaker(s). The SAV-L will be administered in a soundfield test condition since users of auditory devices cannot be assessed with headphones due to microphone placement and feedback problems. Despite this, throughout the development of the test, it was decided to present the stimuli diotically through TDH-50 headphones in a sound-treated room, because soundfield testing is not generally as well controlled and monitored as testing using headphones (Sherwood & Fuller, 1997, p. 100). In the soundfield even slight movements of the head may affect the sound and SNR, if sound is presented from different sources (Cameron, Dillon & Newall, 2006, p. 31). Furthermore, in soundfield testing, listener placement is difficult to replicate exactly between testing, which is a necessity for test consistency (Koehnke & Besing, 1996, ¶4). The use of headphones was seen as acceptable since sound pressure levels generated by headphones can reliably be referred to sound pressure levels in the soundfield (Brinkmann & Richter, 1997, p. 114).

2.2.2. Patient components

Patient components refer to variables directly connected to the participants and are divided into the auditory transmission channel and reponse channel used.

2.2.2.1. Auditory transmission channel

The auditory transmission channel in this context is interpreted as the factors that will affect the reception of the stimulus by the participant, either due to participant variables or the level of the stimulus and will be discussed as such.

2.2.2.1.1. Participant variables

Different sites of lesion in auditory dysfunction can create different clinical pictures, leading to variable results in speech perception tests, even when the pure tone audiograms of two participants are very similar (Taylor, 2003, p. 43). Someone with a sensorineural hearing loss would have a loss of intensity, temporal and frequency cues, leading to reduced clarity of speech perception, whereas a person with a conductive hearing loss would only have a reduction in intensity of sound. Someone with auditory processing disorders (APD) would also most likely experience far more problems understanding speech in the presence of background noise than someone with normal auditory processing skills.

For normative scores to be obtained, participants with normal hearing were therefore selected. Participant criteria had to be carefully considered with regards to peripheral and central hearing status, but also with regards to age. Upper and lower limits of age had to be set in order to exclude early signs of presbyacusis, poor understanding and concentration, since these factors could affect the understanding of test instructions and the ability to stay focussed for the duration of the test.

2.2.2.1.2. Presentation and SNR levels

According to Skinner et al. (1997, p. 3781), one of the concepts in developing a test for the evaluation of postlingually deafened adults, such as cochlear implant candidates, is that the intensity level at which speech perception tests are presented clinically, should represent that of conversational speech. In quiet, this level would be around 60dBSPL (Firszt et al., 2004, p. 376). In noisy situations, noise levels of around 65dBSPL can be considered a worst case scenario (Cochlear, 2008, p. 8), with SNRs of +5 to +15dB being generally maintained at conversation distances (Olsen, 1998, p. 25). For the current test, it was decided to set the babble constant at 45dBHL (which is equivalent to 65dBSPL), with a SNR of +5dB. Although studies in the use of cochlear implants do not have a universal level at which masking is presented, the choice of intensity level in the current study is comparable to levels used in other tests, which vary from 60 to 70dBSPL (e.g. 60dBSPL in studies conducted by Kim et al., 2009, p. 75, and Koch, Soli, Downing & Osberger, 2009, ¶10; 65dBSPL in studies undertaken by Müller-Deile, Kiefer, Wyss, Nicolai & Battmer, 2008, p. 14, and Chung, Zeng & Acker, 2006, p. 2219; 70dBSPL in a study undertaken by Balkany et al., 2007, p. 759). Madell (2010, slide 23) also describes a +5dBSNR as the most realistic SNR to use during speech perception testing.

2.2.2.2. Response channel (mode)

Participants can use written, oral or manual communication to respond to stimuli (Lucks Mendel & Danhauer, 1997, p. 18). Written answers will take longer than verbal responses and rely on the participant’s ability to write and is therefore not generally used. However, writing responses could be considered when someone's pronunciation is not clear. As a standard, it was decided to allow sufficient time between sentences for verbal repetition of the stimulus.

2.2.3. Results

The concept of *results* refers to the means in which a response is measured and compared to the norm and is discussed below:

2.2.3.1. Measurement of the stimulus

Although verbal responses were the choice of response for the developed test, verbal responses can be misheard by the tester, especially if the tester has an expectation of what the participant is going to say. This can fortunately be resolved by lipreading, since most errors made by normal hearers, are place of articulation errors (Dillon & Ching, 1995, p. 313). It is therefore necessary for the tester to use visual and auditory cues when scoring speech perception tests. This requires adequate lighting in the test set-up as well as positioning of the tester in order to see the participant from the front. Fortunately most audiometric test booths are set up as such, including the booths in which testing was undertaken for the development of the SAV-L.

2.2.3.2. Comparison of the response to the correct stimulus

As discussed before, it was decided to score responses on syllable level. The response of the participants had to be exactly correct (e.g. if a singular instead of plural form was used, the participant was penalised in the syllable that reflected the plural form). However, when dialectal differences occurred which carried the exact meaning of the word, the participant was not penalised. An example of this would be a participant saying ‘watse’ instead of ‘watter’ (‘which’), which is purely a dialectal variety of the word.

2.3. RESEARCH QUESTION

On the basis of existing needs in South Africa and the above discussion of various speech perception test characteristics and qualities, it was decided to use the CUNY topic-related sentences (Boothroyd et al., 1985) as a framework for developing an Afrikaans sentence perception test.

Because the CUNY topic-related sentence test contains a large collection of sentence lists, the current study was the first stage of a phased approach to develop the full Afrikaans version of a sentence test in quiet and in noise. The current study focussed on developing 27 sentence lists in three-talker babble, based on careful consideration of the different components of speech perception tests as discussed above. The test would be called the *Sinslyste in Afrikaans vir Volwassenes in Lawaai* [Sentence lists in Afrikaans for Adults in Noise] (SAV-L).

The study served to answer the following research question: Is the developed Afrikaans sentence test a valid and reliable test that is sensitive to the adult test participant’s speech perception abilities, at levels that are representative of normal conversational speech in noise?

CHAPTER 3

METHODOLOGY:

3.1. INTRODUCTION

This chapter presents the aims, design, participants and ethical considerations for the study undertaken. It furthermore gives a detailed description of the test equipment and procedures used.

3.2. AIMS OF THE STUDY

The main aim of the study was to develop a valid, reliable and sensitive Afrikaans sentence perception test in noise for use at a fixed signal-to-noise ratio (SNR). A test like this would be suitable to assess the test participant's speech perception abilities at levels that represent speech in everyday noisy situations, in order to compare it with the score a person with normal hearing would achieve (Taylor, 2003, p. 41). To achieve this aim, the following sub-aims were established:

1. The development and recording of 27 Afrikaans sentence lists in the presence of three-talker babble, based on the CUNY topic-related sentence test (Boothroyd et al., 1985).
2. Establishment of inter-list, test-retest as well as inter-rater reliability of the developed material.
3. Confirmation of performance at the intended test level, i.e. at levels of normal conversational speech in noise, where a 100% score was expected for listeners without any speech perception difficulties.

3.3. RESEARCH DESIGN

The original research paper describing the development of the CUNY topic-related sentences (Boothroyd et al., 1985) is not available anymore (personal correspondence with A. Boothroyd, 2009). However, the test is still very much in use and has been expanded into the Computer-Assisted Speech Perception and Training at the Sentence Level (CASPERSENT, Boothroyd, 2006). The purpose of the test material remains "a means of assessing the probability of word recognition under conditions that simulate conversation" (Boothroyd, 2006, p. 2).

The following characteristics of the CUNY topic-related sentence test were kept during the development of the SAV-L:

1. Varying sentence lengths. In the CUNY topic-related sentence test the sentence lengths vary from 3 to 14 words but during the compilation of the SAV-L, the 14-word sentences were reduced to 12-word sentences, therefore the sentence lengths varied from 3 to 13 words with two 12-word sentences per list. This was because in Afrikaans compound words are generally written as one (Carstens, 2003, p. 232), which can make Afrikaans sentences with the same amount of words as that of English, longer and more taxing on auditory memory.
2. Large number of sentence lists (27 for the purpose of the current study and a further 24 recorded at the same time for standardisation based on the methodology of the current study).
3. Inclusion of 4 statements, 4 commands and 4 questions per list.

4. Inclusion of the following variety of topics per list: food, family, work, clothes, animals and pets, homes, sports and hobbies, weather, health, seasons and holidays, money and music. Appropriate adjustments were made during translation to incorporate the South African culture rather than the American culture, e.g. excluding sentences containing American holidays like Halloween.

A critical evaluation of current and available literature was used to compile an appropriate research methodology. By far the most widely published methodology for the development of recent sentence material, is that of the HINT (Nilsson et al., 1994), which has been translated into several languages, including Canadian French (Vaillancourt, et al., 2005), Cantonese (Wong & Soli, 2005), Mandarin (Wong et al., 2007) and Swedish (Hällgren, et al., 2006). The HINT, together with other available sentence tests like the Bench-Kowal-Bamford Speech-In-Noise (BKB-SIN) test (Niquette et al., 2003) and Quick Speech-In-Noise (Quick-SIN) test (Killion et al., 2004), differ significantly from the CUNY topic-related sentences in that they contain only sentences that are equal in length, and they have no restrictions with regards to specific topics used. As discussed in section 2.2.1.1.7 these two characteristics allow for adjustment of individual sentences first to approximate equal difficulty, after which the individual sentences can be combined to form fairly equivalent lists. In view of the CUNY topic-related sentences having restrictions in terms of differing sentence lengths and specific topics used, individual sentences could not be manipulated first before combining them into lists. Instead, the lists had to be developed as a unity. An adapted methodology therefore had to be devised to develop sentence lists approaching equal difficulty whilst maintaining the restrictive characteristics that the CUNY topic-related sentences have.

The current study used a developmental research design (Thomas, 1987), consisting of three phases, namely the analytical, design- and developmental phases. During the *analytical phase* the problem, i.e. the need for a sentence test with high ecological validity, was identified. The *design phase* consisted of the translation and recording of the sentence lists. Lastly, during the *developmental phase*, the sentences were adjusted in intensity to form lists that scored similarly by participants with hearing within normal limits.

3.4. PARTICIPANTS

Three groups of participants had to be identified. For the design phase, a group of participants were firstly needed to rate the naturalness of the compiled sentence lists. Secondly, a group of speakers were required for the recording of the sentences. A third group of participants was needed for the developmental phase, to evaluate the test material. All the groups entailed a large population and samples therefore had to be used. Ideally a completely random sample is always favoured, but since the parameters that had to be measured were known, a stratified sampling scheme could be adopted (Lyregaard, 1997, p. 42).

3.4.1. Raters of naturalness

The raters of naturalness had to speak Afrikaans as first language with at least a Senior Certificate (i.e. Grade 12/Standard 10) or equivalent. Five participants were selected through snowball sampling. After amendments to the compiled sentence lists were made as suggested by the raters, the sentences were submitted to three more raters selected in the same way. Table 3.1 contains a description of the participants. The participants were given a letter with information about the study (Appendix 1) as well as a letter of consent that they had to sign (Appendix 2). No reward was offered for participation.

3.4.2. Participants in the recording of the test material

The sentences were recorded using a professional male voice actor with Afrikaans as first language. He spoke standard Afrikaans as opposed to having a strong regional accent. For the recording of masking noise, four talkers, two male and two female, were used. Of these, the recordings of two of the males and one of the females were used in the final compilation to make three-talker babble. All had a degree in drama and spoke Afrikaans as first language, with a standard accent.

Table 3.1. Description of participants who rated the naturalness of the sentences

Participant no	Sex	Age	Profession	Highest qualification
1	F	64	Building Consultant	College degree in administration
2	M	67	Retired Health and Safety Inspector	Senior certificate
3	F	42	Estate Agent	Estate agent's qualification
4	F	28	Speech therapist & Audiologist	University degree in Speech Therapy and Audiology
5	F	38	Speech therapist & Audiologist	University degree in Speech Therapy and Audiology
6	F	61	Proof-reader	Bachelors of Arts
7	M	64	Audiologist	Master's degree in Audiology
8	F	64	Audiologist	Master's degree in Audiology

3.4.3. Participants in the evaluation of the recorded test

During the developmental stage, 64 participants (10 persons per round for 5 rounds, 9 persons for another round and 5 participants for a last round) were used to evaluate the test material. Appendix 3 stipulates the distribution of age and gender of the participants. For each round, new participants were recruited in order to avoid the learning effect. Participants were selected through snowball sampling, obtained through word of mouth and through advertising at the involved university's medical campus, a local school and the audiology private practice at which testing was conducted. Prior to advertising at the local school, permission was obtained from the headmistress of the school.

All participants were sent an information letter explaining the purpose of the test and what would be expected from them during testing (Appendix 4). They were asked to sign an informed consent letter, which described the purpose of the study, indicated that their participation was voluntary, that they could withdraw at any point, and that they would remain anonymous. Adults received different consent forms than scholars, who needed additional consent from their parents (Appendices 5a-b). The participants were given R100 to cover travel and other expenses with money from a grant awarded to the Cochlear Implant Team at Tygerberg Hospital for the development of speech perception test materials by the Cochlear Foundation Ltd.

The selection criteria for the participants used for the evaluation of the recorded test are discussed below. Criteria were controlled through a questionnaire that screened for age and factors for auditory processing disorders (APD). A hearing screen was also done to eliminate peripheral hearing problems. Results were recorded on a test form developed for the purpose of the study (Appendix 6).

3.4.3.1. Native language

The participants had to be native speakers of Afrikaans. This was stipulated during recruitment of the participants and confirmed during the completion of the screening questionnaire.

3.4.3.2. Age

The participants had to be above the age of 15 years, an age where concentration were expected to be maintained during testing. They also had to be under the age of

45, to minimise the influence of central presbyacusis, i.e. APD with no known cause other than natural aging (Weinstein, 2009, p. 715; Hanks & Johnson, 1998, p. 1336).

3.4.3.3. Peripheral hearing status

The participants had to have normal otoscopic findings and normal middle ear function, as confirmed through tympanometry using a 226Hz probe tone. A normal tympanogram was accepted as a middle ear pressure of -50 to +50daPa and compliance of 0.3 to 1.6 cm^3 (British Society of Audiology, 1992, p. 257). Participants with Jerger type C (Jerger, 1970, p. 312) tympanograms, i.e. with negative middle ear pressure, were also accepted as candidates in cases where hearing was still within normal limits during pure tone audiometry. Negative pressure in the middle ear space is quite common in many otherwise normal ears (Jerger, 1970, p. 312) and is often consistent with simply a sinus or allergy congestion, or the end-stages of a cold (Duffey, 2007, ¶4).

Peripheral hearing had to be within normal limits, i.e. 20dBHL or better at octave frequencies between 250 and 8000Hz (Jerger 1980, in Hall & Mueller, 1997, p. 104). This was confirmed through pure tone audiometry.

3.4.3.4. Central hearing status

Participants with normal central hearing were preferable, although central problems are more difficult to screen for. A case history was undertaken to exclude a history of excessive middle ear problems, sensitivity to loud sounds or late language development, as these are risk factors for APD (Cleveland, 1997, ¶9, 11 & 12).

3.4.3.5. Exposure to the translated test sentences

The participants were not allowed to have been exposed to the test material prior to testing. This was achieved through the selection of participants who were uninvolved with the development process of the sentence lists.

Table 3.2 contains a summary of the steps of the study as well as the sample sizes used in every phase.

3.5. ETHICAL CONSIDERATIONS

The necessary ethical aspects were considered in order to ensure protection of the participants and testers. These considerations are discussed in the following section.

3.5.1. Approval

Permission for the study was granted by the chairperson of the Health Research Ethics Committee of the Division of Research Development and Support at the University of Stellenbosch (Appendix 7).

3.5.2. Confidentiality

Confidentiality was ensured by using no participant names during reporting. The participants' names were also coded during data analysis.

Table 3.2: Summary of the steps in the study and sample sizes used

Phase	**Aim**	**Steps**	**Sample size per round**
Analytical	Identification of the problem	The need for a sentence test with high ecological validity was identified through communication with the Tygerberg Hospital Cochlear Implant Team.	N/A
Design	Compilation and recording of test material	Sentence lists were compiled.	N/A
		Naturalness of sentences were rated.	5, then 3
		Sentence lists and three-talker babble were recorded and edited	5
Developmental (Part 1)	Establishment of inter-list, test-retest and inter-rater reliability in the presence of babble	Pilot study: established test level of (SNR where 70% score in babble was achieved); determined whether 10 subjects were sufficient per round of testing	10
		5 rounds: Presented sentences in babble at level determined during pilot study and identified adjustments necessary to obtain similar scores across sentence lists Adjusted RMS	10 (only 9 in Round 2)
Developmental (Part 2)	Confirmation of performance at intended test levels in babble.	Presented 9 randomly selected sentences in babble at a +5dB SNR with babble level at 45dBHL on the audiometer	5

3.5.3. <u>Consent</u>

Every participant was asked to give written consent for participation in the study. This ensured that participation was voluntary.

3.5.4. Safety

Participants were protected from physical harm by being exposed to sound stimuli in a controlled environment at levels that were not harmful to hearing. Testing was undertaken by qualified audiologists with good knowledge of the equipment used.

3.5.5. Abnormal clinical findings

In instances where an abnormal clinical finding was found during screening, the participant, or the parent, in the event of a minor, were informed of the results. Recommendations were made with regards to further assessment or referral. In the event of minors with abnormal findings or conductive problems that had to be treated, the participants were telephonically followed up to ensure suitable management took place.

3.5.6. Cultural diversity

Linguistic diversity had to be restricted to using participants with Afrikaans as first language due to the nature of the test being aimed at an Afrikaans population. No restrictions with regards to socio-economic class or ethnic background were set.

3.6. TEST EQUIPMENT AND ROOMS

High quality test equipment was selected to ensure their reliability. The equipment that was used, is discussed below.

3.6.1. Studio

Recording took place in a sound-treated booth at a professional sound studio. The room used was resistant of external noise of up to 76dBSPL and had a balanced frequency response, allowing level fluctuations of no more than 3dBSPL. Recording took place at close range with microphone placement of a standard 3 inches (7.62cm).

3.6.2. Audiometers

Two Grason Stadler clinical audiometers, models GSI16 and GSI61, were used. They are calibrated annually according to the SANS (SysAdmin, Audit, Network, Security) standards of the South African Bureau for Standards (SABS), using SANS 10154-1:2004 for air conduction (SABS, 2004), SANS 10154-2:2004 for bone conduction (SABS, 2004) and SANS 8253-3:1996 (SABS, 1996) for speech testing. They are also checked weekly according to the recommended procedures for routine checks and participant tests on pure tone audiometers (Organisation Internationale de Métrologie Légal, 1993). These checks are applied directly to the speech feature on the equipment as well, which is an acceptable practice, being carried out on the same equipment (Sherwood & Fuller, 1997, p. 92).

3.6.3. Tympanometers

The tympanometers used, were Grason Stadler models GSI Tympstar and GSI38, which are also calibrated annually according to SANS 10154-1:2004 (SABS, 2001).

They are also checked subjectively as well as objectively on a 2cc and 0.5cc coupler on a weekly basis.

3.6.4. Sound Players

High quality domestic compact disc (CD) players (one Sony E-E220 ESP max and one Sanyo MCD-UB275M) were used with which the test material in CD format was presented through the audiometers. Good domestic CD players should easily meet the requirements of being able to handle the dynamic range of speech (about 30dB), with an output impedance that is matched to the input impedance of the audiometer, a harmonic distortion of less than 1% and the flattest possible frequency response over the major speech frequencies of 250Hz to around 4000Hz (Sherwood & Fuller, 1997, p. 94).

3.6.5. Test rooms

Testing was conducted in sound-treated booths. The booths adhered to the standards of the NIST's (National Institute of Standards and Technology, 2003) NVLAP (National Voluntary Laboratory, Accreditation Program), laboratory code 100286-0.

3.7. PROCEDURE

Once the need for the development of a speech-in-noise test was identified, the development of the lists was done as described below.

3.7.1. Design Phase

The design phase of the study entailed the compilation of the sentence lists as well as the recording thereof. These steps are discussed in more detail below.

3.7.1.1. Compilation of Afrikaans sentence lists

The chosen sentences and type of masking noise were carefully considered, in order to make the temporal and level characteristics consistent with those found in everyday listening situations. This consideration would improve the face validity of the test (Bentler, 2000, p 84).

A preliminary translation of the CUNY topic-related sentences into Afrikaans had been done by an experienced audiologist for the Cochlear Implant Unit at Tygerberg Hospital in anticipation of standardisation thereof. This prepared set consisted of a translation of the original 48 CUNY topic-related sentence lists, with the necessary adjustments made to replace sentences that were not culturally appropriate for South Africa, but did not account for the required amount of words per sentence. For the current study, the first 24 sentence lists of this translation were used as a starting point. A concurrent study ran in parallel to develop a larger database of lists, which used the translation of the 24 remaining lists. For the current study, a further three of 12 lists later added to the original CUNY topic-related sentences, labelled as lists 49 to 51 of the CUNY topic-related sentences, were added. These three lists were intended to be used as practice lists, but were then included for full development, in order to get a more extensive library of sentence lists. Three of the concurrent study's lists were used instead as practice lists for the current study (see Appendix 8).

The original lists 49 to 51 of the CUNY topic-related sentences now became lists 25 to 27 of the current study, totalling a number of 27 lists for development. The lists were adjusted to contain the required amount of words per sentence, each list containing one sentence each of three to 13 words, with the exception of two sentences containing 12 words. The sentences were further adjusted to still keep the amount of words the same but to increase or decrease the amount of syllables in order for the syllables of each list to be between 134 and 142, since scoring on syllable level was strongly considered at this point. During compilation of the sentences, particular emphasises was laid on using common vocabulary and sentence structures, and also avoiding words that are confined to particular dialects of the language. This was deemed necessary because uncommon words have a lower intelligibility than common words (Savin, 1963, p. 200).

3.7.1.2. Evaluation of the naturalness of the sentences

The use of commonly encountered words in the SAV-L was verified by getting speakers of Afrikaans to read through and identify unusual words according to their frame of reference, prior to recording of the sentence lists. Five participants were asked to rate the naturalness of the sentences, to ensure high content validity. The five participants were expected to read through the sentences and rate each on a scale from 1 to 5 for their naturalness (Soli, 2003, p. 1). A score of 5 was for a sentence that was considered completely natural and would definitely be encountered in everyday speech. A score of 1, in contrast, was considered completely unnatural. If a sentence scored less than 5, the participants were required to suggest a modification. These modifications were considered and adjusted by the author. Where a sentence was given a score of 4 by only one rater and given a full score of 5 by the other raters, amendments were considered but not necessarily made. If more than one rater gave a score of less than 5 for a sentence, the sentence was adjusted. After changes were made, the sentences were submitted to a further three raters and amendments were made accordingly.

3.7.1.3. Recording of the test material

During recording of the sentence lists, the speaker was instructed to speak as naturally as possible, since naturally spoken sentences have a greater dynamic range, and therefore give a more valid representation of real speech (Villchur, 1982, as cited in BKB-SIN manual, n.d., p. 17). The average intensities of the sentence lists were equalised by adjusting the lists' average root mean square (RMS) values, as had been done in the development of previous sentences tests (Nilsson et al., 1994, p. 1087; Hirsh, Davis, Silverman, Reynolds, Eldert & Benson, 1952, p. 325).

Multi-talker babble using cold running speech was recorded as masking noise. Two female and two male talkers were instructed to speak naturally to maintain level variations (personal correspondence with Carver, 1991, by Killion et al., 2004, p. 2396). Increased level variation makes equalisation more complicated, but it ensures that the application of the test can be extended to evaluate the use of compression hearings aids, which may clamp into a fixed-gain setting, if the masking noise is kept too constant (Killion et al., 2004, p. 2396). Each talker read passages that were recorded independently and three of the recordings were subsequently electronically mixed (Sperry, Wiley & Chial, 1997, as cited in Wilson et al., 2005, p. 500) to form three-talker babble.

The variable level of multi-talker babble was described by Killion et al. (2004, p. 2398) as an "ebb and flow level". This inevitable varying level results in variations during performance in recognition of the stimulus sentence, if the masking noise changes during repeat presentation of the same stimulus. To control this variation in masked test conditions, a set block of the recorded babble was used with the stimulus sentences being superimposed upon them in a second channel, so that the same babble would occur in the background of each particular sentence. The recording was thus *time-locked* (BKB-SIN manual, n.d., p. 19) against the background babble, to ensure repeatability of responses.

The sentences had intervals between them consisting of the length of the particular sentence together with an additional 3 seconds, to allow time for the participants to repeat the sentences.

The recording was exported to a compact disc together with a 1kHz pure tone with a duration of 60 seconds as calibration signal (SANS 8253-3). The calibration signal was added as track and served to control the stimulus level. Recording levels were set so that the average RMS speech levels were the same as the peak amplitude of the calibration tone.

3.7.2. Developmental Phase

The first part of the developmental phase consisted of the adjustments of the sentence lists to approximate equal difficulty in the presence of three-talker babble, i.e. increasing inter-list reliability. This phase also involved a second and third part to determine inter-rater and test-retest reliability. It was completed with confirmation of the intended test level.

Testing in a sound-treated booth served to control extraneous noise as an independent variable. The stimulus and masking noise were presented diotically through TDH-50 supra-aural headphones to ensure that stimulus levels were unaffected by participant placement. Participants were frequently offered time to rest, in order to control for fatigue and concentration as much as possible.

3.7.2.1. Establishment of inter-list equivalence

The sentence lists were presented in rounds to groups of 10 participants at a time (with the exception of Round 2, where only nine participants were used because testing was terminated) to obtain an average percentage score of the sentence lists, which could be compared to establish the lists' equivalence. Following each round of presentation, the average participant scores for every list, together with the overall average scores of the lists were calculated. The lists with maximum and minimum average scores were also identified and the difference between these scores calculated. These differences were used to judge whether any level adjustments were indicated in instances where inter-list variance was too large. Sentences that achieved high scores were reduced in intensity, and sentences that achieved low scores were increased in intensity. The average scores of the lists were used to calculate adjustments in intensity levels of the sentence lists necessary to approximate equal intelligibility. For each subsequent round of testing the increment of adjustment was reduced more and more until sentences approximated equal intelligibility, as was done in studies by Wong et al. (2007), Schafer and Thibodeau (2006), Hällgren et al. (2006), Vaillancourt et al. (2005), Wong & Soli, (2005), Kollmeier and Wesselkamp (1997), Nilsson et al. (1994) and Smoorenburg (1992). These adjustments were aimed at increasing inter-list reliability.

During presentation of the sentences to determine their equivalence, truncation effects, i.e. levels where scores of 0% or 100% are obtained, had to be avoided. The reason for this was that SNRs where truncation effects are present, do not provide useful information about the difficulty of one stimulus in comparison to another stimulus. Scores below 20% are often affected by floor effects since the task difficulty is too great to show participant changes in performance, whereas scores above 80% are often affected by ceiling effects because the task difficulty is too easy to be sensitive for performance changes (McArdle & Hnath-Chisolm, 2009, p. 67). A pilot study was therefore done using five participants to find a SNR where participants scored between 20% and 80%.

The most sensitive region of the PI curve is in the range 65% to 90% (Dillon, 1983, p. 340). It was therefore decided to test at an SNR where approximately 70% was reached. An adaptive procedure, where the sentence intensities in one list are adjusted to find a pre-determined percentage score, could not be used, because the sentences in the test differed in length, and therefore in difficulty. The test was also in its unadjusted form at this stage, therefore the sentence lists were not expected to be of equal difficulty to allow an adaptive test procedure. The SNR level at which approximately 70% was achieved was therefore determined by presenting a whole sentence list at a high SNR of +10dB to a participant whilst getting a percentage correct score. A next list was subsequently presented, while decreasing the SNR by 2dB. This was repeated with subsequent lists until a level was found where an approximate score of 70% was obtained. Another list at the same SNR was presented to confirm the level. The procedure was repeated with four other participants to confirm the SNR. Through this method it was established that a SNR of +1dB was a suitable level to test at, in order to avoid truncation effects.

The lists were subsequently presented at 46dBHL[1] at a +1dB SNR (masking level therefore at 45dBHL) to a group of 10 participants to complete the pilot study. These results were analysed to establish whether a group of 10 participants generated sufficient data to be a statistically significant measure in determining differences in intelligibility of the sentences and sentence lists. Presentation order was randomised across participants, as was done throughout the subsequent rounds of testing, to exclude participant fatigue as a variable (Bamford & Wilson, 1979, p. 159). For all participants clear instructions were given, instructing them, 'Jy/u gaan stemme hoor wat saamgesels en dan gaan 'n manlike spreker tussen hulle sinne van verskillende lengtes sê. Die man se stem waarna jy/u moet luister, is dieselfde stem as wat aan die begin van die sinslys byvoorbeeld 'Sinslys 7' aankondig. Herhaal asseblief elke sin sodra dit gesê is. Selfs as jy/u nie die hele sin gehoor het nie, herhaal asseblief soveel as moontlik wat jy/u dink gesê is, selfs al is dit net 'n deel van 'n woord. Jy/u

1. dBHL refers to the dial setting on the audiometer

is welkom om te raai'. This translates as 'You will hear voices talking together and then a male speaker amongst them saying sentences of varying lengths. The man whose voice you need to listen to, is the same voice that announces for instance 'Sentence list 7' at the beginning of each sentence list. Please repeat each sentence as soon as it is said. Even if you do not hear the whole sentence, please repeat as much of the sentence as you think you heard, even when it is just part of a word. You are welcome to guess'.

During the pilot study, list 15 was used as a practice list and scoring took place at word and syllable level to establish whether one method of scoring rendered much different results in comparison to the other method. Although scoring at word level was strongly considered, Afrikaans tends to follow a conjunctive pattern i.e. having a tendency to write compound words as single words as opposed to a language like English, which follows a disjunctive pattern, i.e. a tendency to write words separately (Carstens, 2003, p. 232). This conjunctive pattern led to uncertainty as to which compound words had to be separated to be scored as separate words rather than one. An example of this would be the word 'drieslaapkamerwoonstel' (three bedroom flat), which, although it is written as one word, for the purpose of scoring at word level could be argued that it has to be scored as three words, because it has three very separate parts to it. Other words like *droogskoonmaker* (*dry-cleaner*), *troeteldiere* (*pets*, which, if directly translated, would be *petting animals*), *waspoeier* (*washing powder*), *slaapkamer* (*bedroom*) and *netbal* (*netball*) could all be divided into smaller meaningful words as well. However, it would be less clear as to whether they need to be scored as one or two words, because even with two concepts in these words, even in English many of these words are written as one.

The average percentage score per list was equated and compared in order to determine inter-list within-participant repeatability/reliability (Nilsson et al., 1994, p. 1089).

Following the pilot study, the individual sentences were adjusted by 0.7dB for every 10% it differed from the average, as was done in the beginning rounds of the HINT's development (Kollmeier & Wesselkamp, 1997; Nilsson et al., 1994; Smoorenburg, 1992).

Round 1 and subsequent rounds were started with three practice lists from the sentence material in the study that ran concurrent with the current study. A number of three practice lists is the amount of lists found to be necessary before performance stabilises (MacLeod & Summerfield, 1990, p. 40).

Following Round 1, the intensities of the sentences were adjusted with 0.5dB for every 10% they differed from the average score. Because scores dropped significantly in Round 1, the SNR during Round 2 was increased to +2dB with the masking noise still kept at 45dBHL. Scores remained poor during Round 2 and only 9 participants were tested in this round since it was decided at this point to investigate why the scores had dropped so significantly.

Prior to Round 3 the RMS levels of the individual sentences were reset to be equal in intensity. Round 3 therefore served as the start of a new comparison. Prior to the pilot round, the average RMS levels of the lists as a whole were adjusted, while prior to Round 3 it was decided to adjust the average RMS levels of the individual sentences. The SNR at which 70% was obtained, therefore had to be re-established with the same procedure as prior to the pilot round. A SNR of -3dB was determined to be suitable. In subsequent rounds the stimulus was therefore presented at 42dBHL with the masking level kept at 45dBHL.

It was not expected that the sentences of different lengths would exactly match in difficulty in view of their differing lengths. The intensities of the lists as a whole were therefore from here on adjusted, as opposed to individual sentences level adjustments, as in the first three rounds. The exception was with individual sentences that already had a score of 0% in instances where a reduction in intensity was indicated, or a score of 100%, in instances where an increase in intensity was indicated. For these sentences the intensities were not adjusted, because the baselines and ceilings in these sentences had been reached, and would have been exacerbated if the intensities had been further reduced or increased. Following Round 3, the sentence intensities were adjusted with 0.5dB for every 10% they difference from the average of that round.

This procedure was repeated for Round 4, after which sentences were adjusted by 0.3dB for every 10% difference from the average score achieved for the lists, finishing with Round 5 as a last comparison post adjustment of list intensities.

3.7.2.2. Establishment of test-retest reliability

Test-retest reliability was established by retesting 10 participants from the pilot round and Round 1 with at least a six month delay to minimise contamination of the results with memory factors (Bamford & Wilson, 1979, p. 162). Variables were controlled as carefully as possible by duplicating them. These included the instructions given, the presentation order of lists, recordings, SNR, testers, room and equipment used, and rescreening the participants' hearing status to ensure hearing status remained unchanged.

3.7.2.3. Establishment of inter-rater reliability

Throughout the first four rounds of testing, a second tester randomly scored participant responses in parallel with the main tester. The scores of the two testers were compared to determine inter-rater reliability by establishing the intra-class correlations.

3.7.2.4. Confirmation of intended test level

In its final form, the sentence lists were presented at the intended test level to five participants, in order to confirm that the scores for persons with normal hearing were 100% as expected. By knowing the scores of persons with normal hearing, the disadvantage of a person with a hearing loss can be established. This data can then be used to establish auditory device candidacy once guidelines are developed.

The level at which the test should be administered, was chosen as the level where the 'standard' group, i.e. persons with normal hearing, were expected to obtain a 100% score. The chosen SNR should not be too easy or difficult, because this could under- or overestimate the person's abilities (Madell, 2010, slide 21; Taylor, 2003, p. 41).

Nine lists were randomly chosen from the final version of the test material and presented to five new participants at 50dBHL with an SNR of +5dBHL, to confirm that a 100% score was obtained. A level of 50dBHL is recommended to maximise audibility and to represent sound levels found in most typical social gatherings, commonly 65 to 85dBSPL (BKB-SIN manual, n.d., p. 10). An SNR of +5dB is a ratio that is encountered in daily listening conditions (Olsen, 1998, p. 25; Firszt et al., 2004, p. 376) and where a 100% score was expected from listeners with normal hearing.

In instances, where the expected 100% score was not obtained, the SNR level was increased to +6dB, keeping the three-talker babble level at 45dBHL. A further 9 lists were presented at this level. A final set of 9 sentences was then presented at +7dB, again keeping the three-talker babble level constant at 45dBHL. An analysis of the mistakes was done to look for trends.

3.8. SUMMARY

Chapter 3 detailed the steps taken in the development of the first phase of the SAV-L. It stipulated how the first 27 lists of the SAV-L were developed and recorded, after which field studies were undertaken to establish the inter-list, inter-rater and test-retest reliability. Lastly the procedure of testing the material at the intended test level was also described.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. INTRODUCTION

Chapter 4 discusses the results, learning experiences and outcomes of the study, which entailed the development of the first 27 sentence lists of the *Sinslyste in Afrikaans vir Volwassenes in Lawaai* [Sentence lists in Afrikaans for Adults in Noise] (SAV-L). The results and outcomes are discussed following the chronological order of the test's development, starting with the design phase, during which the sentence lists were compiled and recorded, followed by the developmental phase, during which the recorded lists were adjusted and tested on participants with normal hearing. Particular emphasis will be placed on the validity, reliability and sensitivity of the test, and how these were influenced by the various processes of the test's development.

4.2. THE DESIGN PHASE

The design phase formed the foundation of the SAV-L's development and consisted of the compilation of sentence lists, the evaluation of the naturalness of the sentences, followed by the recording of the sentences and three-talker babble. The various procedures and strategies employed, choices made as well as their rationales and impacts during the design phase will be discussed.

4.2.1. Compilation of the sentence lists

The sentence lists of the SAV-L were based on the CUNY topic-related sentences. Because the CUNY topic-related sentences are frequently used during cochlear implant evaluations and have been found to be useful (Zwolan, 2009, p. 922), this test format was assumed to be suitable.

Some of the sentences of the SAV-L were a direct translation from the original CUNY topic-related sentences. Others were firstly translated and subsequently changed after they scored poorly during the rating of the sentences' naturalness. A third group of sentences consisted of completely new sentences in instances where the original CUNY sentences used vocabulary that was more applicable for the American culture (e.g. 'You can see *deer* near my country house' and 'The football field is right next to the *baseball* field'). These new sentences still focussed on the same topics as the original CUNY topic-related sentences.

Figure 4.1 illustrates the distribution of the methods of sentence formulation of the total of 324 sentences developed for the test. In 50.93% of the cases (N= 165), new sentences were compiled, maintaining the correct sentence length and same topic but ensuring the use of culturally appropriate words. In 36.73% of cases (N = 119) the original sentences were directly translated and maintained. In 12.35% of cases (N = 40) the sentences were directly translated but adjusted in response to feedback from the raters of naturalness.

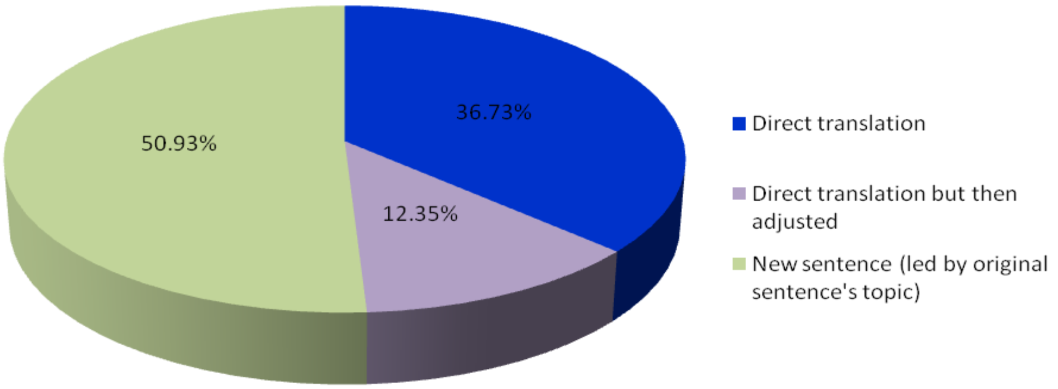


Figure 4.1. Manner of sentence formulation re: original CUNY topic-related sentences

Despite the SAV-L not being a direct translation of the CUNY topic-related sentences, the test features of the CUNY topic-related sentences that were kept, were seen as necessary to ensure a close simulation to natural speech-in-noise listening environments, i.e. obtaining high ecological validity (Mackersie, 2002, p. 395). These test features included the use of sentences that consisted of varying lengths and structures as stimuli, use of more complex vocabulary suitable for adults, topics relevant to the specific culture, as well as the addition of multi-talker babble as masking noise, in this instance three-talker babble.

Careful and correct selection of test construction features will lead to a high degree of construct validity (McAllister, 2003, p. 177). The measurement of speech perception with speech itself as a stimulus, in a format (sentences) that is encountered in everyday life, was seen to strengthen construct validity. A contributing feature was the simulation of the frequently-encountered cocktail party problem (Cherry, 1953, p. 976) by using three-talker babble as masking noise. In this instance achieving high ecological validity (Mackersie, 2002, p. 395) also achieved high construct validity.

The choice of stimuli and masking noise used in the present test were considered to give the test high face validity (Nunnally & Bernstein, 1994, p. 110) for the assessment of speech perception as well. As a sacrifice to obtaining high face validity, the population that the test is used on, must however be restricted to adults with good language skills due to the level of language used in the test. This will ensure that participants are not put at an undue disadvantage, should their language abilities be inadequate.

Conclusions drawn from the design phase of the study were as follows:

1. More than 50% of the sentences of the SAV-L could not simply be directly translated from the CUNY topic-related sentences, but had to be adjusted in order to ensure cultural appropriateness. Even where translations were directly made, these sentences often had to be adjusted again in order for naturalness to be rated highly. Cultural appropriateness therefore seemed to be a significant factor to consider in the compilation of speech perception tests and had to be considered in order to ensure high ecological validity, since it is known that cultural biases can affect a test's validity (Kerkhoff & Vallen, 1984, p. 133).

2. The SAV-L has high ecological and construct validity. It also has high face validity, providing it is used with the intended population.

4.2.2. <u>Evaluation of the naturalness of the sentences</u>

High content validity was ensured by getting the naturalness of the SAV-L's sentences rated by native speakers of Afrikaans, i.e. the language the test was developed in. This is a widely used means of validating sentence lists (Wong et al., 2007; Schafer & Thibodeau, 2006; Hällgren et al., 2006; Vaillancourt, et al., 2005; Wong & Soli, 2005; Nilsson et al., 1994), and was therefore seen as an acceptable way of evaluating the naturalness of the sentences.

The task of the raters was to identify whether the sentences were likely to be encountered in everyday life. Eight independent persons rated the sentences on a scale of 1 to 5 with regards to their naturalness. As explained in 3.7.1.2, a score of 1 was awarded to a sentence that did not sound natural at all, while a score of 5 reflected a very natural sentence, which one would encounter in everyday conversation. The rating took place over two rounds, with adjustments made after each round based on feedback from the raters. The lowest score obtained on any one sentence was 3 (e.g. for the sentence 'Sal jy dit oorweeg om 'n kamer aan die huis aan te bou?' [Will you consider building a room onto the house?].

The sentences were colour-coded to indicate their relative need of adjustment. Green flag sentences were those where all the raters awarded them a 5 out of 5 score. These sentences were not changed, unless other sentences in the particular list affected the distribution of sentence lengths which could not be corrected in any other way than by adjusting some green flag sentences. Sentences were flagged yellow, if only one rater scored them 4 out of 5. In the instance of a yellow flag, a change was considered but not necessarily made. The changes were not made when the changes suggested by the scorers were not considered linguistically correct or judged not to be standard Afrikaans (e.g. 'Koop jy die kos, dan sal ek die drankies vir die braai kry' [If you buy the food, I will get the drinks for the barbeque]). Red flag sentences were those that had to be changed, because they received scores of 3 or less out of 5 from one of the raters, or 4 or less from two or more raters. An example of a changed sentence was 'Jy moet soms jou kinders hulle eie foute laat begaan' [You must sometimes allow your children to make their own mistakes], which was changed to 'Jy moet jou kinders uit hulle eie foute laat leer' [You must let your children learn from their own mistakes]. Figure 4.2 shows the improvement in ratings made over the two rounds, i.e. the reduction in red and yellow flag sentences and the increase in green flag sentences.

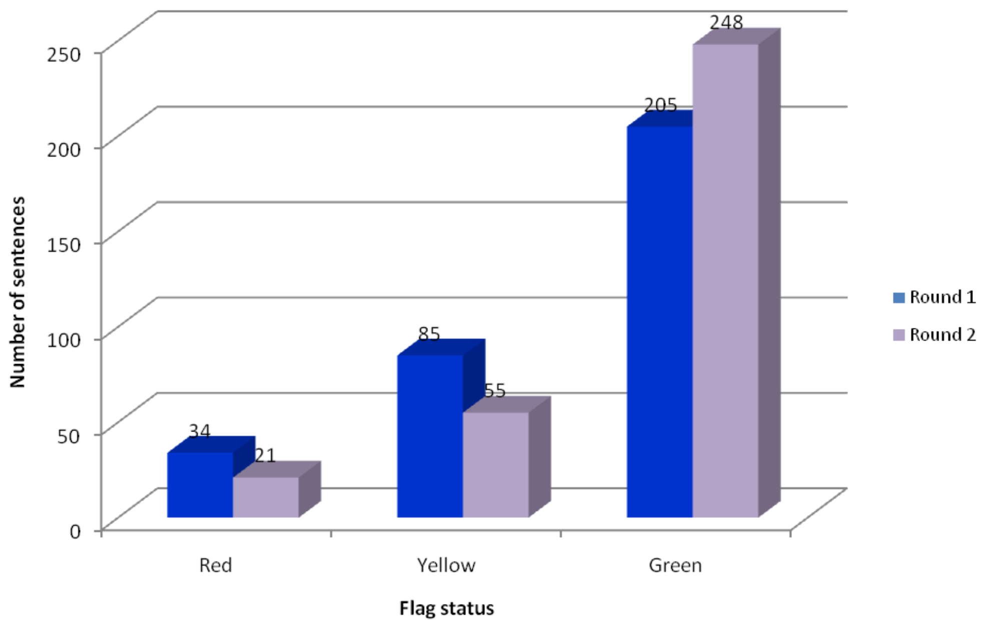


Figure 4.2. Improvement over the two rounds of naturalness rating (total N=324)

Key: Red flag = a sentence that had to be changed because it was rated 3 or less out of 5 by one of the raters

Yellow flag = only one rater scored the sentence 4 out of 5, while the other raters scored the sentence 5 out of 5

Green flag = a sentence that were rated 5 out of 5 by all the raters

Figure 4.3 gives a detailed flowchart of the status of the sentences following each round of corrections made.

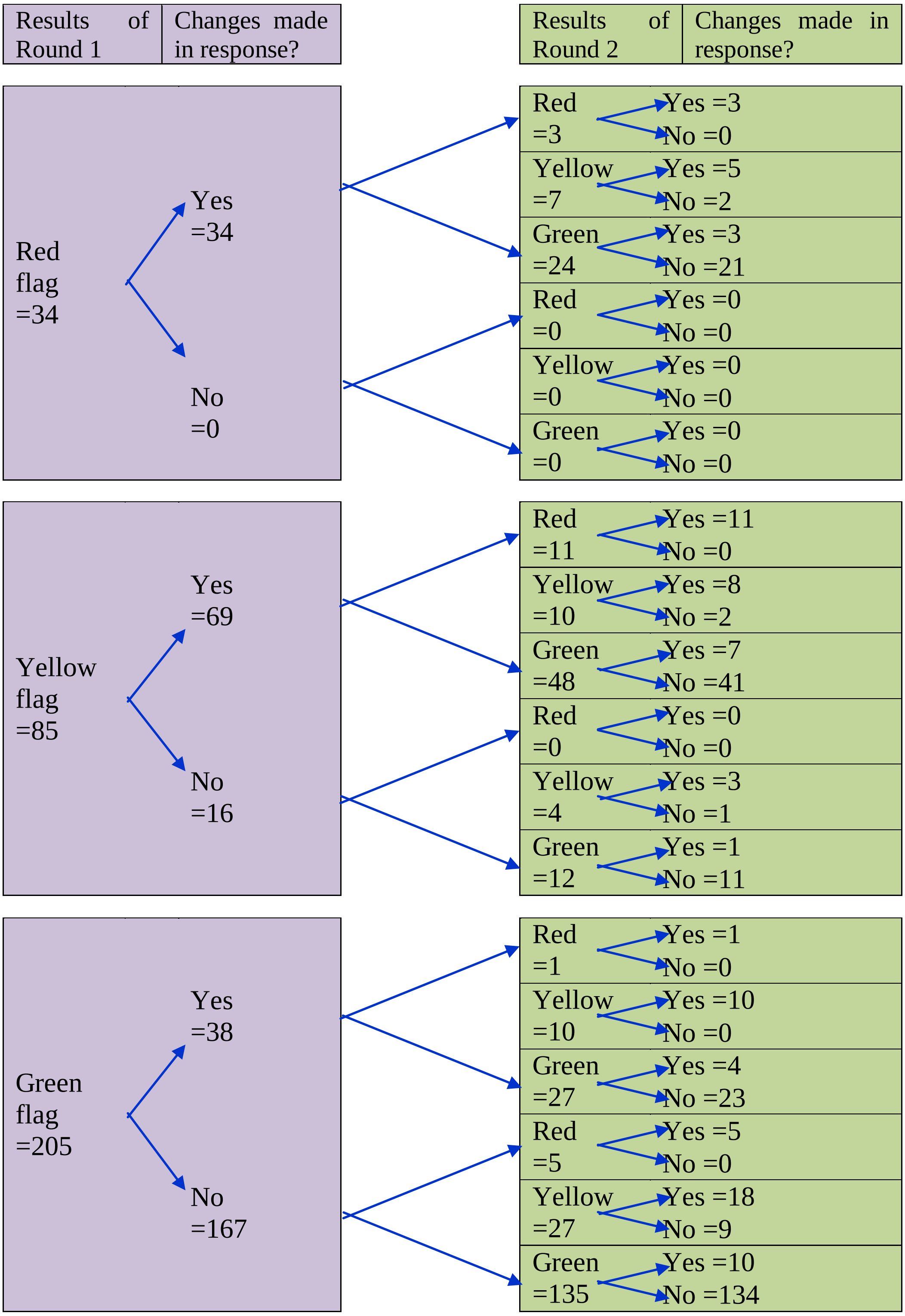


Figure 4.3. Flowchart of the number of flag statuses following two rounds of ratings and subsequent adjustments (total N = 324)

During the first round of rating, 205 sentences were identified as green flag. Of these, 38 were changed to maintain the required distribution of sentence lengths, which were disturbed due to changes to other sentences. After these changes, one sentence became red flag and ten became yellow flag in Round 2. The red flag as well as all the yellow flag sentences were then changed again.

A total of 32 sentences that were rated as green flag during the first round of rating, were rated in Round 2 of rating as yellow or red flag and subsequently had to be considered for further adjustment. These were adjusted in all instances where a red flag was obtained (5 sentences), and most of the times when sentences were flagged yellow (18 out of 27 times), leaving 9 yellow flagged sentences unchanged, because the suggested changes were not deemed linguistically more correct.

During Round 1 of rating, 85 sentences were identified as yellow flag sentences, i.e. possible candidates for change. Of these, 69 sentences were changed when it was found that the suggested changes were linguistically correct. The majority (48) changed to green flag status, of which 7 underwent further adjustments, again to accommodate for changing sentence lengths of other adjusted sentences in the particular lists. Ten maintained their yellow flag status. Of these, eight were changed again. Eleven sentences turned to red flag status in Round 2 and had to be changed again. Sixteen yellow flag sentences were not adjusted. Four remained yellow flag status in Round 2, of which three were subsequently adjusted. Twelve became green flag, of which one was changed following Round 2.

In Round 1 of rating, 34 out of the 324 prepared sentences were identified as red flag sentences, i.e. definitely in need of change. Consideration was given to the suggested changes and all the sentences were amended accordingly, whilst still maintaining the required length of each sentence. The sentences flagged red mostly involved sentences that were deemed too formal, and were converted to sentences that were deemed more standard spoken language. Twenty-four of the sentences changed to green flag status, but then three of these were changed again to accommodate changes in sentence lengths resulting from changes to the yellow and red flag sentences in the particular lists. Three of the original red flag sentences remained red flag and had to be changed again. Seven sentences changed to yellow flag status, of which five sentences were amended again.

In Round 2 of rating, a total of 20 red flags, 58 yellow flags and 246 green flags were obtained. All the sentences with red flags and 44 with yellow flags were corrected. Yellow flags were ignored if the sentence had obtained a green flag in the first round, because it was then not seen as a significant flag in the second round. The yellow flag sentences were also ignored if they had already undergone an adjustment in the previous round. The changes made in the second round were strongly influenced by the suggestions of one particular rater, who is a proof-reader. Following the adjustments, the sentences were considered as the final version for recording. This final version is presented in Appendix 9.

During the rating of the sentences, no problems were identified with regards to topics used, which led to the assumption that the topics were fairly familiar (widely encountered). Later on, during the presentation of the recorded sentence lists to participants, it was however observed that, when the SNR was challenging, certain participants, on mishearing words, would often substitute them with words from topics encountered in their specific daily context. A mother of two children, for example, often replaced other nouns with the word *children* or words from home-related topics, while a building consultant scored much higher on sentences relating to finance, a topic often encountered in her working life. Although this aspect was not specifically assessed, it indicated how lexical access seemed to influence performance. Numerous studies have shown that common words are easier to recognise than rare words, and this is often measured by presenting words in noise (Paivio & Begg, 1981, p. 128). The observed response bias was described by Broadbent (1967, p. 2) as a *pure guessing* bias, where guesses were more likely to be common than rare words. Recognising common words was explained by Carroll (2008, p. 57) as an automatic task, requiring fewer substantial resources from the limited pool of linguistic resources. During pure guessing bias, it could therefore be concluded that fewer of the limited linguistic resources were used when a common word was used as a guess.

During the first round of rating, it was also observed that breaking certain language rules was not identified as unusual or unnatural. An example of such a broken rule was that in Afrikaans, adverbs are required to be presented in the following sequence: time, then manner and lastly place (Carstens, 2003, p. 55). Only in the second round of rating was the offence of this rule identified by one of the raters, who is a proof-reader. Even though it is doubtful that errors like these would affect scores when the test is administered, these observations were regarded as important in striving towards language purism, and were consequently corrected. In contrast, unusual words or phrases identified by the raters were seen as a variable that could potentially affect test scores, and were amended to ensure increased content validity.

Conclusions drawn from the evaluation of the sentences' naturalness were as follows:

1. High content validity was ensured through the process of rating the naturalness of the sentences.
2. The topics of the test were considered appropriate. However, sentences with more familiar topics might have been more advantageous to the participants than those with less familiar topics.
3. Breaking certain language rules will not necessarily affect perceived naturalness of sentences.

4.2.3. Recording of the sentences

Following the recording of the sentence lists by the main speaker, the average root mean square (RMS) levels of all the sentences lists were adjusted to be at an equal average level of -4.59dBFS[1]. This allowed a starting point from which the intelligibility of each sentence list could be assessed.

The compilation of the three-talker babble led to an average RMS level of -6.19dBFS. The varying level of the babble was not adjusted in any way but kept in its original form to simulate real-life conversation as closely as was allowable in a controlled speech perception test. The difference in the average RMS levels of the

1. See 'Definition of Terms' in Chapter 1 (p. 15) for discussion of dBFS

sentences and masking noise was not seen as problematic as long as the stimulus sentences were time-locked against the three-talker babble.

During the recording, the main speaker's voice had a steadier average RMS in comparison to the background speakers. This was attributed to him being a very experienced voice artist. However, the reduced variation of his voice was not seen as an obstacle, because he still followed instructions to speak naturally.

Although it was decided to use three talkers for the multi-talker babble, a fourth speaker was also recorded in case it was found that three-talker babble led to too many quiet periods. However, as no excessive quiet periods were present with three-talker babble, the fourth speaker was not needed. A sufficient energetic masking noise component was therefore present with three talkers whilst an optimum informational masking noise component was still maintained (Carhart et al., 1975, p. S35).

The following conclusions were made from the recording of the sentences:

1. Time-locking the stimulus sentences against multi-talker babble was essential to control the varying intensity levels (the energetic masking component) and linguistic distractions (the informational masking component) of the babble.
2. Three-talker babble provides sufficient energetic masking whilst maintaining optimal informational masking

4.3. THE DEVELOPMENTAL PHASE

The developmental phase consisted firstly of presenting the sentence lists over six rounds to participants in order to establish the average percentage scores. Adjustments in the intensity of the lists were made following each round based on these scores, in order to attempt approximating equal intelligibility of the lists, therefore increasing inter-list reliability. The data of the first four rounds was also used to evaluate inter-rater reliability by getting a second scorer to randomly co-score responses to some of the sentence lists.

A seventh round of testing was undertaken to evaluate test-retest reliability, while the eighth and final round of testing was used to present the sentences at their intended test level in order to confirm the performance of persons with normal hearing.

Table 4.1 summarises the work that was done over the eight rounds of testing undertaken during the developmental phase. The results and outcomes of all of these rounds are discussed next.

Table 4.1. Summary of testing undertaken during the developmental phase

Goal	Method	Round of testing involved
Establish inter-list reliability	Presentation of test material to participants and adjusting intensities based on scores obtained, in order to approximate equal intelligibility	Pilot round and Rounds 1 to 5
Establish inter-rater reliability	Scoring of some of the sentence lists by 2 raters in order to establish influence of inter-rater variance on scores	Pilot round and Rounds 1 to 3 (at random)
Establish test-retest reliability	Presenting test material to the same participants with at least a 6 month period between testing, duplicating test conditions	1st testing: Pilot round and Round 1 Repeat testing: Round 6
Confirmation of intended test level	Presentation of the test material in its final form at an SNR of +5dB, with babble at 45dBHL and stimulus sentences at 50dBHL	Round 7

4.3.1. <u>Inter-list reliability</u>

During the first six rounds of testing (the pilot round and Rounds 1 to 5), inter-list variability was determined and sentence intensities were adjusted to reduce this variability. The results from the different rounds of testing are consequently discussed.

4.3.1.1. Inter-list reliability of the pilot round, Rounds 1 and 2

Following the pilot round**,** the average scores of the individual lists were calculated at word as well as syllable level (see Figures 4.4 and 4.5). Results showed similar results between word and syllable scores, with syllables scoring slightly higher with an average score of 72.17% versus an average score of 71.47% for words. The variability of the average scores of the lists, i.e. the difference between the average maximum and minimum scores of the lists, was 49.34% at syllable level, with list 2 scoring the highest at 86.72%, and list 24 scoring the lowest at 37.38%. At word level the difference between the maximum score (list 6 at 85.80%) and minimum score (list 24 at 35.40%) was 50.40%. The maximum scores were comparable with the average test scores, but the minimum score of list 24 was particularly low. Closer examination showed no trend to the kind of mistakes made. Two sentences in particular did however score consistently low (the sentences 'Pos jou kaartjies betyds' [Post your cards on time] and 'Koop daardie nuwe klavier terwyl die winkel so 'n groot uitverkoping het' [Buy that new piano while the shop has such a big sale]). No apparent explanation was found for the poor score of these sentences and the rest of the list. Acoustic analysis of these sentences in order to explain the poor score of these particular sentences was seen as beyond the scope of the study.

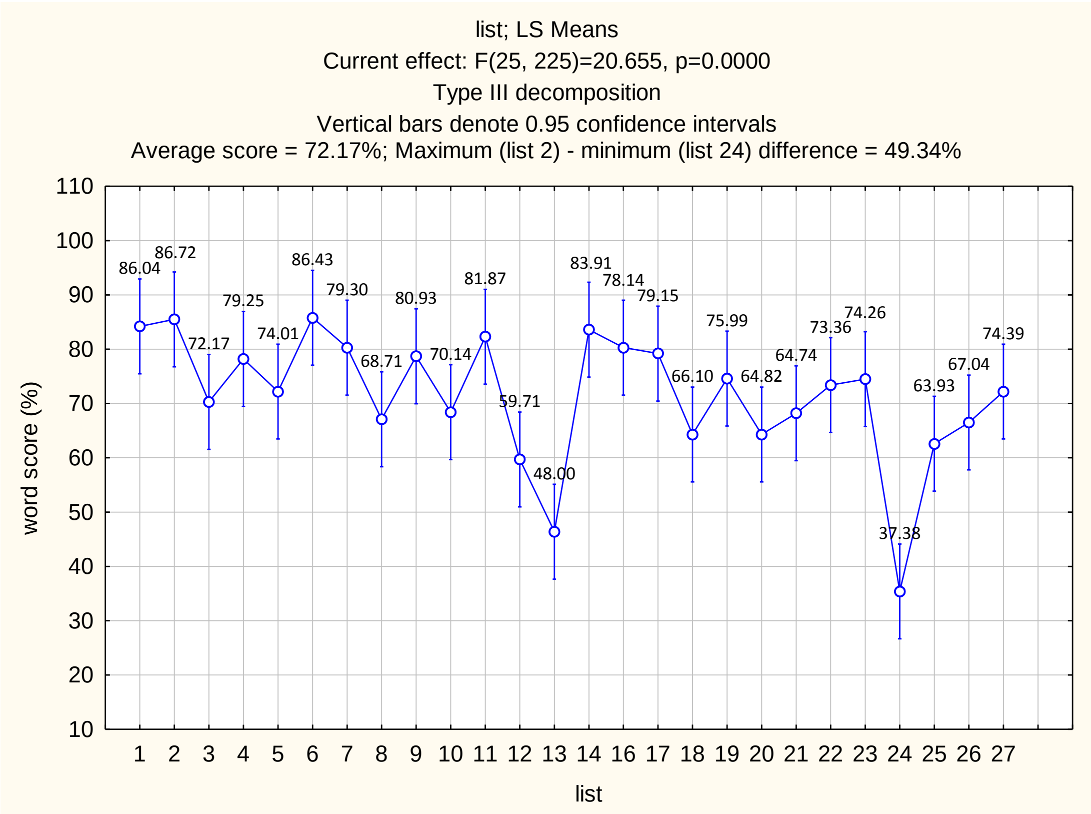


Fig. 4.4. The average scores for the sentence lists following the pilot study, scored at word level

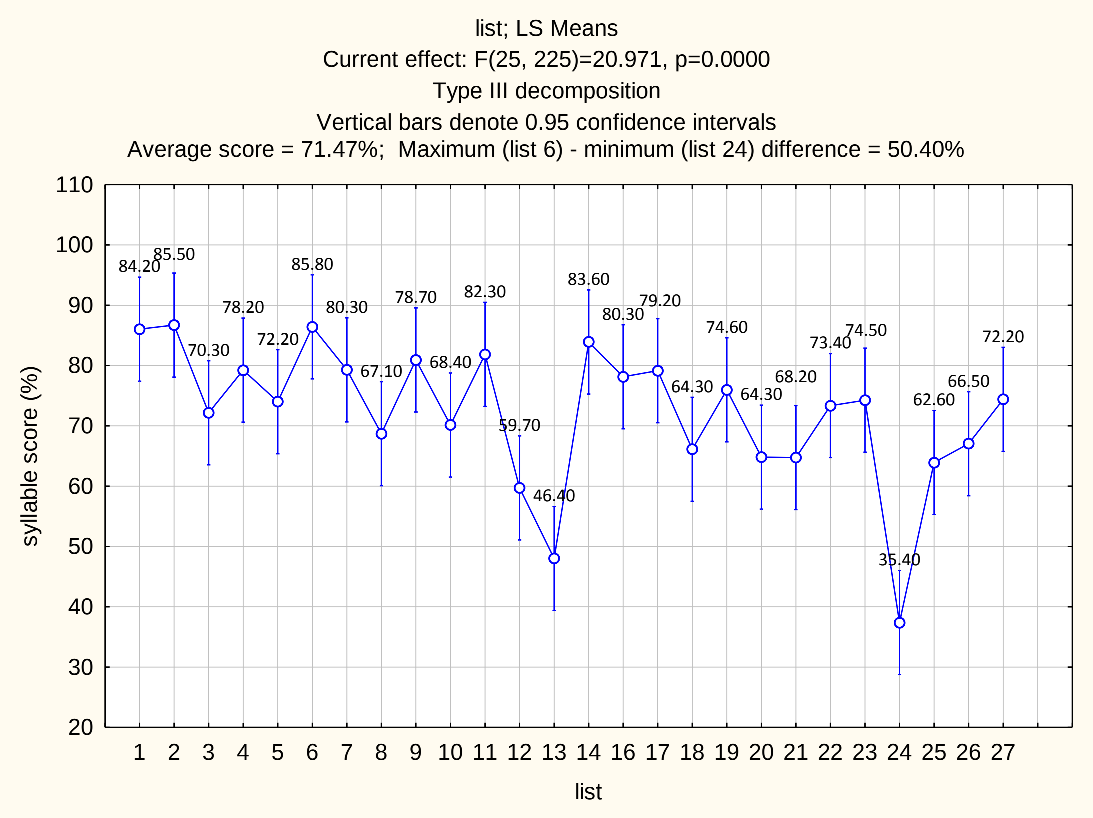


Fig. 4.5. The average scores for the sentence lists following the pilot study, scored at syllable level

When comparing the average scores of the lists when scoring at word versus syllable level, the largest difference in scores was for list 21 with a difference of 3.46% (see Table 4.2). Because of the small differences in scores on word and syllable level, as well as the uncertainty of what to label as individual words, when scoring at word level, due to the conjunctive nature of Afrikaans semantics, it was decided to score only at syllable level in subsequent rounds. In addition, the choice of syllables as scoring units kept the amount of test items as high as possible, which should minimise variance (Gelfand, 1998, p. 1088). Scoring at syllable level had already been introduced to other recently developed sentence tests, including the Beautifully Efficient Speech Test (BEST) (Schmitt, 2004) and an Afrikaans test for sentence recognition in noise (Theunissen, 2008).

Table 4.2. Comparison of differences between average scores of the individual lists, scored at word and syllable level
(List 15 was excluded because it was used as a practice list)

List number	Word average scores (%)	Syllable average scores (%)	Difference (%)
1	84.20	86.04	-1.84
2	85.50	86.72	-1.22
3	70.30	72.17	-1.87
4	78.20	79.25	-1.05
5	72.20	74.01	-1.81
6	85.80	86.43	-0.63
7	80.30	79.30	1.00
8	67.10	68.71	-1.61
9	78.70	80.93	-2.23
10	68.40	70.14	-1.74
11	82.30	81.87	0.43
12	59.70	59.71	-0.01
13	46.40	48.00	-1.60
14	83.60	83.91	-0.31
16	80.30	78.14	2.16
17	79.20	79.15	0.05
18	64.30	66.10	-1.80
19	74.60	75.99	-1.39
20	64.30	64.82	-0.52
21	68.20	64.74	**3.46**
22	73.40	73.36	0.04
23	74.50	74.26	0.24
24	35.40	37.38	-1.98
25	62.60	63.93	-1.33
26	66.50	67.04	-0.54
27	72.20	74.39	-2.19
Max difference			3.46

In response to the variation of average list scores, the individual sentences that differed from the average percentage score, as found during scoring at syllable level, were identified. The intensities of the sentences were adjusted by 0.7dB for every 10% they differed from the average scores as indicated in Table 4.3. The need for adjustments like these were to be expected, since intelligibility is not only determined by signal level but also by the contents of the utterance (sentence context, phonetic and prosodic factors), level variations within utterances, noise interference and word familiarity (Hällgren et al., 2006, p. 229; Versveld et al., 2000, p. 1673; Danhauer et al., 1984, p. 164; Kalikow, Stevens & Elliott, 1977, pp. 1337-1338).

During the pilot study, it was found that the intervals between stimulus sentences, which consisted of the time it took for the speaker to utter the sentence together with an additional three seconds, did not allow sufficient time for the participants to repeat the stimuli. This affected scores negatively in instances where participants were still in the process of giving a response while the next stimulus was already presented, resulting in their own talking masking the next stimulus sentence. This could not be rectified during testing by repeating the stimuli, since informal data collection by the author indicated that scores increased, when a stimulus was repeated (29.11% improvement on monosyllabic words during word identification testing). Repeating some stimuli therefore posed the risk of improving scores. It was consequently decided to increase the interval between sentences in subsequent rounds to the stimulus sentence presentation time plus an additional five seconds. This proved to be sufficient to allow enough time for repetition of the stimulus by the participants.

Table 4.3. Adjustments made to sentence intensities following the pilot round of testing

	Sentence no & adjustment made (dBFS)											
List no	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.9	-1.5	-0.7	-1.0	-0.3	0.5	-1.1	-1.4	-1.5	-0.2	-1.8	-0.8
2	-1.6	-1.5	-1.9	1.3	-0.9	-0.4	-0.5	-1.7	-1.2	0.2	-1.9	-1.5
3	-1.4	-1.5	1.6	-1.7	-0.9	2.3	1.7	4.6	-0.8	0.0	-1.1	-0.6
4	-1.9	-1.3	-0.9	-1.9	1.1	-1.1	0.5	0.6	0.1	-1.9	2.6	-0.9
5	-1.2	-0.3	0.0	-0.4	0.4	5.1	-1.8	1.7	-1.2	0.5	0.2	-1.9
6	-1.0	2.9	-1.3	-1.3	2.1	-0.3	-1.7	-1.5	0.3	-1.4	-1.6	-1.3
7	-1.9	-0.4	-0.7	0.4	1.9	-0.7	-1.9	0.0	-0.7	-1.6	-0.3	0.7
8	-0.2	-1.3	-0.5	0.7	1.6	2.9	-1.3	2.6	-0.9	0.5	0.2	-1.9
9	-0.6	-1.9	-1.4	-1.1	4.8	2.7	-0.8	0.1	-1.7	-0.9	-1.6	-1.6
10	0.9	-0.8	-0.7	1.6	3.5	0.8	-1.7	1.0	3.2	0.3	-1.4	-1.1
11	-1.6	-1.0	1.3	-1.4	-0.4	0.7	0.1	-1.1	-0.4	-1.0	-1.7	1.5
12	1.6	-0.8	-1.6	1.8	-0.9	2.7	2.9	-1.8	-0.1	0.8	2.7	1.7
13	0.2	1.0	1.7	-1.2	0.3	3.6	4.7	3.1	4.3	1.1	1.7	0.3
14	-1.3	2.8	0.1	-1.3	-1.9	-0.2	-1.9	-0.5	0.3	-1.1	-1.3	-1.8
16	-1.9	0.4	-0.1	-1.1	2.9	-1.9	-0.4	-0.5	-1.2	-0.9	-1.4	-0.9
17	-1.6	-1.4	-1.7	-1.1	0.9	2.3	-0.9	0.0	-0.1	1.9	-1.8	-1.2
18	0.7	4.3	0.3	-1.0	-0.6	4.6	-1.1	-1.5	-1.9	-1.4	1.2	-0.7
19	0.6	0.1	-0.5	-0.8	-1.6	0.3	1.4	-1.2	1.0	-1.3	0.2	-1.7
20	0.4	-1.9	1.9	2.0	1.9	-1.4	0.3	0.3	1.9	-0.3	2.8	-1.0
21	-1.9	-0.6	-1.4	3.2	4.2	-1.9	-0.6	5.1	1.4	-0.1	0.3	-0.4
22	-0.7	-1.1	0.3	-1.8	3.6	0.9	5.1	-0.7	-0.8	-0.2	0.4	0.6
23	2.9	-0.2	-1.2	1.2	-0.1	5.1	-0.7	-1.6	-0.2	-0.9	-1.7	-1.1
24	2.7	1.7	2.8	1.3	2.1	3.7	1.3	2.0	3.3	4.4	0.5	4.3
25	1.6	-1.9	-1.5	-0.4	4.7	2.4	1.7	-0.6	-1.9	3.0	-1.7	1.9
26	1.7	-0.2	0.8	2.7	0.0	2.6	-1.1	0.2	-1.5	1.8	-0.3	0.6
27	-1.5	1.2	0.6	1.5	2.0	0.3	0.2	0.1	-1.0	-1.6	-0.7	-0.7

As indicated earlier, from Round 1 onwards scoring was only done at syllable level. Figure 4.6 shows the average of participant scores for the different lists following Round 1.

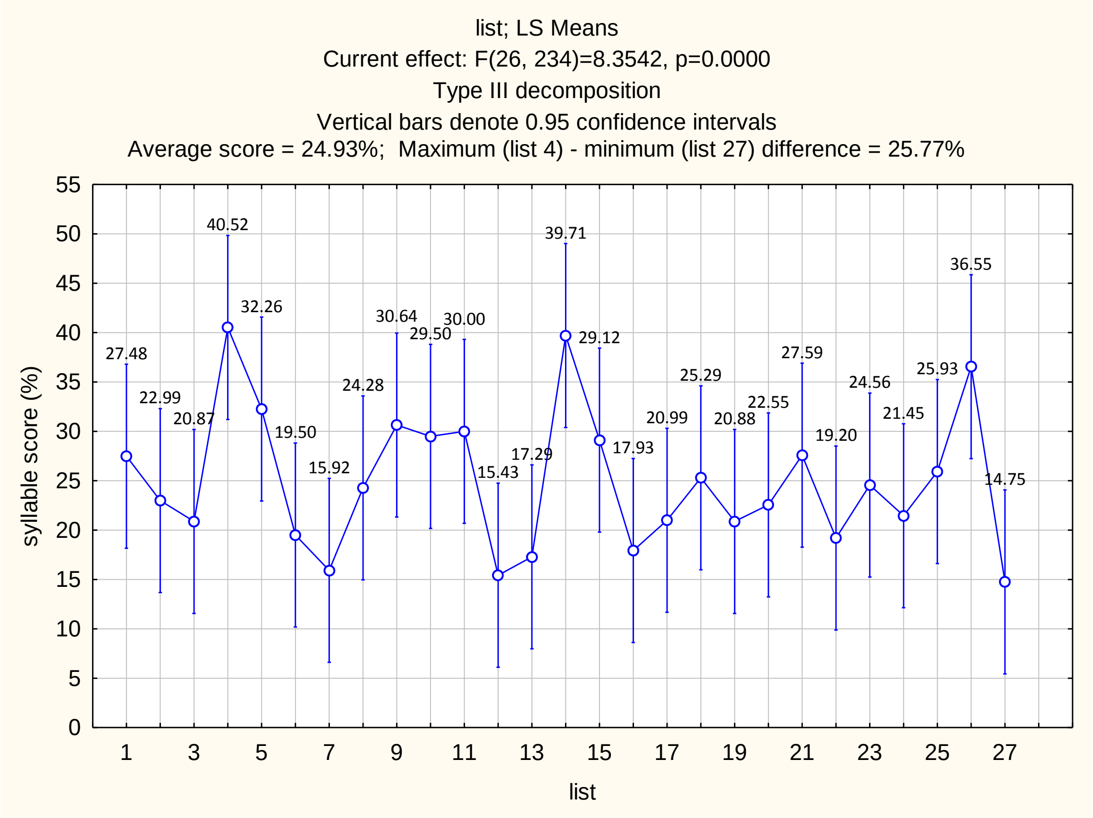


Fig. 4.6. The average scores for the sentence lists following Round 1 of testing

During Round 1 of testing, the average scores of the participants dropped from an average of 72.17% in the pilot study to an average of 24.93%. This was considered to either have been due to a weaker group of subjects in Round 1 than the preceding round, or due to changes made to the lists from the pilot round to Round 1. Although the subject selection criteria were carefully set out and controlled, there are variables, such as central auditory processing disorders (CAPD), which cannot be controlled on the basis of a screening procedure only. Certain behavioural symptoms, such as difficulty hearing in noise, can be considered risk factors for CAPD, but a full evaluation in order to make a diagnosis of CAPD should only be considered following measures of cognition, learning, speech and language (Bellis, 2003, p. 146), which clearly is a lengthy process. Although testing for CAPD was not done, the stratified sampling method was, however, seen as enhancing the representativeness of the population and the sampling allowed for even distribution of potential auditory disorders left undetected. A variable that was not controlled for, was familiarity with the test environment, as some of the participants of the pilot study were regular attendees at the audiology practice where testing was undertaken. Familiarity of the test environment might have positively influenced the test results.

An alternative explanation for the reduction in scores between the pilot round and Round 1, could be the changes in the sentence lists of the two rounds. Two major changes were made between the pilot round and Round 1. The first one was to increase the time intervals between individual sentences to allow more time for the participants to repeat the stimuli. The second change was the adjustment of intensities of the individual sentences by 0.7dB for every 10% they differed from the average.

Although the stimulus sentences were time-locked against a continuous background babble, this was disturbed when the time intervals between individual sentences were increased after the pilot round to allow more time for the participants to repeat the stimulus sentences. The stimulus sentences were therefore at this stage time-locked against a different part of the background babble. This could potentially have a large effect on scores, since the nature of multi-talker babble is that of fast varying energetic as well as informational masking levels. The increase in time between sentences would have changed the amount of energetic and informational masking present with every sentence, rendering the calculation of the amount of level adjustment following the pilot round inaccurate.

Despite the reduction in overall scores in Round 1, the variability of scores in this round was much smaller than that of the pilot round. The difference between the maximum and minimum scores of the pilot study dropped from 49.34% to a difference of 25.77% in Round 1, where list 4 scored the highest average of 40.52%, and list 27 scored the lowest average of 14.75%. Although the adjustments in intensity levels following the pilot round were not regarded as meaningful due to the change in time intervals between stimulus sentences, the results of Round 1 were still used as basis for a further round of testing due to the reduced variability in scores of this round. An adjustment of 0.5dB for every 10% deviation from the average score of 24.93% was made following Round 1, as depicted in Table 4.4.

Round 2 was presented at a SNR of +2dB, instead of +1dB as was used in the previous rounds, since the scores of Round 1 were so poor. However, the participants continued to score poorly with an average score of 44.43%. In response, testing in this round was terminated after only nine participants were tested, instead of the usual ten. The data of this round was still analysed to establish whether the adjustment in the sentence intensities had a positive effect on the homogeneity of scores.

Table 4.4. Adjustments made to sentence intensities following Round 1 of testing

	Sentence no & adjustment made (dBFS)											
List no	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.8	0.3	0	-2.7	1.7	-1.4	1.1	0.7	-0.3	-1.4	1.2	0.6
2	-0.5	0.8	1.2	0.1	1.9	-0.1	0.4	0.2	-0.5	-1.5	0.5	1.3
3	0.6	1.5	-0.1	1.2	2	0.8	-0.5	-1.3	1.3	-0.1	-1.7	1.1
4	0.4	-0.5	-2.2	1.4	0.7	-0.9	1.3	-2.6	-1.2	-0.7	-0.8	0.4
5	0.7	-1.4	0.1	0.3	1	-1.8	-1.2	-1.8	0.6	-1.2	0.1	1.1
6	-0.2	-0.4	0.6	0.2	1.5	1	1	1.5	0.7	-0.2	-0.3	1.3
7	0.4	0.7	0.7	1.5	-0.2	-0.6	1.5	0.9	1.5	1.2	0.6	0.8
8	0.6	1.3	-2.4	1.4	0.8	-0.3	1.1	-0.4	1	-2.1	-0.3	1.4
9	-0.5	1.4	-1.3	0.4	-1.3	-0.3	1	-0.6	0.3	-0.2	1.5	0.1
10	1.5	0.8	1.1	1.4	0.3	-2.4	0	-0.8	-0.4	-2.9	0.9	1.3
11	0.6	-0.7	-0.1	-0.8	1.3	0.9	0.3	1.3	0.8	-1.4	-1.5	0.1
12	0.3	1.2	0.2	0.2	1.4	1.4	-0.7	1.5	1.2	0.9	1.5	0.5
13	1	0.8	0.5	1.1	1.5	-0.9	-0.4	0.3	1.5	-0.2	1.3	1.4
14	-0.2	-1.9	-1.2	0.4	1.1	-1.3	-1.9	0.1	0.5	-0.9	0.1	-0.8
15	-0.1	1.5	0.2	1.5	1.3	1.3	-0.1	-0.2	-0.4	-2.8	-0.1	1.3
16	-0.1	1.2	0.8	0.9	0.6	0.7	0.2	0.7	0.3	0.5	0.8	-0.1
17	-0.2	1.2	0.9	1.5	1.5	-0.7	0.2	0.2	0.1	0.8	0.2	1.5
18	0	-0.5	-0.7	-2.8	0.5	-0.7	0.4	1.2	1.5	0.8	1.3	1.5
19	-0.6	0.8	-0.2	0.7	1.2	0.6	-0.2	1.1	1.3	0.6	0.3	-1.9
20	0.6	1.3	1.3	-1.1	0.5	0.8	1	1.1	1.5	-0.9	-2.5	1.5
21	1.1	0.5	0.9	0.5	-0.3	0.9	0.1	0.7	-1.3	-0.3	0.1	-0.9
22	1.4	0.4	0.1	1	-0.1	1.1	-0.1	0.3	-1.2	0.3	1.3	1.1
23	0.5	0.8	1.5	-0.8	1	-2.2	0	-2.4	0.8	-0.6	0.4	1
24	0.5	0.9	-0.6	-0.3	1.5	0.7	0.3	0.8	1.3	-0.4	-0.2	0.6
25	0.8	0.6	1.3	0.2	1.4	-1.6	-0.4	-0.7	1.5	-1.4	0	1.3
26	0	-0.9	1.3	0.6	-2	-1.2	-0.3	-2.9	0.1	-0.5	0.6	-1.4
27	0.4	1	-0.7	1.5	1	0.8	1	0.5	0.8	1.4	0.5	0

Figure 4.7 shows the results of Round 2. A slight improvement from 25.77% in Round 1 to 22.32% in Round 2 in the difference between the maximum and minimum scores can be seen, with list 21 obtaining the maximum score of 58.15% and list 3 obtaining the minimum score of 35.83%.

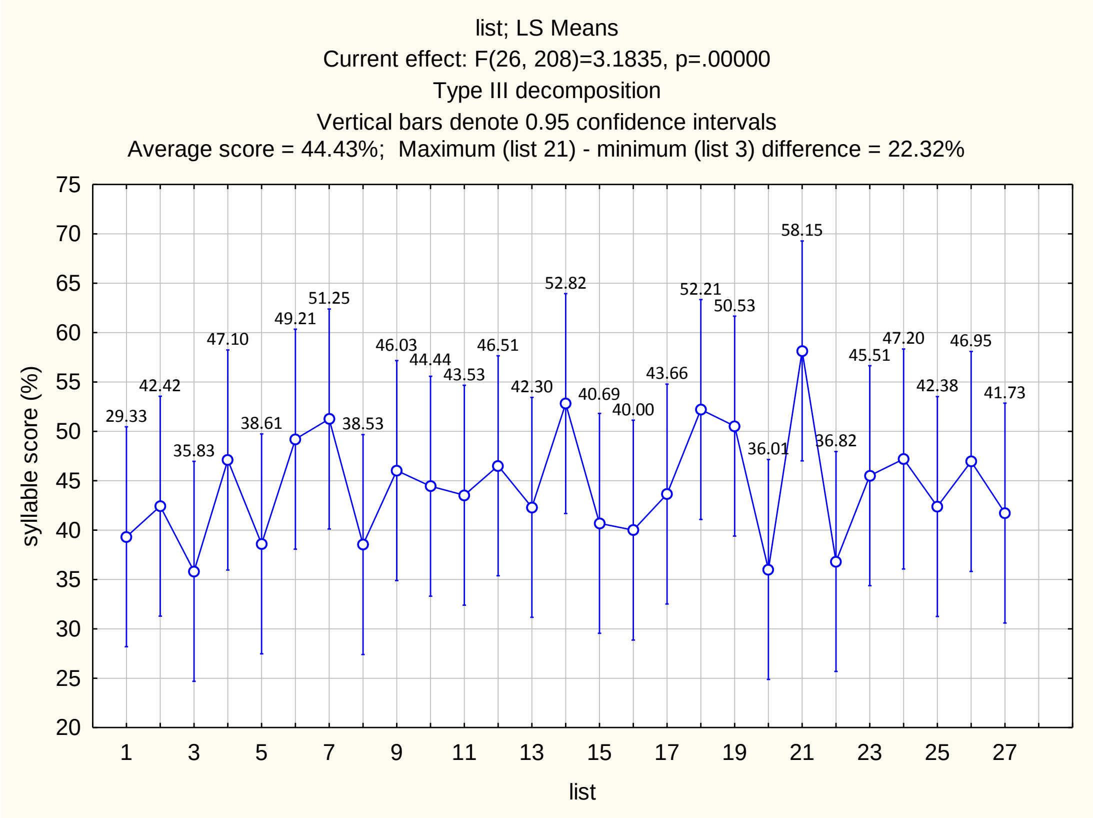


Fig. 4.7. Average scores for the sentence lists following Round 2 of testing

It was assumed at this stage that testing was not taking place at a very sensitive part of the PI function, since scores were below 65% (Dillon, 1983, p. 340), with an average score of 44.43%. Low scores like these meant that the task was too difficult and, if the task was felt to be too difficult, subjects might have been discouraged to perform at their best.

Despite the challenges during the first three rounds of testing, the following observations and conclusions were made:

1. Prior to increasing the time intervals between the sentence stimuli, test scores were affected negatively by the participants' responses sometimes overlapping with the next stimulus sentence. Because the SNR used to test

the sentences was so low, it was difficult for the participants to identify the start of the stimulus sentences. Where multi-talker babble is used as masking noise, one should consider incorporating a cue for the start of each stimulus sentence. These cues could be drawn as a parallel to visual cues, which are absent in an auditory-only test, such as the SAV-L. A warning tone prior to the presentation of the stimulus sentences might not be heard. A more favoured alternative solution is to not keep the babble continuous, but to stop it between stimulus sentences. The babble should then start shortly before and finish shortly after each stimulus sentence. This interruption of the masking noise is used in the HINT (Nilsson et al., 1994).

2. To ensure optimal performance, testing should take place at the most sensitive part of the PI function, described by Dillon (1983, p. 340) as the level, where 65% to 80% is achieved.

3. Unfamiliarity with test conditions seemed to be the most probably explanation of the large reduction in scores between the pilot round and Round 1, and should be considered when interpreting performance in the clinical set-up.

4.3.1.2. Inter-list reliability of Rounds 3 to 5

Prior to Round 3, the RMS levels of all the sentences were reset to be equal in intensity due to the discrepancies of previous rounds. The sentences were all reset to -12.00dBFS, the average RMS level of the individual sentences.

The average score for the lists of Round 3 was 70.32%, which showed a return to a sensitive area of the PI function. Figure 4.8 shows the average participant scores of the different sentence lists following Round 3.

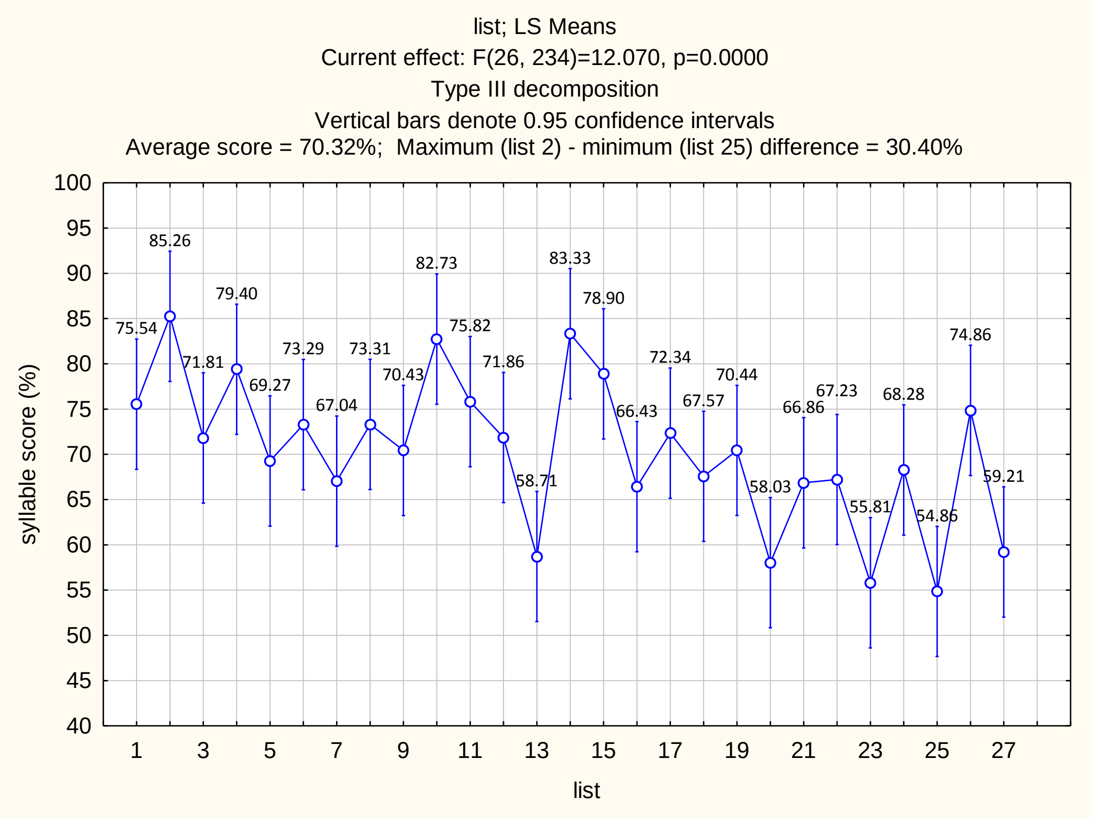


Fig. 4.8. Average scores for the sentence lists following Round 3 of testing

The difference in performance between maximum and minimum scores in Round 3 was 30.40% with list 2 obtaining the maximum score of 85.26% and list 25 obtaining the minimum score of 54.86%.

After comparison of participant scores obtained during Round 3, every list (as opposed to every individual sentence) was adjusted by 0.5dB for every 10% it differed from the average of 70.32%. Table 4.5 shows the adjustments made.

Table 4.5. Adjustments made to sentence list intensities following Round 3 of testing

List no	dBFS adaptation
1	-0.3
2	-0.8
3	-0.1
4	-0.5
5	0.0
6	-0.2
7	0.1
8	-0.2
9	0.0
10	-0.6
11	-0.3
12	-0.1
13	0.6
14	-0.7
15	-0.4
16	0.2
17	-0.1
18	0.1
19	0.0
20	0.6
21	0.2
22	0.1
23	0.7
24	0.1
25	0.8
26	-0.2
27	0.5

During Round 4, the average scores of the sentence lists proved to be much more homogenous. The difference between the maximum and minimum scores was 20.84%, as opposed to a difference of 30.40% in Round 3. The maximum score of 79.05% occurred in list 21 and the minimum score of 58.21% occurred in list 25. Figure 4.9 shows the average scores of all the lists. From the improvement in scores with regards to their variability, it was concluded that, although intensity adjustment was not the only predictor of subsequent performance, it did play an important role in making performance scores more homogenous. The decision to adjust the intensities of the sentence lists as a whole as opposed to adjusting individual sentences were therefore seen as acceptable in view of the improvement in variability of scores.

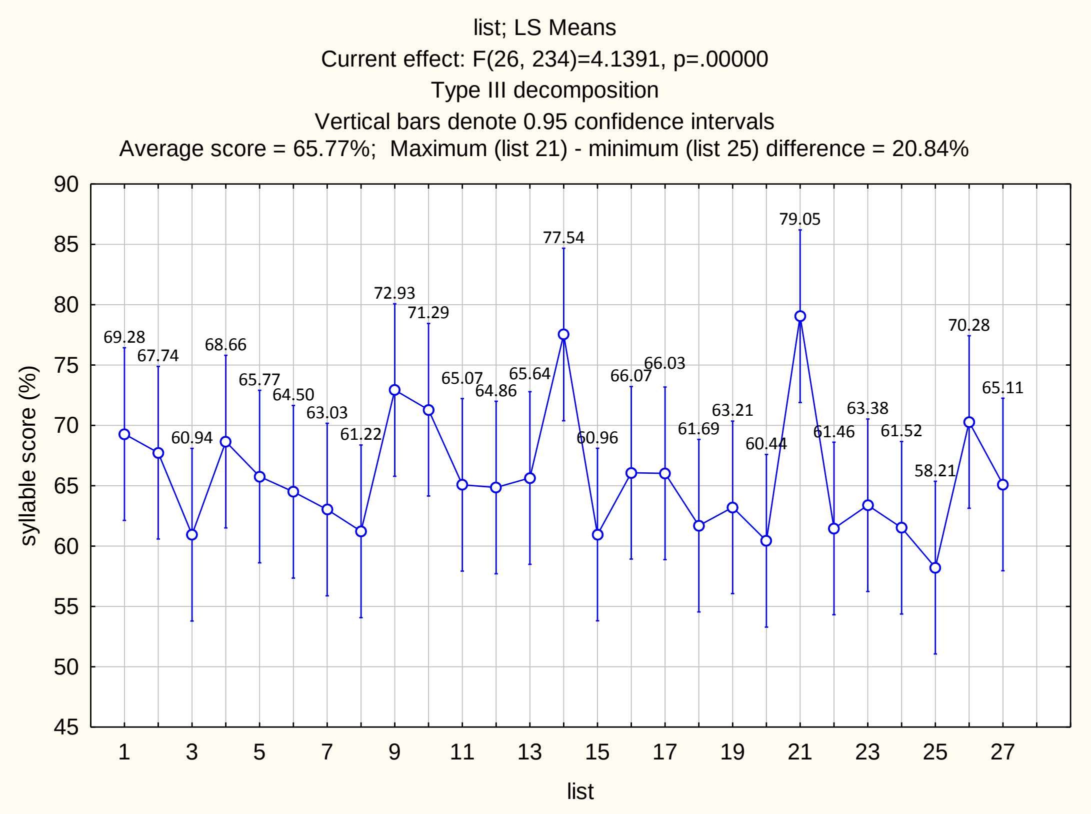


Fig. 4.9. Average scores for the sentence lists following Round 4 of testing

After a comparison of the average intensities of the lists from Round 4, the lists were adjusted with 0.3dB for every 10% they differed from their average of 65.77%, as illustrated in Table 4.6.

Table 4.6. Adjustments made to sentence list intensities following Round 4 of testing

List no	dBFS adaptation	Exceptions
1	-0.1	
2	-0.1	
3	0.2	
4	-0.1	Sentence 7 stayed unchanged
5	0.0	
6	0.0	
7	-0.1	Sentence 4 stayed unchanged
8	0.1	
9	-0.2	
10	-0.2	
11	0.0	
12	0.0	
13	0.0	
14	-0.3	
15	0.2	
16	0.0	
17	0.0	
18	0.1	Sentence 10 stayed unchanged
19	0.1	Sentences 9 and 10 stayed unchanged
20	0.2	
21	-0.4	
22	0.1	
23	0.1	Sentence 8 stayed unchanged
24	0.1	
25	0.2	Sentence 9 stayed unchanged
26	-0.1	
27	0.0	
28	0.0	

The average sentence scores of Round 5 are shown in Figure 4.10. The average score of this round dropped from 65.77% in Round 4 to 59.69%. This reduction was assumed to be insignificant, since test conditions and procedures were maintained, with the only difference being the participants. The difference between the maximum and minimum scores did, however, have a positive reduction from 20.84% in Round 4 to 13.87% in Round 5, with list 21 obtaining the maximum score of 66.72% and list 20 scoring the minimum score of 52.85%. Adjustment of sentence intensities was terminated at this point, because the adjustments in intensities were becoming marginal.

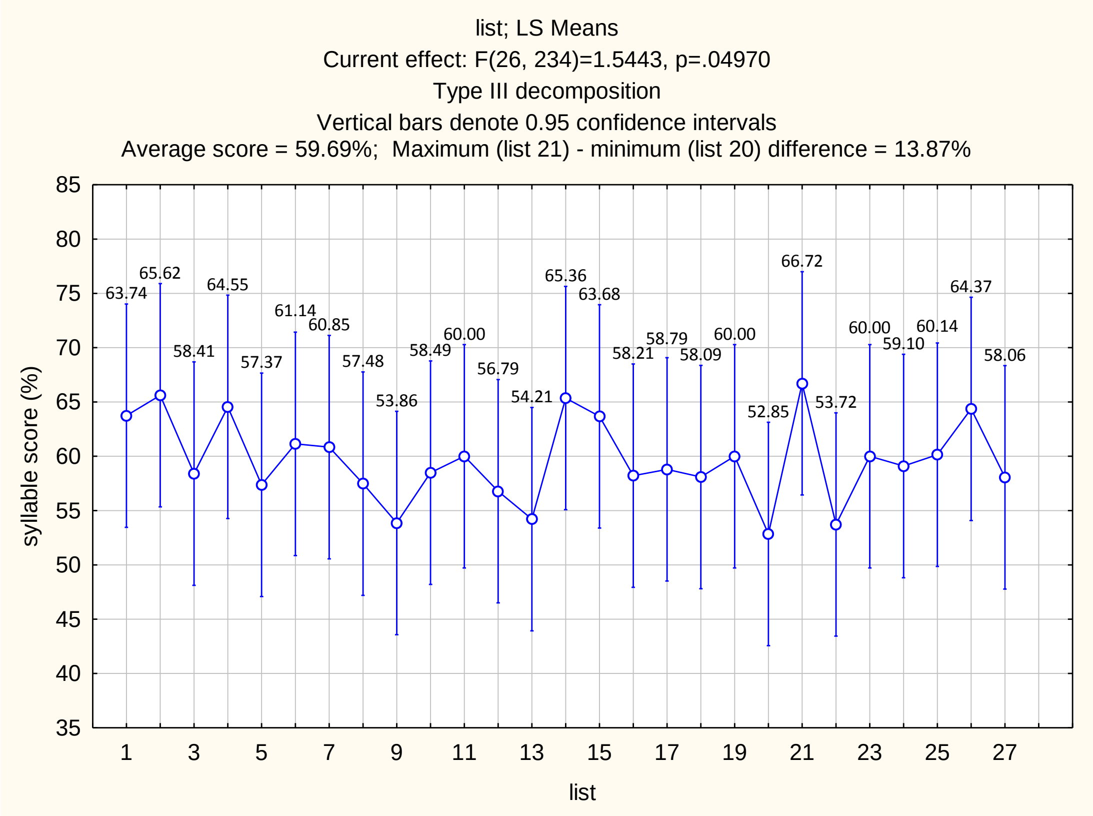


Fig. 4.10. Average scores for the sentence lists following Round 5 of testing

It was noticeable that the level adjustments had a positive effect on the performance with regards to the similarity of scores between lists. This was seen through the reduction of the variability between sentence lists from 30.4% in Round 3 to 13.87% in Round 5. However, other variables may also have had an influence on the performance, since the changes in intensity made, were not directly related to the subsequent performance. In other words, one could not predict the scores of the subsequent round based on the amount of intensity adjustments made following the previous round. This was seen when the scores of Round 3 were compared with those of Round 4 (Table 4.7). Although the intensity of the lists was adjusted with 0.5dB for every 10% the scored in Round 4 varied from a 2.5 dB to a -5dB adjustment for every 10% they were adjusted (Figure 4.11). This implied that there were uncontrolled variables other than stimulus intensity that affected the score of any particular sentence list, and that the level of increase or decrease in intensity of a list could not be used to predict subsequent scores. A discussion of uncontrolled variables is made in 5.2.1.

Table 4.7. Comparison between the intended and actual adjustment in intensity of the sentence lists

List number	Round 3 list average (%)	Round 4 list average (%)	Intensity adjustment (dB) following Round 3 (0.5dB for every 10% difference)	Difference in score between Rounds 3 and 4 (%)	Equivalent to dB per 10%
1	75.54	69.28	-0.3	-6.26	0.44
2	85.26	67.74	-0.8	-17.52	0.44
3	71.81	60.94	-0.1	-10.87	0.08
4	79.40	68.66	-0.5	-10.75	0.44
5	69.27	65.77	0.0	-3.50	-0.10
6	73.29	64.50	-0.2	-8.79	0.19
7	67.04	63.03	0.1	-4.01	-0.37
8	73.31	61.22	-0.2	-12.09	0.14
9	70.43	72.93	0.0	2.50	-0.09
10	82.73	71.29	-0.6	-11.44	0.56
11	75.82	65.07	-0.3	-10.75	0.27
12	71.86	64.86	-0.1	-7.00	0.13
13	58.71	65.64	0.6	6.93	0.81
14	83.33	77.54	-0.7	-5.80	1.15
15	78.90	60.96	-0.4	-17.94	0.25
16	66.43	66.07	0.2	-0.36	-5.00
17	72.34	66.03	-0.1	-6.31	0.19
18	67.57	61.69	0.1	-5.88	-0.21
19	70.44	63.21	0.0	-7.23	0.03
20	58.03	60.44	0.6	2.41	2.48
21	66.86	79.05	0.2	12.19	0.13
22	67.23	61.46	0.1	-5.77	-0.24
23	55.81	63.38	0.7	7.57	0.94
24	68.28	61.52	0.1	-6.76	-0.13
25	54.86	58.21	0.8	3.36	2.26
26	74.86	70.28	-0.2	-4.58	0.53
27	59.21	65.11	0.5	5.90	0.91

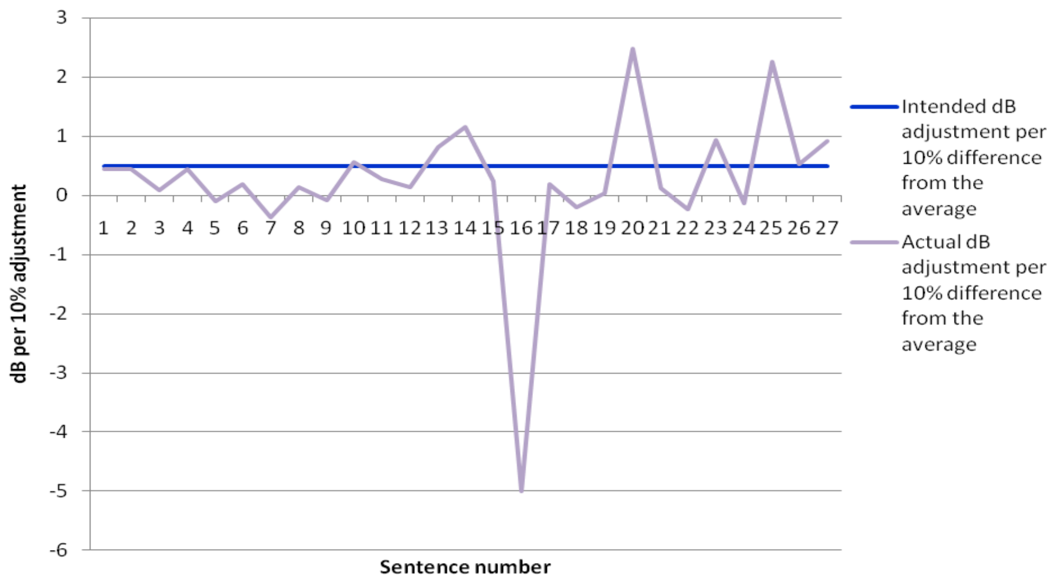


Fig. 4.11. Comparison of the intended level adjustment following Round 3 and the actual result in level adjustment as measured in Round 4

The following conclusions were drawn from Rounds 3 to 5:

1. Adjusting the intensity of sentence lists in a systematic way had a positive effect on the equalisation of the lists, and can be used as an effective tool in the development of speech perception tests.

2. Variables other than level adjustment had an influence on test scores. These variables are further discussed in Chapter 5.

3. Where the sentences in a particular list are not of equal difficulty, e.g. when sentences differ in length, it is effective to adjust the sentence list as a whole as opposed to the individual sentences.

4. The variability amongst list scores for the developed sentence lists was improved to 13.87%. This implies that a participant needs to show at least this amount of improvement in performance with this version of the SAV-L

in order for it to be recognised as an improved score, and therefore a positive outcome during rehabilitation.

4.3.1.3. Summary of the pilot round and Rounds 1 to 5:

As a summary of testing undertaken, the adjustments and outcomes made from the pilot round to Round 5 are presented in Table 4.8. It shows the SNRs at which testing was done throughout the different rounds of testing, as well as average scores obtained, differences between maximum and minimum scores and adjustments made following each round of testing.

Table 4.8. Summary of the data from the pilot round to Round 5

Round name	SNR (dB)	Average score (%)	Max-min difference (%)	dB per 10% adjustment following the round
Pilot – syllable score	+1	71.47	49.34	0.7
Pilot – word score		72.17	50.40	
1	+1	24.93	25.77	0.5
2	+2	44.43	22.32	N/A - levels reset to equal RMS levels
3	-3	70.32	30.40	0.5
4	-3	65.77	20.84	0.3
5	-3	59.69	13.87	N/A - testing completed

Table 4.9 presents the standard deviations of the mean scores of Rounds 3 to 5. Comparison of these rounds revealed that the standard deviation of Round 5 was the smallest, making this version of the sentence lists the most homogenous with regards to performance. The test version used in Round 5 was therefore accepted as the final version of the test.

Table 4.9. Breakdown of the standard deviations of Rounds 3, 4 and 5

Round	Means Score Mean (%)	Means Score Standard Deviation
3	70.32	8.15
4	65.77	5.08
5	59.69	3.80

4.3.2. <u>Test-retest reliability</u>

The test-retest reliability of performance was established by retesting ten participants in Round 6 under identical conditions to the conditions used before. The results from this round of testing are shown in Figure 4.12. An intra-class correlation (ICC) agreement of 0.859 was found. The test results showed an improvement when the participants were retested, which suggested the existence of some kind of bias. This bias seemed to be consistently present with an ICC consistency of 0.907.

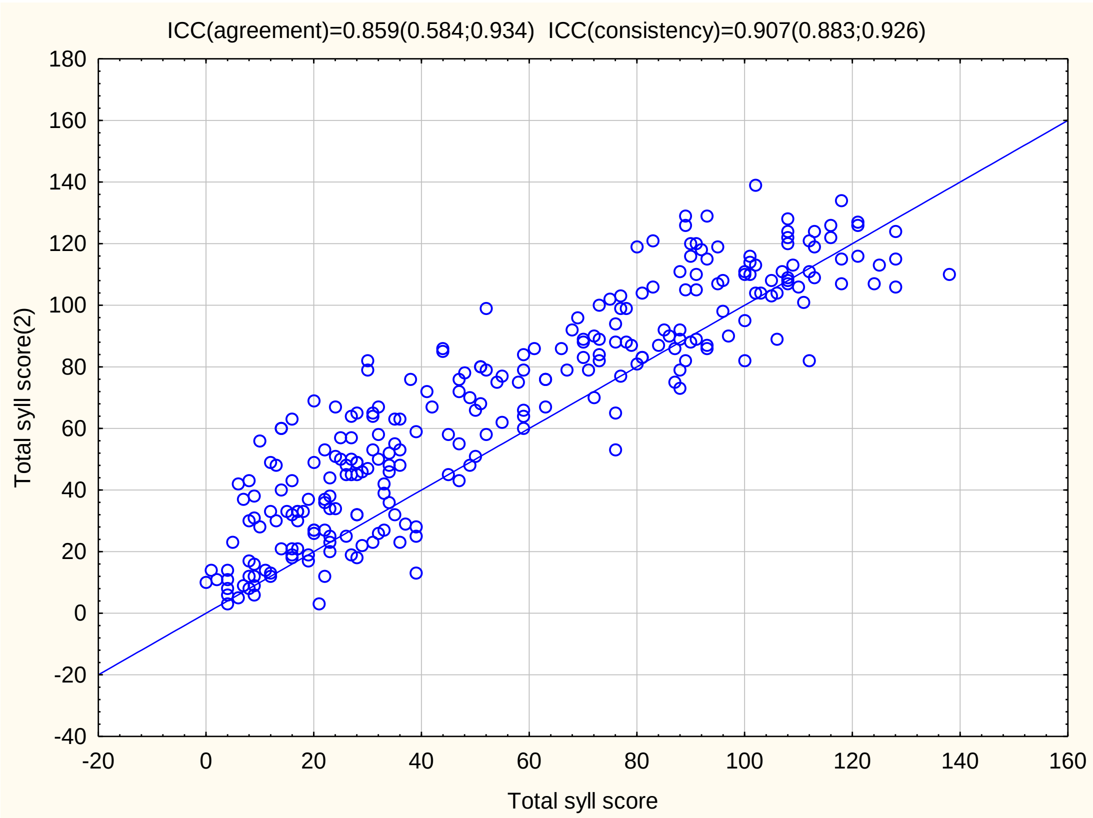


Fig. 4.12. Test-retest reliability

Given the variability in scores obtained during testing, the intra-class correlation agreement of 0.859 was better than anticipated. It was assumed that the test-retest reliability of the participants, who were completely unfamiliar with audiological test procedures during their first round of testing, had improved during their second round of testing, because they were more familiar with the test surroundings, i.e a conformational bias affected the results (Williams & Ward, 2003, Carr, 1980, p. 223). This assumption seemed viable, because this could also, at least partly, have explained the large improvement in scores from the pilot round, where some participants were familiar with the test environment, to Round 1, where the participants were completely unfamiliar with the test environment. A third round of testing after a significant interval would prove valuable, in order to observe trends over a longer period of time. This third round of testing would establish if this assumption is accurate.

The following conclusions were made from this round of testing:

1. Participant performance might have been negatively influenced by an unfamiliar clinical environment and improved as they became more familiar with the test environment.

2. The SAV-L showed an improvement in scores over a short period of time, but longer term measures of test-retest reliability still need to be done. Test-retest reliability was, however, acceptable.

4.3.3. Inter-rater reliability

Throughout the first four rounds of testing, a second scorer randomly co-scored the performances in order to establish inter-rater reliability. The comparison of the scores of the rounds, where two raters were used, showed the maximum inter-rater difference on syllable level during the pilot study to be 4.38%. This improved to a difference of 1.49% in subsequent rounds after underlining the first letter of each new syllable on the test forms as identification of the scoring units. Inter-rater reliability in subsequent rounds, whilst scoring at syllable level, had an ICC agreement of 0.999 with an ICC consistency of 0.999 as well (Figure 4.13). A very high inter-rater reliability was therefore found to be present.

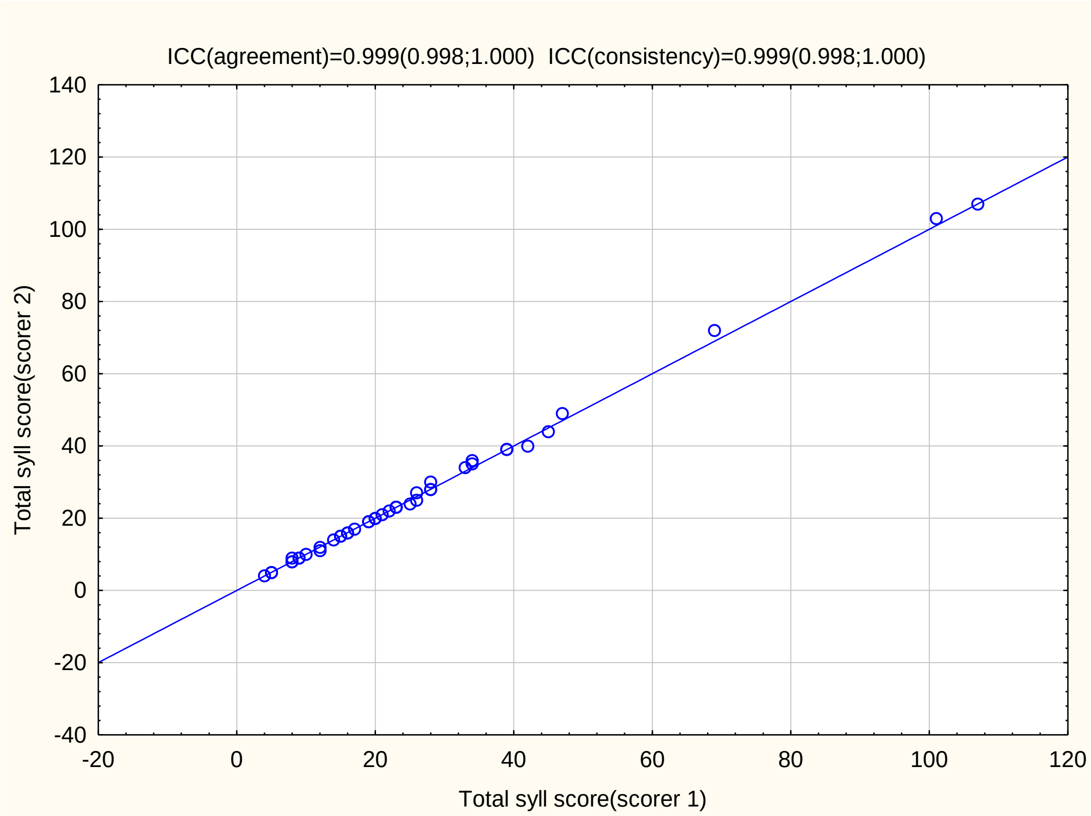


Fig. 4.13. Inter-rater reliability, after scoring units were identified on the test forms

In view of the high inter-rater reliability, the following conclusions were drawn:

1. Scoring at syllable level was acceptable with regards to inter-rater reliability.

2. Clear test forms were achieved by underlining the scoring units (syllables in this instance). This improved inter-rater reliability and therefore overall reliability of the test.

3. The SAV-L has a very high inter-rater reliability. Variability in scores due to inter-rater differences can therefore largely be excluded.

4.3.4. Confirmation of intended test level

Table 4.10 shows the distribution of scores obtained from testing at the intended test level of 50dBHL with an SNR of +5dB, which were then increased to +6dB and lastly to +7dB.

Table 4.10. Scores following Round 7 (presentation of the sentence lists at the intended test level)

List no (in order they were presented in)	SNR (dB)	Subj 1 (%)	Subj 2 (%)	Subj 3 (%)	Subj 4 (%)	Subj 5 (%)	Average of different SNRs (%)
15	+5	100.00	96.32	100.00	98.53	100.00	
27	+5	100.00	96.40	100.00	99.28	100.00	
4	+5	100.00	98.51	94.78	100.00	100.00	
23	+5	100.00	97.79	100.00	100.00	100.00	
8	+5	100.00	97.10	97.83	100.00	100.00	
19	+5	100.00	100.00	100.00	100.00	100.00	
12	+5	100.00	92.14	99.29	100.00	100.00	
16	+5	100.00	97.86	100.00	99.29	100.00	
11	+5	100.00	98.51	97.01	99.25	100.00	99.11
20	+6		100.00	98.54	100.00		
7	+6		100.00	99.30	100.00		
24	+6		98.62	98.62	98.62		
3	+6		100.00	98.55	100.00		
1	+6		99.28	97.12	100.00		
26	+6		100.00	99.30	100.00		
5	+6		98.54	100.00	99.27		
22	+6		100.00	100.00	100.00		
9	+6		98.57	100.00	99.29		99.39
18	+7		100.00	100.00	100.00		
13	+7		98.57	100.00	100.00		
14	+7		99.28	98.55	97.83		
17	+7		100.00	100.00	98.58		
10	+7		97.12	97.84	100.00		
21	+7		91.97	99.27	100.00		
6	+7		97.14	100.00	100.00		
25	+7		100.00	100.00	100.00		
2	+7		100.00	100.00	100.00		99.12
Minimum score		100.00	91.97	94.78	97.83	100.00	
Maximum score		100.00	100.00	100.00	100.00	100.00	
Average overall score		100.00	98.29	99.11	99.63	100.00	99.40

Two out of the five participants (participants 1 and 5) scored 100%. The other three participants scored 91.97% or above. The average overall score was 99.40%. At a SNR of +5dB the average score was 99.11%. At a SNR of +6dB, the average score was 99.39% and at a SNR of +7dB it was 99.12%.

A total of 44 out of 324 sentences contained wrong responses made by the participants. Forty of the sentences consisted of 10 words or more. The remaining 4 sentences contained three, four, seven and eight words per sentence respectively. A total of 83.02% of the mistakes were made by 2 participants, while a third participant made the remaining 16.98% of mistakes.

The two participants that made most of the mistakes were a mother and a daughter. These two participants showed poor concentration, and their minds were reportedly preoccupied with other responsibilities. These influences were suspected to have affected their scores negatively. This supports the assumption that hearing is a rather passive function, which needs to be supplemented by the more active cognitive processes of listening, comprehension and communication for full language understanding (Pichora-Fuller & Singh, 2006, in Rudner, Foo, Rönnberg & Lunner, 2009, p. 405).

The observation that no improvement of scores was observed with increase of the SNRs, supported the conclusion that the linguistic load or other factors, such as concentration and auditory memory span, influenced scores, rather than the interference of the three-talker babble at SNRs of +5dB or higher. Although factors like these could easily influence performance in the real world as well as in the clinical set-up, the sample for this round of testing was too small to make conclusions about the extent of influence these factors might have had on testing.

The following conclusions were drawn from this round of testing:

1. The SAV-L's linguistic load or demands on auditory memory seemed to affect scores, even when the task was made easier by increasing the SNR. Test results therefore appeared to be influenced by the participants' cognitive skills. The sample used was small and this observation needs to be supported by more data.

2. A SNR of +5dB was seen as a sufficient fixed SNR to be tested at, since any higher SNR did not improve scores.

3. Persons with normal hearing could be expected to score 91.97% or better at a favourable SNR of +5dB on this test. This score needs to be confirmed with data from a larger sample.

4.4. SUMMARY

Chapter 4 presented the results and outcomes of the present study, of which the aim was to develop a sentence test in three-talker babble. The results and outcomes gained from the design phase, i.e. the compilation and recording of the sentence lists, were presented. This was followed by the presentation of the developmental phase of the study, i.e. the field studies with the test. These studies were aimed at improving the inter-list reliability as well as establishing the test-retest and inter-rater reliability. Lastly they served to confirm the intended clinical test level.

The linguistic load of the sentences used for the test, together with the choice of masking noise, made the SAV-L a valid test of speech-in-noise performance for the target population of adults with developed oral language. The choice of three-talker babble was felt to provide sufficient energetic masking whilst maintaining informational masking.

The choice of differing sentence lengths for the test material, increased the test's construct, face and ecological validity but reduced the test's reliability, since it is known that differing sentence difficulty (which would result from differing sentence lengths) can reduce reliability (Dillon, 1983, p. 336). Nevertheless, inter-list reliability was reduced to a variation of 13.87% in scores between sentence lists. To establish whether this score makes the test sensitive enough, further testing needs to be done.

Although test-retest reliability was acceptable, it did seem to be affected by the familiarity of the participants with the test conditions. Additional repeated assessment is required to monitor test-retest reliability over a longer period of time. Inter-rater reliability was found to be very good, which is important in instances where more than one clinician is involved.

The intended test level was established as 50dBHL with a +5dBSNR, therefore the masking level being presented at 45dBHL. In this set-up a person with normal hearing is expected to score 91.97% or more from the limited data obtained so far. A larger group of participants needs to be assessed to evaluate the reliability of this score and to establish whether this score makes the test sufficiently sensitive.

CHAPTER 5

FURTHER CONSIDERATIONS, RECOMMENDATIONS AND CONCLUSION

5.1. INTRODUCTION

This chapter discusses aspects of the study identified for further consideration. It also discusses the clinical use of the developed test. Recommendations for further research and a conclusion complete the chapter.

5.2. ASPECTS OF THE SAV-L FOR FURTHER CONSIDERATION

The SAV-L provides needed material for the assessment of candidacy as well as monitoring of benefit and progress with auditory devices during audiological rehabilitation in the South African context. However, during the development of the SAV-L, two aspects were identified as needing further consideration. The first aspect entailed variables that could not be controlled, while the second aspect entailed the analysis of individual sentences that proved to be problematic. These two areas will consequently be discussed.

5.2.1. Uncontrolled variables

Because participant scores could not precisely be predicted based on level adjustments made to previous rounds, it was assumed that additional uncontrolled variables had an effect on performance. These variables were identified as falling into the categories of internal, cognitive and extra-linguistic variables. They need to be kept in mind when conclusions about performance with speech-in-noise tasks are drawn during evaluation for auditory device candidacy and will consequently be discussed.

5.2.1.1. Internal variables

Internal variables influence cognitive capacity, and therefore performance, in all domains of life, whether it is in the workplace or during audiological evaluation conditions. These variables include emotion, motivation, effort, conflict, fatigue, stress, concentration and energy, and can degrade performance quickly if present (Gaillard, 2008, p. 59). They cannot be controlled or may only be controlled to a very small degree, e.g. making sure that test persons only attend in the mornings instead of after work/school, in order to minimise activities that could have caused mental fatigue. During the development of the SAV-L, these variables were attempted to be maintained by testing participants in one session rather than over two, so that the conditions were more or less kept constant for the individual. Frequent breaks were also allowed as required by the individual, in order to reduce the impact of fatigue on performance.

The coincidental observation of lapses in variables, such as concentration, served as a reminder that the challenge to the SAV-L, as much as to any other test, is whether the performance on the given day is a true reflection of the participant's abilities. It is therefore essential that this test be used as part of a battery of tests that include additional subjective and objective audiological tests, a full case history and discussion or evaluation of subjective disability. If a test result does not clearly show an improvement or reduction in abilities, decisions regarding (further) management can then be taken based on these results in conjunction with other results or discussions with the participant and significant others. This approach is in line with the cross-check approach (Stach, 2010, p. 198) used by audiologists in practice. Saunders, Chisolm and Abrams (2005, p. 162) agree that speech perception testing should not be used in isolation to predict hearing aid benefit, although a valid test can indicate improvement in performance.

5.2.1.2. Cognitive variables

Speech-in-noise tests as part of the test battery for assessment of auditory device candidacy, are generally only used to consider the peripheral aspect that impairs performance. However, additional factors such as higher auditory processing abilities can also affect performance, but these processes are not addressed by hearing aids and cochlear implants. If these additional factors are completely disregarded, unrealistic expectations about the success of an auditory device may be formed and wrong management decisions may also be made.

Central auditory processing disorders (CAPD) can heavily affect test results during speech-in-noise testing. This is due to the increase in stimulus uncertainty in the presence of CAPD, which results in reduced spoken language understanding (Chermak & Musiek, 1997, p. 237).

Additionally, higher-level neurocognitive processes such as linguistic knowledge, can also affect performance (Bellis, 2003, p. 53). In easy listening conditions, bottom-up peripheral auditory processes are used (Kramer, et al., 2009, p. 507), but in difficult listening conditions, such as speech-in-noise testing, the need for top-down processes increases (Kramer et al., 2009, p. 507).

The participants' knowledge of syntax (structure of a sentence; Owens, 2001, p. 19), morphology (internal organisation of words; Owens, 2001, p. 21), phonology (rules of the structure, distribution and sequencing of speech sounds; Owens, 2001, p. 22) as well as semantics (rules governing the meaning of words; Owens, 2001, p. 23) was assumed to play an important role in performance during testing with the SAV-L, since the test material contained all these aspects.

Although speech-in-noise tests are broadly accepted as being the most appropriate measures to assess an individual's ability to understand speech in noisy daily life situations (Kramer, Kapteyn, Festen & Toby, 1996, p. 286), the influence of these top-down processes must be considered. If implicit cognitive processes are insufficient to decode the meaning conveyed by the speech stream, performance will be poor (Rönnberg, 2003, p. S71). The assessment of separate contributions of cognitive (top-down) processes in speech comprehension is actually indicated as part of the audiological test battery (Kramer et al., 2009, p. 508). When the SAV-L is therefore used as an assessment tool, additional skills required, such as higher processing abilities, should be kept in mind in order to draw the correct conclusions.

The influence of linguistic knowledge can be controlled by choosing the most appropriate test battery for the individual. A second way of controlling linguistic variables in speech perception tests, would be by controlling the linguistic load of the test. Simple vocabulary and syntactic structures for children and even for adults are often used. An example of this would be the BKB sentence tests (Bench et al., 1979) that use vocabulary and grammar from hearing-impaired children and are approximately at a grade 1 reading level (BKB-SIN manual, n.d., p. 17). The HINT sentence are also based on the sentences of the BKB sentence test (Nilsson et al., 1994, p. 1086). However, test material that is too easy, could lead to over-performance (Madell, 2010, slide 21). The SAV-L with its more advanced language structures is therefore attractive to use for assessing adults in order to prevent over-performance.

A third way of controlling the linguistic variables of a speech perception test might be to consider a dynamic assessment. Dynamic assessment is an approach known to speech and language therapists and uses a test-teach-retest model (Gillam, Peña & Miller, 1999, p. 36). This form of assessment can be applied to sentence perception tests by having sentence lists with a specific topic. With the test-teach-retest model of dynamic assessment, a baseline measure would be obtained in the first instance. This is done by presenting a list of unfamiliar sentences of a certain topic to a participant. This would be followed by a teaching phase, consisting of a discussion of the topic to introduce its vocabulary. This phase could be visually supported by a picture relating to the topic. Directly after the training phase, another list is presented and the participant's scores are compared to the original scores. If a significant improvement is present, linguistic variables could be identified as having had a significant influence on the baseline scores. This kind of assessment is an especially important consideration in South Africa with its diverse socio-economic and linguistic context, since test bias makes many standardised tests nearly unusable for differential diagnosis, where culture or experience is an issue (Gillam et al., 1999, p. 33). Currently such a test is not available in audiology, but might warrant further consideration.

5.2.1.3. Extra-linguistic variables

Extra-linguistic components can affect the understanding of language and therefore also play a role in any speech perception test. Extra-linguistic elements include non-linguistic cues, such as gestures, body posture, facial expressions, eye contact, head and body movements and proxemics, i.e. physical distance (Owens, 2001, p. 13). These cues were excluded from the SAV-L by making it an auditory-only test. Although these variables were therefore neutral throughout the test, lack of them could affect speakers to a greater or lesser degree. The absence of non-linguistic cues could for instance put a person with a severe hearing impairment, who relies heavily on these cues, at a greater disadvantage than a person with normal hearing or a milder loss. The ability of a test participant to make good use of non-linguistic cues might heavily affect management decisions and should not be excluded as a consideration.

The absence of this variable can once again be controlled through the choice of a test battery through consideration of the individual's history and abilities.

The following conclusions were drawn from the consideration of uncontrolled internal, cognitive and extra-linguistic variables:

1. The SAV-L should be used as part of a battery of tests, in order to get a complete picture of a test participant's abilities.
2. Higher cognitive abilities may affect a participant's performance in a speech-in-noise task and their impact should be considered when an intervention plan is compiled.
3. Extra-linguistic cues, or more precisely, lack thereof, can affect scores in speech perception tests. Although not directly assessed during an auditory-only speech perception tests, the ability of a participant to make use of these aspects need to be considered in the development of an intervention plan.
4. The linguistic load of a speech-in-noise test will dictate the population it is suitable for. A one-test-fits-all approach is not appropriate, and linguistic abilities should be considered for the particular individual.
5. Dynamic assessment is an unknown technique in audiology, but may be a consideration to address the wide cultural and linguistic differences in a country such as South Africa.

5.2.2. Problematic sentences

Low scores obtained in Round 5, where an SNR of -3dB was present, were compared with the low scores obtained in Round 7, where an SNR of +5dB or more were present. Low scores were defined in Round 5 as having received an average score of less than 65%, i.e. falling below the most sensitive area of the PI function (Dillon, 1983, p. 340). In Round 7, low scores were defined as any individual score of less than 100%, since a 100% score was expected during this round of testing.

During Round 5 of testing, 165 out of the total of 324 sentences received average scores below 65%, whereas 44 out of 324 sentences in Round 7 scored below 100%. Only 8 sentences scored poorly in both of these rounds.

Table 5.1 compares the scores of poor-performing sentences from Rounds 5 and 7. It was assumed that the 157 sentences that scored poorly at low SNRs but not at high SNRs, were difficult due to the masking interference of the three-talker babble. The 36 sentences that scored poorly only at higher SNRs but not at lower SNRs were unexpected, and judged to be due to lapses in concentration. The 8 sentences that scored poorly at higher as well as low SNRs were suspected to have linguistic components that might have caused the poor performance.

Table 5.1. Explanation of poorly scored sentences in Rounds 5 and 7

Round 5 sentences score low?	Round 7 sentences score low?	Explanation	Total sentences with this pattern
No	Yes	Masking interference	157
Yes	No	Concentration	36
No	No	Linguistic interference	8

The sentences that scored low in both rounds were compared to look for a cause of under-performance. Factors like syntax, word predictability and topics used, were all considered as possible causes of under-performance of these eight sentences. However, it was felt that these linguistic attributes were no more pronounced in these sentences than in the other sentences of the SAV-L. The only noticeable trend was that, with the exception of one of these sentences (which contained 10 words), the sentences had 12 or more words per sentence. Sentence length, therefore, seemed to be the only common factor in these sentences, which were concluded to be the cause of under-performance.

A more detailed investigation of problematic sentences still remains to be done because testing at high SNRs was only done with a small group of five participants. A poor score of as little as one individual would have been classified as a sentence with a poor score.

The following conclusions were drawn from analysing the problematic sentences:

1. Sentences containing 12 or more words might affect test scores, presumably due to the demands they place on auditory memory.
2. No sentences in the developed sentence lists stood out as being more difficult than others due to being linguistically more complex.

5.3. CLINICAL USE OF THE TEST

The SAV-L was designed to fill a gap in the battery of speech audiometric tests currently available in South Africa. It can be used to fulfil both of the two broad classes of applications of speech perception tests (Dillon & Ching, 1995, p. 323). The first application is to test a subject under two or more conditions, e.g. monitoring progress over time with the use of a particular auditory device, or comparing different auditory devices. The second application is to compare how the participant scores in comparison to the ‘usual’, i.e. the normative score for the population, for instance, when auditory device candidacy is to be established.

The strongest asset of the SAV-L is believed to be the three-talker babble used as masking noise, which can most closely assess the cocktail party problem under controlled conditions in a clinical set-up (refer to 2.1.1.1.9 for discussion of the cocktail party problem). However, the cocktail party problem is multidimensional and includes not just sensory aspects but also perception, attention, cognition and interference of masking (Hoen, 2010, slide 5). This brings one back to the important aspect that has to be kept in mind, i.e. that speech perception test results are a mixture of different aspects of speech processing, and no single speech perception test should be used in isolation to determine auditory device candidacy.

Another asset of the present test is that it uses language more suitable for adults, in contrast with other sentence in noise tests such as the HINT (Nilsson et al., 1994), BKB sentences (Bench et al., 1979) and an Afrikaans test for sentence recognition in noise (Theunissen, 2008). The more complex language prevents over-performance in adults due to test material being too simple.

It should be kept in mind that the test's development utilised a relatively small number of participants. No normative data on which to base cochlear implant candidacy, is therefore yet available for this test, since the test scores have not been compared to similar speech-in-noise tests. Until comparative studies have been done, this test should only be seen as suitable for monitoring of progress during rehabilitation, or for the comparison of different amplification systems or strategies, rather than for assessment of auditory device candidacy.

Its high construct validity makes the SAV-L a valuable tool in assessment during audiological rehabilitation, especially since it is planned to increase the sentence lists with an additional 24 lists. This would allow the lists to be administered over a long period of time without the patient's memory corrupting the results. However, it is not advisable to use the test as an auditory training tool, which would make the test unsuitable for assessment due to the participant becoming familiar with the test material.

The sensitivity of a test can be increased by establishing a detailed description of certain aspects, such as its purpose, the population of individuals for whom the test is designed, reliable and equivalent alternative test forms, and lastly a clear procedure for test administration, scoring and interpretation (Bilger, 1984, pp. 2-7). Appendix 10 describes all of these aspects together with a description of additional practical aspects, in order to ensure that the test is administered as intended.

5.4. RECOMMENDATIONS FOR FURTHER RESEARCH

This study was the first stage in the development of the SAV-L. Recommendations for further research to improve and advance this test are indicated below.

- Because users of auditory devices, especially cochlear implants, are evaluated over a long period of time, the sentence lists would benefit from expansion into more lists. This process has already been started by the compilation and recording of a further 24 sentence lists, in addition to the lists developed for this study.

- The concurrent validity of the developed sentence lists has not been established yet. The test could be administered in addition to existing sentence lists that are widely used, such as the Afrikaans translation of the CID sentences, for concurrent validity to be established.

- Running the test in conjunction with another accepted and already used test, will also give an opportunity to develop norms for the SAV-L.

- A longitudinal study with persons undergoing audiological rehabilitation, e.g. cochlear implant or hearing instrument users, should be undertaken to monitor subjective progress versus scores obtained from the test. This will establish whether the test is sensitive enough to show improved scores, when subjective improvement is reported.

- During the development of the SAV-L, the question arose as to what lower age limit or language skill level should be set for the test, so that the participant's language abilities would not adversely affect test scores. Some language structures can develop up to the age of 15 or 16 (Foster-Cohen, 1999, p. 174). A study should therefore be undertaken to establish the lower age limit or level of linguistic abilities where these abilities might start affecting the scores.

- As much as the lower age limit can affect test results, word scores have been found to be significantly better for young adults than for older adults (Nittrouer & Boothroyd, 1990, p. 2710). Especially in noise, younger listeners tend to perform better (Pichora-Fuller, Schneider & Daneman, 1995, p. 607) and the effect of upper age limit on the SAV-L should therefore also be established.

- The sampling method used for the development of the SAV-L drew only few participants from lower socio-economic groups. The test should be administered to members of lower socio-economic groups to assess the influence of socio-economic status on test scores.

- The final version of the sentence lists should be presented to a larger group of participants, to establish the standard error of means for the test. Guidelines as to the critical difference necessary to interpret test results as reflecting improved performance is a valuable, necessary tool. Appendix 11 contains an example of such a tool in the form of the SPRINT chart (Thibodeau, 2010). The 95% confidence interval for improved scores is indicated, and only if a score falls outside this area, can one interpret the differences in scores as significant. Development of a similar tool for the SAV-L should prove to be very valuable.

- The sentence lists should be presented for a third time to the group of participants used during establishment of test-retest reliability, in order to establish reliability over a longer period of time.

- Normative data for the population of interest, i.e. the hearing impaired population, should be established. This is particularly important since information processing can be affected in certain populations, such as the elderly and hearing-impaired (Baldwin & Ash, 2010, p. 5; Rönnberg, 2003, p. S72). If the validity and reliability of the test has been established on the population for whom the test was designed, this will allow the test to perform its purpose more accurately (Lucks Mendel & Danhauer, 1997, p.7). One must, however, keep in mind that this standardisation will come with its own challenges, since it is very hard to establish norms due to the varying nature of the hearing-impaired population (Killion et al., 2004, p. 2401). The undertaking of this aspect should also be carefully considered, because the opposing opinion exists that the performance of a person with hearing impairment should be compared to that of persons with normal hearing, since the normal-hearing world (e.g. a mainstream school, but which could also be expanded to the normal-hearing society), is where the hearing-impaired often needs to function in (Madell, 2010).

- Recently it has been suggested that the traditional level of 70dBSPL to determine cochlear implant candidacy is too loud, since normal conversational speech is at levels of 50 to 60dBSPL (Kirk & Choi, 2009, p.

194). Cochlear implant candidacy criteria based on softer levels of 50 to 60dBSPL have been recommended. Evaluation of performance of the SAV-L at softer levels is, therefore, also recommended.

- The influence of cognitive processes, such as auditory working memory on the test material, should be evaluated more in depth, in order to exclude sentences that are too demanding with regards to the use of higher level processes. The test should be presented to a large group of participants at a comfortable level, i.e. presenting it in quiet at a comfortably audible listening level. Any decrease in scores in such a listening situation could be attributed to the sentence being too taxing on top-down cognitive processes and might need to be excluded from the test material.

- The concept of dynamic assessment is an aspect that should be considered in communities, where cultural diversity is present. This could potentially increase control of linguistic variables and might proof to be a useful frame to develop an assessment tool for use in a culturally diverse population.

5.5. CONCLUSION

The present study focussed on the development of an Afrikaans sentence test for use in the presence of the accompanying three-talker babble. The test was specifically designed for assessment of adults and is called the *Sinslyste in Afrikaans vir Volwassenes in Lawaai* [Sentence lists in Afrikaans for Adults in Noise] (SAV-L). The test provides much needed material for the assessment of candidacy for auditory devices and rehabilitation of hearing impaired adults in Afrikaans, a language for which inadequate sentence test material currently exists. The test was specifically developed with the cochlear implant candidate in mind, since sentence-in-noise testing is already routinely done for this population (Zwolan, 2009, p. 922)

The study led to the development of 27 sentence lists, containing sentences varying in length from 3 to 13 words per list and containing twelve particular topics per list. The test was standardised in, and therefore should be presented in, the presence of the accompanying three-talker babble as masking noise.

One of the strengths of the SAV-L was identified as it being specifically a test for adults due to its more advanced sentences structures, topics and vocabulary used. Three-talker babble as masking noise was also identified as a strong feature of the test. A potential weakness of the test is the increase in inter-list variability due to the use of sentences with different lengths.

Just as any clinician is expected to select the most suitable diagnostic test based on the patient’s history, signs and symptoms, an audiologist is expected to select the most appropriate outcome measure, based on the patient’s communication needs and established treatment goals (Saunders et al., 2005, p. 165). The SAV-L adds a valuable measure to consider when compiling a test battery that is most suitable for assessment and/or monitoring of any particular individual.

REFERENCES

American Speech-Language-Hearing Association (1988). *Determining threshold level for speech* [Guidelines]. Available from www.asha.org/policy.

Arlinger, S. & Hagerman, B. (1997). The Swedish approach to speech audiometry. In M. Martin (Ed.), *Speech audiometry* (2nd ed.) (pp. 278-286). London, England: Whurr Publishers Ltd.

Baldwin, C.L. & Ash, I.K. (2010). Impact of sensory acuity on auditory working memory span in young and older adults. *Psychology and Aging*, pp. 1-7. Advance online publication. Doi: 10.1037/a0020360. Retrieved September 11, 2010, from http://web.ebscohost.com.ez.sun.ac.za/ehost/results?vid=2&hid=104&sid=d6ac4610-5775-4d67-af11-97b847b32a88%40sessionmgr111&bquery=(JN+%22Psychology+and+Aging%22+AND+IK+%2220100816+00000+00000%22)&bdata=JmRiPXBkaCZ0eXBlPTEmc2l0ZT1laG9zdC1saXZl.

Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., et al. (2007). Nucleus Freedom North American clinical trial. *Otolaryngology – Head and Neck Surgery*, 136, 757-762.

Bamford, J. & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 147-190). London, England: Academic Press.

Bellis, T.J. (2003). *Assessment and management of central auditory processing disorders in the educational setting: From science to practice* (2nd ed.). Canada: Thomson Delmar Learning.

Bench, J. & Bamford, J. (Eds.) (1979). *Speech-hearing tests and the spoken language of hearing-impaired children*. London, England: Academic Press.

Bench, J., Kowal, A. & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology*, 13(3), 108-112.

Bentler, R.A. (2000). List equivalency and test-retest reliability of the Speech in Noise Test. *American Journal of Audiology*, 9, 84-100.

Berger, K.W. (1971). Speech Audiometry. In D.E. Rose (Ed.), *Audiological Assessment* (pp. 207-240). Upper Saddle River, New Jersey: Prentice-Hall Inc.

Bilger, R.C. (1984). Speech recognition test development. In E. Elkins (Ed.), Speech recognition by the hearing impaired. *ASHA reports*, 14, 2-15.

BKB-SIN manual. Retrieved June 8, 2009, from http://www.etymotic.com/pdf/bkbsin-user-manual.pdf.

Boothroyd, A. (1968). Developments in speech audiometry. *Sound*, 2, 3-10.

Boothroyd, A. (1995). Speech perception tests and hearing-impaired children. In G. Plant & K. Spens (Eds.), *Profound deafness and speech communication* (pp. 345-371). London, England: Whurr Publishers Ltd.

Boothroyd, A. (1998). The perception of speech by children with hearing loss. In A. Weisel (Ed.), *Issues unresolved. New perspectives on language and deaf education* (pp. 103-116). Washington, D.C.: Gallaudet University Press.

Boothroyd, A. (2006). *Caspersent 3.4: Computer-Assisted Speech Perception Testing and Training: Sentence-level, with F0-only option*. Retrieved October 31, 2010, from http://www-rohan.sdsu.edu/~aboothro/files/CASPERSENT/CASPERSENT_MANUAL.pdf.

Boothroyd, A., Hanin, L. & Hnath, T. (1985). *A sentence test of speech perception: reliability, set equivalence, and short-term learning*. Internal Report RCI 10. New York: City University of New York.

Brinkmann, K. & Richter, U. (1997). Ensuring reliability and comparability of speech audiometry in Germany. In M. Martin (Ed.), *Speech Audiometry* (2nd ed.) (pp. 106-130). London, England: Whurr Publishers Ltd.

British Society of Audiology (1992). Recommended procedure for tympanometry. *British Journal of Audiology*, 26, 255-257.

Broadbent, D.E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74(1), 1-15.

Bronkhorst, A.W. & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of Acoustical Society of* America, 92(6), 3132-3139.

Brungart, D.S., Simpson, B.D., Ericson, M.A. & Scott, K.R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of Acoustical Society of America*, 110(5), 2527-2538.

Cameron, S., Dillon, H. & Newall, P. (2006). Development and evaluation of the listening in spatialized noise test. *Ear & Hearing*, 27(1), 30-42.

Carhart, R., Johnson, C. & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *Journal of Acoustical Society of America*, 58(S1), 535.

Carhart, R. & Tillman, T.W. (1970). Interaction of competing speech signals with hearing losses. *Archives of Otolaryngology*, 91, 273-279.

Carr, J. (1980). The organisation of short courses. In W. Yule & J. Carr (Eds.), *Behaviour modification for the mentally handicapped* (pp. 213-225). Baltimore, Maryland: University Park Press.

Carroll, D.W. (2008). *Psychology of language* (5th ed.). Belmont, California: Thomson Wadsworth.

Carstens, W.A.M. (2003). *Norme vir Afrikaans: Enkele riglyne by die gebruik van Afrikaans* [Norms for Afrikaans: A few guidelines for the use of Afrikaans] (4th ed.). Pretoria, South Africa: Van Schaik.

Chermak, G.D. & Musiek, F.E. (1997). *Central auditory processing disorders: New perspectives*. San Diego, California: Singular Publishing Group, Inc.

Cherry, E. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of Acoustical Society of* America, 25, 975-979.

Chung, K., Zeng, F. & Acker, K.N. (2006). Effects of directional microphone and adaptive multichannel noise reduction algorithm on cochlear implant performance. *Journal of Acoustical Society of America*, 120(4), 2216-2227.

Clark, G. (2003). *Cochlear implants: Fundamentals and applications*. New York, New York: Springer-Verlag New York Inc.

Cleveland, S. (1997). *The ADHD Report Volume Five, Number 5*. Retrieved June 13, 2009, from http://www.ldonline.org/article/Central_Auditory_Processing_Disorder:_ When_is_Evaluation_Referral_Indicated%3F.

Cloete, J. (1997). *'n Afrikaanse vertaling van die BKB-sinne* [An Afrikaans translation of the BKB sentences]. Unpublished manuscript, Department of Speech Therapy and Audiology, University of Stellenbosch.

Cochlear (2008). Nucleus Freedom for previous generation implants. *Implantable Hearing Solutions*, 1, 7-11.

Cox, R.M. (1993). On the evaluation of a new generation of hearing aids. *Journal of Rehabilitation Research & Development*, 30(3), 297-304.

Creston, J.E., Gillespie, M. & Krohn, C. (1966). Speech audiometry: Taped vs live voice. *Archives of Otolaryngology*, 83, 14-17.

Danhauer, J.L., Crawford, S. & Edgerton, B.J. (1984). English, Spanish and bilingual speakers' performance on a nonsense syllable test (NST) of speech sound discrimination. *Journal of Speech and Hearing Disorders*, 49, 164-168.

De Vaus, D. (2005). *Research design in social research*. London, England: Sage Publications Ltd.

Dillon, H. (1983). The effect of test difficulty on the sensitivity of speech discrimination tests. *Journal of the Acoustical Society of America*, 73(1), 336-344.

Dillon, H. & Ching, T. (1995). What makes a good speech test? In G. Plant & K. Spens (Eds.), *Profound deafness and speech communication* (pp. 305-344). London, England: Whurr Publishers Ltd.

Duffey, J. (2007). *Common types of audiograms*. Retrieved August 22, 2010, from http://www.audiologyonline.com/askexpert/display_question.asp?question_id=451.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151-190). Berlin, Germany: Walter de Gruyter GmbH & Co.

Kerkhoff, A. & Vallen, T, (1984). Cultural biases in second language testing of children. In G. Extra & T. Vallen (Eds.). *Ethnic minories and Dutch as a second language* (pp. 133-146). Dordrecht, The Netherlands: Foris Publications Holland.

Firszt, J.B., Holden, L.K., Skinner, M.W., Tobey, E.A., Peterson, A., Gaggle, W., et al. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and Hearing*, 25, 375-387.

Flynn, M.C. (2003). *Sailing out of the windless sea of monosyllables: The use of speech perception tests in aural rehabilitation*. Retrieved April 3, 2009, from http://www.hearingreview.com/issues/articles/2003-04_02.asp.

Foster-Cohen, S.H. (1999). *An introduction to child language development*. Harlow, England: Addison Wesley Longman Ltd.

Freyman, R.L., Balakrishnan, U. & Helfer, K.S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of Acoustical Society of America*, 115(5), 2246-2256.

Gaillard, A.W.K. (2008). Concentration, stress and performance. In P.A. Hancock & J.L. Szalma (Eds.), *Performance under stress* (pp. 59-76). Hampshire, England: Ashgate Publishing Ltd.

Garcia Lecumberri, M.L. & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of Acoustical Society of America*, 119(4), 2445-2454.

Gatehouse, S. (1992). The time course and magnitude of perceptual acclimatisation to frequency responses: Evidence from monaural fitting of hearing aids. *Journal of Acoustical Society of America.* 92(3), 1258-1268.

Gatehouse, S. & Robinson, K. (1997). Speech tests as measures of auditory processing. In M. Martin (Ed.), *Speech audiometry* (2nd ed.) (pp. 74-88). London, England: Whurr Publishers Ltd.

Gelfand, S.A. (1998). Optimizing the reliability of speech recognition scores. *Journal of Speech, Language and Hearing Research*, 41(5), 1088-1102.

Gelfand, S.A. (2009). *Essentials of Audiology* (3rd ed.). New York, New York: Thieme.

Gernsbacher, M.A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113(2), 256-281.

Gillam, R.B., Peña, E.D. & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders*, 20(1), 33-47.

Green, R. (1997). The uses and misuses of speech audiometry in rehabilitation. In M. Martin (Ed.), *Speech Audiometry* (2nd ed.) (pp. 150-175). London, England: Whurr Publishers Ltd.

Hall, J.W. & Mueller, H.G. (1997). *Audiologists' Desk Reference, Volume I: Diagnostic audiology principles, procedures, and practices*. San Diego, California: Singular.

Hällgren, M., Larsby, B. & Arlinger, S. (2006). A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology*, 45 (4), 227-237.

Hanks, W.D. & Johnson, G.D. (1998). HINT list equivalency using older listeners. *Journal of Speech, Language and Hearing Research*, 41(6), 1335-1340.

Hannley, M. (1986). *Basic principles of auditory assessment*. London: Taylor & Francis Ltd.

Hawley, M.L., Litovsky, R.Y. & Culling, J.F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of Acoustical Society of America*, 115(2), 833-843.

Hazan, V. (1997). Speech perception tests for profoundly deaf listeners. In M. Martin (Ed.), *Speech Audiometry* (2^{nd} ed.) (pp. 195-207). London, England: Whurr Publishers Ltd.

Herzog, J. & Rehwinkel, K. (2008). Telephone use in a group of Nucleus cochlear implant recipients. *Implantable Hearing Solutions*, 2(1), 52-54.

Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E. & Benson, R.W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders*, 17, 321-337.

Hoen, M. (2010). Speech comprehension and the cocktail party phenomenon: Dissecting the informational masking effect in speech-in-speech comprehensions, psycholinguistic and neuroscientific approaches. *Phonak 2010 Virtual Conference.* Retrieved September 6, 2010, from https://www.phonakpro.com/com/b2b/en/events/proceedings/conf_virt_audiology_2010.html.

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology*, 92(4), 311-24.

Jerger, J. & Hayes, D. (1977). Diagnostic speech audiometry. *Archives of Otolaryngology*, 103, 216-222.

Johnstone, P.M. & Litovsky, R.Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *Journal of Acoustical Society of America*, 120(4), 2177-2189.

Kalikow, D.N., Stevens, K.N & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61, 1337-1351.

Katz, B. (2002). *Mastering audio: The art and the science*. Canada: Focal Press.

Killion, M.C. (1997). The SIN report: Circuits haven't solved the hearing-in-noise problem. *The Hearing Journal*, 50(10), 28-32.

Killion, M.C. & Niquette, P.A. (2000). What can the pure-tone audiogram tell us about a patient's SNR loss? *The Hearing Journal*, 53(3), 46-53.

Killion, M.C., Niquette, P.A., Gudmundsen, G.I., Revit, L.J. & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of Acoustical Society of America*, 116(4), 2395-2405.

Kim, L., Jang, Y.S., Choi, A., Ahn, S., Park, J., Lee, Y., et al. (2009). Bilateral cochlear implants in children. *Cochlear Implants International*, 10(S1), 74-77.

Kirk, K.I. & Choi, S. (2009). Clinical investigations of cochlear implant performance. In J.K.C. Niparko (Ed.), *Cochlear implants: Principles and practices* (2nd ed.) (pp. 191-222). Philadelphia, Pennsylvania: Lippinkott Williams & Wilkins.

Koch, D.B., Soli, S., Downing, M. & Osberger, M.J. (2009). Simultaneous bilateral cochlear implantation: Prospective study in adults. *Cochlear Implants International*, 11(2), 84-99.

Koehnke, J. & Besing, J. (1996). A procedure for testing speech intelligibility in a virtual listening environment. *Ear & Hearing*, 17(3), 211-217.

Kollmeier, B. & Wesselkamp, M. (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *Journal of the Acoustical Society of America*, 102(4), 2412-2421.

Kowal, A. (1979). Sentence list construction and pilot test. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 111–146). London, England: Academic Press.

Kramer, S.E., Kapteyn, T.S. & Festen, J.M. (1998). The self-reported handicapping effect of hearing disabilities. *Audiology*, 37(5), 302-312.

Kramer, S.E., Kapteyn, T.S., Festen, J.M. & Tobi, H. (1996). The relationship between self-reported hearing disability and measures of auditory disability. *Audiology*, 35(5), 277-287.

Kramer, S.E., Zekveld, A.A. & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology*, 50(5), 507-515.

Kreul, E.J., Bell, D.W. & Nixon, J.C. (1969). Factors affecting speech discrimination test difficulty. *Journal of Speech and Hearing Research*, 12(2), 281-287.

Kruger, B. & Kruger, F.M. (1994). Future trends in hearing aid fitting strategies: With a view towards 2020. In M. Valente (Ed.), *Strategies for selecting and verifying hearing aid fittings* (pp. 300-342). New York, New York: Thieme Medical Publishers, Inc.

Lange-Seidl, A. (1977). *Approaches to theories for nonverbal signs*. The Netherlands: The Peter De Ridder Press.

Laubscher, A.M.U. & Tesner, H.E.C. (1969). *Enkellettergrepige woordelyste in Afrikaans* [Monosyllabic word lists in Afrikaans]. Unpublished manuscript, Department of Speech Therapy and Audiology, University of Pretoria.

Lehiste, I., & Peterson, G.E. (1959). Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustical Society of America*, 31, 280-286.

Lucks Mendel, L. & Danhauer, J.L. (Eds.). (1997). *Audiologic evaluation and management and speech perception assessment*. San Diego, California: Singular Publishing Group Inc.

Lutman, M.E. (1997). Speech tests in quiet and noise as a measure of auditory processing. In M. Martin (Ed.), *Speech audiometry* (2^{nd} ed.) (pp. 63-73). London, England: Whurr Publishers Ltd.

Lyregaard, P. (1997). Towards a theory of speech audiometry tests. In M. Martin (Ed.), *Speech audiometry* (2^{nd} ed.) (pp. 34-62). London, England: Whurr Publishers Ltd.

Mackersie, C.L. (2002). Tests of speech perception abilities. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 10, 392-397.

MacLeod, A. & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for users. *British Journal of Audiology*, 24, 29-43.

Madell, J.R. (2010). *Pediatric amplification: Using speech perception to achieve best outcomes*. Retrieved October 2, 2010, from http://www. audiologyonline. com/ceus/recordedcoursedetails.asp?class_id=17148.

Madell, J.R. & Flexer, C.A. (2008). *Pediatric audiology: Diagnosis, technology, and management*. New York, New York: Thieme.

Martin, M. (Ed.). (1997). *Speech Audiometry* (2nd ed.). London, England: Whurr Publishers Ltd.

Mathers, C.D., Sadana, R., Salomon, J.A., Murray, C.J.L. & Lopez, A.D. (2000). *Estimates of DALE for 191 countries: Methods and results*. World Health Organisation. Retrieved September 25, 2010, from http://www.who.int /healthinfo/paper16.pdf.

McAllister, R. (2003). A speech comprehension measurement tool: Normalization and measurement of perceptual foreign accent [Electronic version]. *Phonum 9*, 177-180.

McArdle, R. & Hnath-Chisolm, T. (2009). Speech Audiometry. In J. Katz, L. Medwetsky, R. Burkard & L. Hood (Eds.), *Handbook of Clinical Audiology* (6th ed.) (pp. 64-79). Baltimore, Maryland: Lippincott Williams & Wilkins.

Medical Research Council (2008). *POCIA support notes*. Retrieved September 25, 2010, from http://www.ihr.mrc.ac.uk/legacy/prostheses/pocia/.

Middelweerd, M.J. & Plomp, R. (1987). The effects of speechreading on the speech reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, 82(6), 2145-2147.

Mueller, G.H. (2001). Speech audiometry and hearing aid fittings: Going steady or casual acquaintances? *Hearing Journal*, 54(10), 19-28.

Müller, A.M.U. & De Stadler, L. (1989). *Afrikaanse vertaling van die CID sinne* [Afrikaans translation of the CID sentences]. Unpublished data.

Müller-Deile, J., Kiefer, J., Wyss, J., Nicolai, J. & Battmer, R. (2008). Performance benefits for adults using a cochlear implant with adaptive dynamic range optimization (ADRO): A comparative study. *Cochlear Implants International*, 9(1), 8-26.

National Institute of Standards and Technology (2003). *Sound absorption test*. Retrieved September 18, 2010, from http://www.perfmetalpanels.com/AS-SA2359A.pdf.

Nilsson, M., Soli, S.D. & Sullivan, J.A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of Acoustical Society of America*, 95(2), 1085-1096.

Niparko, J.K.C., Lingua, C. & Carpenter, R.M. (2009). Assessment of candidacy for cochlear implantation. In J.K.C. Niparko (Ed.), *Cochlear implants: Principles and practices* (2nd ed.) (pp. 137-146). Philadelphia, Pennsylvania: Lippinkott Williams & Wilkins.

Nittrouer, S. & Boothroyd, A. (1990). Context effects in phoneme and word recognition of young children and older adults. *Journal of the Acoustical Society of America*, 87, 2705-2715.

Niquette, P., Arcaroli, J., Revit, L., Parkinson, A., Staller, S., Skinner, M., et al. (2003). *Development of the BKB-SIN Test.* Paper presented at the annual meeting of the American Auditory Society, Scottsdale, Arizona.

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

Olivier, J.M. (2000). *Die vertaling en ontwikkeling van sinsmateriaal vir die evaluasie van spraakpersepsie by Xhosa-sprekendes* [The translation and development of sentence material for the evaluation of speech perception in Xhosa-speaking persons]. Unpublished manuscript, Department of Communication Pathology, University of Pretoria.

Olsen, W.O. (1998). Average speech levels and spectra in various speaking/listening conditions. *American Journal of Audiology*, 7, 21-25.

Olsen, W.O. & Matkin, N.D. (1991). Speech Audiometry. In W.F. Rintelmann, (Ed.), *Hearing Assessment* (2nd ed.) (pp. 39-140). Needham Heights, Massachusetts: Allyn and Bacon.

Organisation Internationale de Métrologie Légale (1993). *International recommendation: Pure-tone audiometers*, OIML R104. Retrieved June 30, 2009, from http://www.oiml.org/publications/R/R106-e93.pdf.

Owens, E. (1961). Intelligibility of words varying in familiarity. *Journal of Speech and Hearing Research*, 4, 113-129.

Owens, R.E. (2001). *Language development: An introduction* (5th ed.). Needham Heights, Massachusetts: Allyn & Bacon.

Paivio, A. & Begg, I. (1981). *Psychology of language*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Pichora-Fuller, M.K., Schneider, B.A. & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1), 593-608.

Plant, G. (1997). Testing visual and auditory-visual speech perception. In M. Martin (Ed.), *Speech audiometry* (2nd ed.) (pp. 208-232). London, England: Whurr Publishers Ltd.

Plomp, R. (1994). Noise, amplification, and compression: Considerations of three main issues in hearing aid design. *Ear and Hearing*, 15(1), 2-12.

Rhebergen, K.S., Versfeld, N.J. & Dreschler, W.A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *Journal of Acoustical Society of America*, 118(3), 1274-1277.

Roeser, R.J. & Clark, J.L. (2007). Clinical masking. In R.J. Roeser, M. Valente & H Hosford-Dunn (Eds.), *Audiology. Diagnosis* (2nd ed.) (pp. 261-287). New York, New York: Thieme.

Roets, J. (2006). *Spraakoudiometrie in Suid-Afrika: Ideale kriteria teenoor kliniese praktyk* [Speech audiometry in South Africa: Ideal criteria against clinical practice]. Unpublished manuscript, Department of Communication Pathology, University of Pretoria.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*. 42(S1), S68-S76.

Royal National Institute for the Deaf (2002). *Best practice standards for adult Audiology*. London: RNID.

Rudner, M., Foo, C., Rönnberg, J. & Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal of Psychology*, 50(5), 405-418.

Sanders, J.W. (1991). Clinical masking. In W.F. Rintelmann (Ed.), *Hearing Assessment* (2nd ed.) (pp. 141-178). Needham Heights, Massachusetts: Allyn and Bacon.

Sarant, J.Z., Blamey, P.J., Cowan, R.S. & Clark, G.M. (1997). The effect of language knowledge on speech perception: What are we really assessing? *The American Journal of Otology*, 18, S135-S137.

Saunders, G.H., Chisolm, T.H. & Abrams, H.B. (2005). Measuring hearing aid outcomes – Not as easy as it seems. *Journal of Rehabilitation Research and Development*, 42(4), 157-168.

Savin, H.B. (1963). Word-frequency effects and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35(2), 200-206.

Schafer, E.C. & Thibodeau, L.M. (2006). Speech recognition in noise in children with CIs while listening in bilateral, bimodal, and FM-system arrangements. *American Journal of Audiology*, 15(2), 114-126.

Schmitt, N. (2004). *A new speech test (BEST test). Practical training report*. Sydney, Australia: National Acoustics Laboratory.

Sherwood, T. & Fuller, H. (1997). Equipment for speech audiometry and its calibration. In M. Martin (Ed.), *Speech Audiometry* (2nd ed.) (pp. 89-105). London, England: Whurr Publishers Ltd.

Shore, I., Bilger, R.C. & Hirsh, I.J. (1960). Hearing aid evaluation: Reliability of repeated measurements. *Journal of Speech and Hearing Disorders*, 25(2), 152-170.

Silverman, S.R. & Hirsh, I.J. (1955). Problems related to the use of speech in clinical audiometry. *Annals of Otology, Rhinology and Laryngology*, 64, 1234-1244.

Simpson, S. & Cooke, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of N. *Journal of Acoustical Society of America*, 118(5), 2775-2778.

Skinner, M.W., Holden, L.K, Holden, T.A., Demorest, M.E, & Fourakis, M.S. (1997). Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants. *Journal of Acoustical Society of America*, 101(6), 3766-3781.

Smoorenburg, G.F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *Journal of Acoustical Society of America*, 91(1), 421-436.

Soli, S.D. (2003). *Protocols for the development of HINT in a new language*. Los Angeles, California: House Ear Institute.

South African Bureau of Standards (1996). *SANS 5253-3:1996 Acoustics – Audiometric test methods part 3: Speech audiometry*. Pretoria, South Africa: SABS.

South African Bureau of Standards (2004). *SANS 1054-1:2004 Calibration of pure-tone audiometers Part 1: Air Conduction.* Pretoria, South Africa: SABS.

South African Bureau of Standards (2004). *SANS 1054-2:2004 Calibration of pure-tone audiometers Part 2: Bone Conduction.* Pretoria, South Africa: SABS.

Speaks, C., Jerger, J. & Trammell, J. (1970). Comparison of sentence identification and conventional speech discrimination scores. *Journal of Speech and Hearing Research*, 13, 755-767.

Stach, B.A. (2010). *Clinical Audiology: An introduction*. San Diego, California: Singular Publishing Group, Inc.

Standards South Africa (1996). *South African National Standard SANS 8253-3: Acoustics – Audiometric test methods Part 3: Speech audiometry*. Pretoria, South Africa: Standards South Africa.

Statistics South Africa (2001). *Census 2001*. Retrieved June 28, 2009, from http://www.southafrica.info/about/people/language.htm.

Strom, K.E. (2006). The HR 2006 dispenser survey [Electronic version]. *Hearing Review*, June 2006. Retrieved September 18, 2010, from http://www.hearingreview.com/issues/articles/2006-06_11.asp.

Taylor, B. (2003). Speech-in-noise tests: How and why to include them in your basic test battery. *The Hearing Journal*, 56(1), 40-45.

Theunissen, M. (2008). *Development of an Afrikaans Test for Sentence Recognition in Noise*. Unpublished manuscript, Department of Communication Pathology, University of Pretoria.

Thibodeau, L.M. (2010). *SPRINT Chart for 25 words*. Retrieved September 11, 2010, from www.utdallas.edu/~thib.

Thomas, E.J., 1987. Developmental approach to research. *Encyclopedia of Social Work*, 18(1), 382-388.

Tobias, J.V. (1964). On phonemic analysis of speech discrimination tests. *Journal of Speech and Hearing Research*, 7, 98-100.

Tye-Murray, N. & Tyler, R.S. (1988). A critique of continuous discourse tracking as a test procedure. *Journal of Speech and Hearing Disorders*, 53 (3), 226-231.

Understanding dB. Retrieved October 9, 2010, from http://www.jimprice.com/prosound/db.htm

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Prophy, A., et al. (2005). Adaptation of the HINT (Hearing In Noise Test) for adult Canadian Francophone populations. *International Journal of Audiology*, 44, 358-369.

Van Engen, K.J. & Bradlow, A.R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of Acoustical Society of America*, 121(1), 519-526.

Versveld, N.J., Daalder, L., Festen, J.M. & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception thresholds. *Journal of the Acoustical Society of America*, 107(3), 1671-1684.

Walden, B.E. (1984). Validity issues in speech recognition test development. In E. Elkins (Ed.), *Speech recognition by the hearing impaired* (pp. 16-18). ASHA reports 14. Rockville, Maryland: American Speech-Language-Hearing Association.

Walden, T.C. (2006). Evaluating and treating hearing loss in the older patient: Clinical case studies. *Seminars in Hearing*, 27, 311-319.

Walden, T. & Walden, B. (2004). Predicting success with hearing aids in everyday living. *Journal of American Academy of Audiology*, 15(5), 342-352.

Weinstein, B.E. (2009). Hearing loss in the elderly: A new look at an old problem. In J. Katz, L. Medwetsky, R. Burkard & L.J. Hood (Eds.), *Handbook of clinical audiology* (6th ed.) (pp. 712 – 725). Baltimore, Maryland: Lippincott Williams & Wilkins.

Williams, A.M. & Ward, P, (2003). Perceptual expertise: Development in sport. In J.L. Starkes & K.A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 219-250). Champaign, Illinois: Human Kinetics.

Wilson, R.H., Burks, C.A. & Weakley, D.G. (2005). A comparison of word-recognition abilities assessed with digit pairs and digit triplets in multitalker babble. *Journal of Rehabilitation Research & Development*, 42(4), 499-510.

Wilson, R.H. & McArdle, R. (2005). Speech signals used to evaluate functional status of the auditory system. *Journal of Rehabilitation Research and Development*, 42(4), 79-94.

Wong, L.L & Soli, S.D. (2005). Development of the Cantonese Hearing In Noise Test (CHINT). *Ear and Hearing*, 26, 276-289.

Wong, L.L.N., Soli, S.D., Liu, S., Han, N. & Huang, M. (2007). Development of the Mandarin Hearing In Noise Test (MHINT). *Ear & Hearing*, 28, 70S-74S.

World Health Organisation (n.d.). *Disabilities*. Retrieved September 25, 2010, from http://www.who.int/topics/disabilities/en/.

Yotis, J. (2005). *Optimisation of the measurement of speech perception performance using the Beautifully Efficient Speech Test and evaluation of the word Intelligibility of a single Beautifully Efficient Speech Test list*. Unpublished manuscript, Department of Otolaryngology, University of Melbourne.

Zwolan, T.A. (2009). Cochlear Implants. In J. Katz, L. Medwetsky, R. Burkard & L. Hood (Eds.), *Handbook of Clinical Audiology* (6th ed.) (pp. 913-933). Baltimore, Maryland: Lippincott Williams & Wilkins.

APPENDIX 1

INFORMATION LETTER FOR RATERS OF NATURALNESS

BEOORDELAARS BENODIG VIR NAVORSINGSPROJEK.

Titel van navorsingsprojek: ***Development of an Afrikaans sentence perception test based on the CUNY topic-related sentences – Phase 1: Sentence Perception in Noise***

(*Ontwikkeling van 'n Afrikaanse sinspersepsietoets gebaseer op die CUNY onderwerp-verwante sinne - Fase 1: Spraakpersepsie in Geraas*)
Hoofnavorser: Jolanda Scourfield

Baie dankie dat u ingestem het om my te help met my navorsingsprojek. Ek is besig om 'n toets te ontwikkel om uitermatig gehoorgestremde mense te toets vir hulle geskiktheid ten opsigte van 'n kogleêre inplanting en om hulle vordering te monitor ná hulle 'n inplanting gekry het. Voor die sinne opgeneem word, sal ek dit opreg waardeer as u my kan help om die natuurlikheid van die sinne te beoordeel. U deelname is heeltemal vrywilliglik.

Wat van u verwag word:

- Aangeheg is 27 sinslyste van 12 sinne elk (genommer 1 – 24 en 49 – 51) wat u op u eie tyd moet deurlees.
- Vir elke sin moet u 'n telling uit 5 gee deur 'n nommer van 1 tot 5 langs elke sin te omkring, op grond van of u dink dat die sinne natuurlik is, met ander woorde in alledaagse Afrikaanse gesprekke gehoor kan word.
- Tellings:
    - 1 = Heeltemal onnatuurlik, hierdie sin sal mens glad nie teëkom nie
    - 2 = Effens onnatuurlik, dit is onwaarskynlik dat mens so 'n sin sal teëkom
    - 3 = Die sin is ongewoon, maar mens mag dit dalk hoor
    - 4 = Mens behoort so 'n sin in alledaagse gesprekke te hoor
    - 5 = Mens sal verseker so 'n sin kan teëkom in alledaagse gesprekke
- Indien 'n sin minder as 'n telling van 5 kry volgens u, moet u asseblief 'n alternatiewe sin voorstel wat meer natuurlik is.
- Ek sal dit waardeer as u teen die 19e Oktober 2009 kan klaar wees. Ek sal u dan kontak om te reël om die lyste van u terug te kry.

Baie dankie vir u hulp met my studie; dit word opreg waardeer. Kontak my asseblief direk indien u enige verdere vrae het of enige probleme ondervind.

Vriendelike groete

Jolanda Scourfield
Oudioloog
W: **********
S: **********
E-pos: *****************

APPENDIX 2

LETTER OF CONSENT FOR RATERS OF NATURALNESS

Verklaring:

Met die ondertekening van hierdie dokument stel ek myself / stel ek myself nie *(krap dood wat nie van toepassing is nie)* **beskikbaar vir deelname aan die navorsingsprojek getiteld** ***Development of an Afrikaans sentence perception test based on the CUNY topic-related sentences – Phase 1: Sentence Perception in Noise*** **(*Ontwikkeling van 'n Afrikaanse sinspersepsietoets gebaseer op die CUNY onderwerp-verwante sinne - Fase 1: Spraakpersepsie in Geraas*), om sinne te beoordeel vir hulle natuurlikheid.**

Ek verklaar dat:

- ek hierdie inligtings- en toestemmingsvorm gelees het.
- ek verstaan dat deelname aan hierdie navorsingsprojek **vrywillig** is en dat daar geen druk op my geplaas word om deel te neem nie.
- ek voorneem om die vorms teen die versoekte tyd te voltooi, om sodoende te verseker dat die navorsing stiptelik kan plaasvind. Ek verstaan egter dat ek op enige tyd kan onttrek en dat ek nie op enige wyse daardeur benadeel sal word nie.
- indien ek verkies om te onttrek, sal ek die hoofnavorser so gou as moontlik in kennis stel.

Geteken te *(plek)* .... op *(datum)* .... 20.....

..
Naam van deelnemer

..
Handtekening van deelnemer

Kontaktelefoonnommer(s) van deelnemer **..**

APPENDIX 3

DISTRIBUTION OF AGE AND GENDER OF THE PARTICIPANTS IN THE EVALUATION OF THE RECORDED TEST

Round number	Participant number	Participant age (years)	Average age	Participant gender (F = Female; M = Male)	Gender distribution
Pilot	1	29	29.4	F	F=7; M=3
	2	34		F	
	3	41		F	
	4	24		F	
	5	25		F	
	6	31		M	
	7	28		F	
	8	25		F	
	9	29		M	
	10	28		M	
1	1	37	29.3	F	F=8; M=2
	2	32		F	
	3	23		F	
	4	35		M	
	5	24		F	
	6	32		F	
	7	34		M	
	8	27		F	
	9	29		F	
	10	20		F	
2	1	27	28.2	M	F=7; M=2
	2	33		F	
	3	23		F	
	4	21		F	
	5	34		F	
	6	25		F	
	7	34		F	
	8	22		F	
	9	35		M	
3	1	28	25	M	F=5; M=5
	2	16		M	
	3	16		M	
	4	36		F	
	5	32		M	
	6	34		F	
	7	38		F	
	8	17		F	
	9	15		F	
	10	18		M	

4	1	17	22.6	F	F= 5; M=5
	2	36		M	
	3	18		M	
	4	16		M	
	5	17		M	
	6	34		F	
	7	18		F	
	8	19		F	
	9	24		M	
	10	27		F	
5	1	15	24.1	M	F=6; M=4
	2	39		F	
	3	17		F	
	4	24		F	
	5	38		F	
	6	17		M	
	7	16		M	
	8	44		F	
	9	15		F	
	10	16		M	
6	1	24	30.6	F	F=7; M=3
	2	42		F	
	3	24		F	
	4	23		F	
	5	29		M	
	6	38		F	
	7	35		M	
	8	29		F	
	9	29		M	
	10	33		F	
7	1	27	31	F	F=5; M=0
	2	15		F	
	3	40		F	
	4	37		F	
	5	36		F	

APPENDIX 4

INFORMATION LETTER TO PARTICIPANTS IN EVALUATION OF THE TEST MATERIAL

Beste Potensiële Deelnemer

DEELNEMERS BENODIG VIR NAVORSINGSPROJEK.
Titel van navorsingsprojek: ***Development of an Afrikaans sentence perception test based on the CUNY topic-related sentences – Phase 1: Sentence Perception in Noise***

(*Ontwikkeling van 'n Afrikaanse sinspersepsietoets gebaseer op die CUNY onderwerp-verwante sinne - Fase 1: Spraakpersepsie in Geraas*)

Hoofnavorser: Jolanda Scourfield
Kontaknommer: *********

Baie dankie vir u bereidwilligheid om my te help in my navorsing. Ek is besig om 'n toets te ontwikkel om uitermatig gehoorgestremde mense te toets vir hulle geskiktheid ten opsigte van 'n kogleêre inplanting en om hulle vordering te monitor ná hulle 'n inplanting gekry het. Ek het hiervoor die hulp nodig van ongeveer 40 tot 50 mense tussen die ouderdom van 15 en 45 jaar en sal dit baie waardeer as u vir my 2 ure van u tyd kan gee. U sal met R100* vergoed word vir u reisonkostes.

Lees asseblief hierdie inligtingsblad op u tyd deur, aangesien die detail van die navorsingsprojek hierin verduidelik word. Indien daar enige deel van die navorsingsprojek is wat u nie ten volle verstaan nie, is u welkom om die hoofnavorser daaroor uit te vra. Dit is baie belangrik dat u ten volle moet verstaan wat die navorsingsprojek behels en hoe u daarby betrokke kan wees. U deelname is ook volkome vrywillig en u sal op geen wyse hoegenaamd negatief beïnvloed word indien u sou weier om deel te neem nie. U mag ook te eniger tyd aan die navorsingsprojek onttrek, selfs al het u ingestem om deel te neem.

Hierdie navorsingsprojek is deur die Komitee vir Mensnavorsing van die Universiteit Stellenbosch goedgekeur en sal uitgevoer word volgens die etiese riglyne en beginsels van die Internasionale Verklaring van Helsinki en die Etiese Riglyne vir Navorsing van die Mediese Navorsingsraad (MNR).

Wat behels hierdie navorsingsprojek?

- *Die projek behels die ontwikkeling van 'n sinstoets wat gebruik gaan word om mense met uitermatige gehoorverliese te toets vir hulle kandidaatskap vir kogleêre inplantings. Dit sal ook gebruik kan word tydens die evaluasie en monitor van vordering van gehoorapparaatgebruikers.*
- *U sal een van ongeveer 40-50 persone wees wat die Oorinstituut, Bellville (hoek van Mike Pienaarrylaan en Oosterzeestraat) een keer sal besoek. By die besoek gaan die volgende gedoen word:*
    - *Daar gaan in u ore gekyk word.*
    - *Daar gaan 'n middeloortoets gedoen word waar 'n proppie teen u oorkanaal vir 'n paar sekondes gehou word terwyl die lugdruk liggies verander word, om te kyk of u oordrom gesond beweeg.*
    - *U gehoor gaan getoets word deurdat u na sagte klanke deur oorfone luister en u moet aandui as u die klanke hoor.*
    - *Sinne met spraakstemme as agtergrondsgeraas gaan aan u gespeel word deur oorfone en u gaan die sinne moet terugherhaal.*
- *Enkele persone mag vooraf gevra word om die sinstoets vir 'n tweede keer te doen.*

Waarom is u genooi om deel te neem?

- *U is spesifiek genader omdat u binne gerieflike afstand van die Oorinstituut is en u binne die ouderdomsgroep val wat vereis word vir die toets.*

- *Die toets moet op normaalhorende persone uitgevoer word. Indien u weet dat u 'n gehoorverlies het, is u dus nie 'n geskikte kandidaat nie. Indien u egter wel aan die studie deelneem en daar bevind word dat daar 'n probleem is met u ore/gehoor, sal u volledig in kennis gestel word en die nodige aanbevelings gemaak word. In hierdie geval sal u nie die geldelike vergoeding ontvang nie.*

Is u 'n geskikte deelnemer?

- *Om seker te maak dat u 'n geskikte deelnemer is, maak asseblief seker dat u aan die volgende vereistes voldoen:*
    - *Afrikaans moet u eerste taal wees.*
    - *U moenie nie 'n gehoorverlies hê sover u weet nie.*
    - *U moenie 'n geskiedenis van uitermatige middeloorprobleme as 'n kind hê nie (nie meer as 3 stelle pypies/grommets nie).*
    - *U moenie lááť taal begin ontwikkel het as 'n kind nie.*
    - *U moenie oorsensief vir baie harde klanke wees nie.*

Wat sal u verantwoordelikhede wees?

- *Daar sal van u verwag word om u weg na die Oorinstituut te maak. Toetsing sal plaasvind op 'n weeksmiddag vanaf 16.00 of op 'n Saterdag, afhangend wat vir u meer geskik is.*

Sal u voordeel trek deur deel te neem aan hierdie navorsingsprojek?

- *U sal met R100* vergoed word vir u reiskostes.*
- *U identiteit sal op geen manier gebruik word in die rapportering van die resultate nie, alhoewel die navorsingsdata deur studieleiers en assessors bestudeer mag word.*

Is daar enige risiko's verbonde aan u deelname aan hierdie navorsingsprojek?

- *Die toets sal op goeie, werkende apparaat aangebied word deur gekwalifiseerde oudioloë, en die kanse dat u aan te gevaarlik harde klanke blootgestel word, is dus hoogs onwaarskynlik. Daar is geen ander risiko's vir u verbonde aan die projek nie.*

Sal u betaal word vir deelname aan die navorsingsprojek en is daar enige koste verbonde aan deelname?

- *U sal met R100* vergoed word vir u reiskostes indien u aan die vereistes as deelnemer voldoen.*

Is daar enigiets anders wat u moet weet of doen?

- *U kan Jolanda Scourfield kontak by telefoonnommer ********** indien u enige verdere vrae het of enige probleme ondervind.*
- *U kan die Komitee vir Mensnavorsing kontak by 021-938 9207 indien u enige bekommernis of klagte het wat nie bevredigend deur u studieleier hanteer is nie.*
- *U moet die onderstaande nommer kontak om 'n geskikte tyd te reël.*
- *U moet asseblief die aangehegde vorm invul en dit na u afspraak bring.*

**R100 sal aan u betaal word op die dag van toetsing. Slegs persone wat aan die minimum vereistes voldoen vir die studie soos hierbo uiteengesit, kan vergoed word.*

Vriendelike groete

Jolanda Scourfield
Oudioloog
W: **********
E-pos: ***********

APPENDIX 5a

INFORMED CONSENT FORM FOR ADULTS WHO PARTICIPATED IN EVALUATION OF THE TEST MATERIAL

<u>Verklaring:</u>

Met die ondertekening van hierdie dokument, stem ek in vir deelname aan die navorsingsprojek getiteld *Development of an Afrikaans sentence perception test based on the CUNY topic-related sentences – Phase 1: Sentence Perception in Noise* (*Ontwikkeling van 'n Afrikaanse sinspersepsietoets gebaseer op die CUNY onderwerp-verwante sinne - Fase 1: Spraakpersepsie in Geraas*)

Ek verklaar dat:

- ek hierdie inligtings- en toestemmingsvorm gelees het.
- ek verstaan dat deelname aan hierdie navorsingsprojek **vrywillig** is en dat daar geen druk op my geplaas word om deel te neem nie.
- ek te eniger tyd aan die navorsingsprojek mag onttrek en dat ek nie op enige wyse daardeur benadeel sal word nie.
- ek in kennis gestel mag word as my deelname aan die studie nie meer benodig word nie.

Geteken te *(plek)* .... op *(datum)* ....

....
Handtekening van deelnemer

<u>Verklaring deur navorser (om voltooi te word op die dag van toetsing):</u>

Ek *(naam)* .... verklaar dat:

- ek die inligting in hierdie dokument verduidelik het aan ....
- ek hom/haar aangemoedig het om vrae te vra en voldoende tyd gebruik het om dit te beantwoord.
- ek tevrede is dat hy/sy al die aspekte van die navorsingsprojek soos hierbo bespreek, voldoende verstaan.

Geteken te *(plek)* .... op *(datum)* .... 2010.

....
Handtekening van navorser

☐ R100 ontvang (geteken) .... Datum ....

APPENDIX 5b

<u>INFORMED CONSENT FORM FOR SCHOLARS WHO PARTICIPATED IN THE EVALUATION OF THE TEST MATERIAL AND NEEDED CONSENT FROM THEIR PARENTS</u>

Verklaring deur ouer:

Met die ondertekening van hierdie dokument gee ek / gee ek nie *(krap dood wat nie van toepassing is nie)* **toestemming vir deelname van my kind, ..., aan die navorsingsprojek getiteld** ***Development of an Afrikaans sentence perception test based on the CUNY topic-related sentences – Phase 1: Sentence perception in Noise*** **(*Ontwikkeling van 'n Afrikaanse sinspersepsietoets gebaseer op die CUNY onderwerp-verwante sinne - Fase 1: Spraakpersepsie in Geraas*)**

Ek verklaar dat:

- ek hierdie inligtings- en toestemmingsvorm gelees het.
- ek verstaan dat deelname aan hierdie navorsingsprojek **vrywillig** is en dat daar geen druk op my kind geplaas is om deel te neem nie.
- my kind te eniger tyd aan die navorsingsprojek mag onttrek en dat hy/sy nie op enige wyse daardeur benadeel sal word nie.
- ek in kennis gestel mag word as my kind se deelname aan die studie nie meer benodig word nie.

Geteken te *(plek)* .... op *(datum)* .... 20.....

....	....
Handtekening van deelnemer	**Naam en Handtekening van ouer**

Kontakpersoon en telefoonnommer(s):

Naam: .... Verhouding tot kind: ....

Kontaktelefoonnommer(s): ....

Verklaring deur navorser (om voltooi te word op die dag van toetsing):

Ek *(naam)* .... verklaar dat:

- ek die inligting in hierdie dokument verduidelik het aan ....
- ek hom/haar aangemoedig het om vrae te vra en voldoende tyd gebruik het om dit te beantwoord.
- ek tevrede is dat hy/sy al die aspekte van die navorsingsprojek soos hierbo bespreek, voldoende verstaan.

Geteken te *(plek)* .... op *(datum)* .... 2010.

....
Handtekening van navorser

☐ Geld ontvang (geteken) .... Datum ....

APPENDIX 6

SCREENING FORM

Amptelike gebruik

Subject number	
Round number	

******Dankie vir u deelname aan die studie. Beantwoord asseblief die 7 vrae in skuinsskrif en oorhandig die vorm aan die oudioloog wat u gaan toets. ******

1. Wat is u naam? ____________________ ***2. Wat is u telefoonnommer?*** ____________

---

3. Is Afrikaans u eerste taal? Ja / Nee ***4. Wat is u ouderdom?***

5. ***Het u 'n geskiedenis van uitermatige middeloorprobleme/-pypies?*** (het u meer as 2 stelle pypies/grommets as jong kind gehad?) Ja/Nee

6. ***Het u láát taal ontwikkel?*** (moes u 'n spraakterapeut sien vir taalterapie of was u eerste woorde later as die ouderdom van 18 maande gesê?) Ja/Nee

7. ***Is u sensitief vir harde klanke?*** (onvermoë om sportdae, winkelsentrums, ens. te besoek weens ongemak vir harde klanke) Ja/Nee

Otoscopy

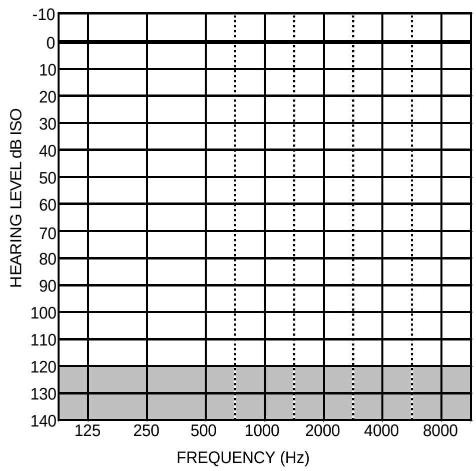
R	
L	

Tympanometry

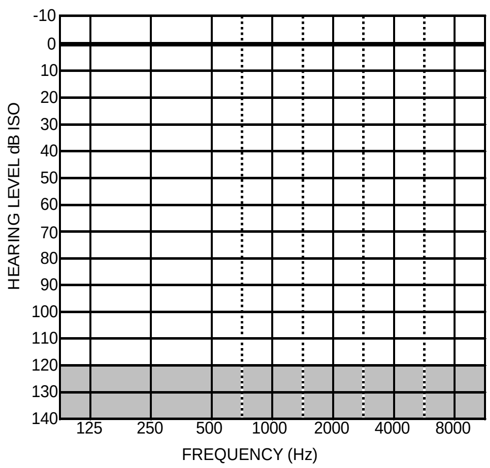
	R	L
ECV		
MEP (-50 - +50 daPa)		
Compliance (0.3 - 1.6cm^3)		

Air Conduction Pure Tone Audiometry

RIGHT O



LEFT X



Subject Number ____________ Round Number ________________ Date ____________________

APPENDIX 7

APPROVAL FROM STELLENBOSCH UNIVERSITY HEALTH RESEARCH ETHICS COMMITTEE

22 February 2010 **MAILED**

Ms J Scourfield
Dept of Speech Language and Hearing Therapy
4th Floor, Teaching Building
Stellenbosch University
Tygerberg campus
7505

Dear Ms Scourfield

"Translation of the CUNY sentence test into Afrikaans-Phase 1: Sentence recognition in noise."

ETHICS REFERENCE NO: N09/09/240

RE : AMENDMENT

Your e-mail dated 17 February 2010 refers.

The Chairperson of the Health Research Ethics Committee approved the amended documentation, provided the word 'gratis' is removed. A copy of the amended advertisement is required for our records.

Yours faithfully

MRS EL ROHLAND
RESEARCH DEVELOPMENT AND SUPPORT
Tel: 021 938 9677 / E-mail: elr@sun.ac.za
Fax: 021 931 3352

APPENDIX 8

PRACTICE LISTS FOR EVALUATION OF TEST MATERIAL

	SIGNAL DB	**List 28 (Track 29)**	**Syllable count**
1	55	Wil jy vanaand by daardie nuwe restaurant gaan eet?	14
2	54	Woon hy nog by sy ouers?	7
3	53	Ek moet uniforms vir die kinders koop.	11
4	52	Kort jy 'n nuwe das om by daardie hemp te pas?	13
5	51	Die eekhorings in die park eet baie graag die voëls se saad.	15
6	50	Sy moet iemand kry om haar woonstel mee te deel.	12
7	50	Ek probeer elke dag oefen.	8
8	49	Daar was vir jare nie 'n groot storm in daardie a-rea nie.	17
9	49	Sit jou hand voor jou mond wanneer jy hoes anders versprei jy kieme.	17
10	48	Koop liewer 'n kamera van beter kwaliteit.	13
11	48	Pas jou geld op.	4
12	48	Speel jou kitaar.	4
			135
		List 29 (Track 30)	
1	47	Sal jy 'n boud vir môre gaarmaak?	9
2	47	Is jou susters nog enkel of is hulle getroud?	13
3	47	Dink jy sy kan al die reëlings vir die vergadering self tref?	16
4	46	Laat jou broek was anders sal jy dit nie kan dra nie.	13
5	46	Ons het na die dieretuin gegaan en die reptiele gesien.	16
6	46	Die sentrale verhitting moet reggemaak word.	12
7	45	Die skare in die stadion het gejuig toe die span die wedstryd wen.	18
8	45	Hoor jy die wind?	4
9	45	Gaan dokter toe en kry 'n bietjie raad.	10
10	44	Leen vir my daardie boek.	6
11	44	Betaal die rekening.	6
12	44	Hulle koop altyd kaartjies vir die jazz- en die popkonserte.	16
			139
		List 30 (Track 31)	
1	43	Verkies jy sop of 'n toebroodjie vir middagete?	14
2	43	Wie kyk na jou kinders wanneer hulle van die skool af kom?	15
3	43	Maak seker dat jy al die toerusting afskakel voor jy die kantoor verlaat.	20
4	43	Trek asseblief jou stewels aan en knoop jou jas toe.	13
5	43	Ek het nog nooit 'n wit slang gesien nie.	10
6	43	My vriende het 'n ou huis gekoop en begin dit nou uitverf.	16
7	42	Waar het jy leer swem?	5
8	42	Het jy gisteraand die maan gesien?	9
9	42	Roep asseblief die verpleegster.	8
10	42	Verf 'n prentjie.	4
11	42	Dis baie duur om die hele familie fliek toe te neem.	15
12	42	Sy rig musiek af by die skool.	8
			137

APPENDIX 9

FINAL SENTENCE LISTS OF THE SAV-L

Set 1

1 Het jy al geëet?
2 Hoeveel van jou broers bly nog by jou ouers?
3 Hoeveel jaar se opleiding het dit geneem om 'n verpleegster te word?
4 Waar kan ek my klere laat was?
5 Katte is maklike troeteldiere, want hulle kom en gaan soos hulle wil.
6 Sy het in 'n drieslaapkamerwoonstel ingetrek.
7 Ek wil graag tennis speel.
8 Ek hoop dit reën, want my tuin is redelik droog.
9 Doen wat die dokter gesê het, anders gaan jy sieker word.
10 Sny die koek.
11 Betaal die tjek voor môre in.
12 Onthou asseblief om vir 'n uur te oefen voor jou volgende klavierles.

Set 2

1 Wil jy miskien vanaand by ons huis kom braai?
2 Wanneer laas het jy by jou ouers op die plaas gaan kuier?
3 Hoe laat hou jy op met werk?
4 Moet asseblief nie weer te veel waspoeier in die masjien gooi nie.
5 Jy kan beeste naby ons huis sien wei.
6 Daardie huis is te koop.
7 Die nuwe netbalbaan is sowaar reg langs die skool.
8 Kan julle onthou wanneer laas ons 'n erge hittegolf gehad het?
9 Drink jou medisyne.
10 Hang die versierings aan die boom.
11 Moenie jou kredietkaart gebruik as jy nie die rekening kan betaal nie.
12 Ek onthou daardie liedjie.

Set 3

1 Wil jy gebraaide hoender of eerder kerrie en rys vir ete hê?
2 Hoe lank is jou dogter al getroud?
3 As jy weer gereed is om te gaan werk, laat my weet.
4 Koop daardie swembroek, want dit pas jou perfek.
5 Die hond wil heeldag speel.
6 Ons gaan die naweek ons huurhuis begin uitverf.
7 Hoe ver hardloop jy elke dag as voorbereiding vir 'n maraton?
8 Het dit gereën?
9 Onthou om jou vitamienes te drink.
10 Koop jy die kos, dan sal ek die drankies vir die braai kry.
11 Daardie winkel is duur.
12 Die klassieke musiekkonserte in die park is baie gewild.

Set 4

1 Wil jy sous oor jou aartappels hê?
2 Indien enigsins moontlik, moet jy jou hele familie na die onthaal nooi.
3 Sorg dat jy betyds by die werk is.
4 Gooi jou ou skoene weg.
5 My katte hou daarvan om na die voëls te kyk.
6 Hoe lank het dit jou geneem om die mat te lê?
7 Draf jy graag?
8 Wat is volgende week se weervoorspelling?
9 Moenie te lank in die son bly nie, anders gaan jy lelik verbrand.
10 Somer is uiteindelik hier.
11 Ek moet vandag bank toe gaan voor dit toemaak.
12 Hy het altyd op gehoor gespeel en eers onlangs musiek leer lees.

Set 5

1 Gaan haal asseblief die inkopies uit die motor en pak dit weg.
2 Onthou om hierdie week jou ma te bel.
3 Voltooi asseblief dadelik die verslag.
4 Neem die rok terug en koop iets waarvan jy hou.
5 Het julle in die moeilikheid beland omdat julle 'n hond aanhou?
6 Waar's jou kamer?
7 Wanneer het jy gaan perdry?
8 Weet jy of dit dié tyd van die jaar hier rond baie reën?
9 Sy het maaggriep.
10 Ek hou van bergklim in die herfs en lente.
11 Baie van my vriende moes al weer studielenings hierdie jaar uitneem.
12 Hulle vorige liedjie was 'n groot treffer.

Set 6

1 Moenie weer die vleis en groente verbrand nie.
2 Neem jou ma kerk toe.
3 As jy wil bedank, gee darem minstens een maand kennis.
4 Die winkel het 'n nuwe klerereeks gekry vir die somer.
5 Slaap jou hond?
6 Hoeveel kamers het jou nuwe huis?
7 Het jy al ooit daaraan gedink om jou ou boeke te begin verkoop?
8 Trek jou jas aan.
9 Dit lyk asof al my vriende slegte verkoues kry.
10 Dit voel of ons baie jare laas 'n warm somer gehad het.
11 Hy koop nog aandele in die maatskappy.
12 Het jy geweet dat hulle ekstra kaartjies vir die konsert gekry het?

Set 7

1 Skakel asseblief die ketel aan.
2 Jy moet jou kinders uit hulle eie foute laat leer.
3 As mens by die lughawe werk, kry jy goeie reisvoordele.
4 Knope is duur.
5 Hoeveel katte en honde het jy?
6 Het jy verlede jaar ook oorweeg om 'n kamer aan te bou?
7 Skop die bal ver.
8 Ry versigtig oor die brug as die wind waai.
9 Wanneer kinders begin skool gaan, is hulle geneig om makliker siek te raak.
10 Ek eet gans te veel oor naweke.
11 Watter tipe spaarplan het hulle gesê sal die beste vir jou wees?
12 Was jy gister by daardie sanger se konsert?

Set 8

1 Was die groente en vrugte af voor jy dit eet.
2 My broer en sy twee kinders kom môre by my kuier.
3 Hy werk graag.
4 Ek moet 'n nuwe uitrusting koop.
5 Het jy toe jy verlede week gaan stap het enige mooi voëls gesien?
6 Maak die kombuis skoon.
7 Moenie probeer hardloop tensy jy ordentlike skoene het nie.
8 Ek neem altyd my sonbril saam wanneer ek op sonskyndae bestuur.
9 Jy sal meer gesond moet begin eet.
10 Wil jy hierdie jaar aan die einde van die somer verlof neem?
11 Doen jy elke maand jou begroting so deeglik?
12 Hou jy van kitaar speel?

Set 9

1 Daardie restaurant se nagereg is baie lekker, maar gans te duur.
2 Sy's my dogter.
3 Hy hou van sy nuwe werkgewer.
4 Ek moet regtig môre my broek en hemde by die droogskoonmaker gaan haal.
5 Luister na die voëls.
6 Skuif die stoele na die muur oorkant die venster.
7 Sê vir die span dat vanaand se oefening weens reën gekanselleer is.
8 Trek warm aan, want dis koud buite.
9 Gaan jy na 'n gimnasium toe of oefen jy by die huis?
10 Sal jy vanaand saam met my gaan fliek?
11 Kan ek asseblief geld leen?
12 Dink jy hy sal klavier speel by môre se partytjie?

Set 10

1 Slaai is lekker.
2 My broer het 'n nuwe meisie.
3 Hy het bevordering gekry en nou wil hy glad nie meer weggaan nie.
4 Is die rok nuut?
5 Moet jy so naby aan die leeuhok staan?
6 Onthou om al die deure en vensters te sluit wanneer jy uitgaan.
7 Neem jou kamera saam na die wedstryd toe.
8 Al die stad se paaie was gister heeldag gesluit weens swaar oorstromings.
9 Ken jy 'n goeie dokter hier rond?
10 Waarheen gaan jy vir Kersfees?
11 Hoeveel rente betaal jy elke maand op jou huislening?
12 Kom ons luister na die nuuste album van daardie bekende orkes.

Set 11

1 Sy het 'n vars slaai gemaak.
2 Hy deel 'n tweeslaapkamerwoonstel in die stad met sy twee broers.
3 Is jou werk interessant?
4 Is dit moeilik om jou broek kleiner te maak?
5 Vra jou bure om jou hond te versorg terwyl jy weg is.
6 Ruim die huis op voor die naweek.
7 Dit verg jare se oefening om 'n goeie sokkerspeler te word.
8 As die temperatuur daal, is dit koud buite.
9 Watter medisyne moet jy drink?
10 Wil jy vandag saam met my gaan Kersinkopies doen?
11 Jy behoort elke maand meer van jou salaris te kan spaar.
12 Speel daardie liedjie.

Set 12

1 Ons het gister weer gebraaide lam, wortels en kool vir ons middagete gehad.
2 Waar is jou seun?
3 Hoe lank gaan jy nog vir hierdie maatskappy werk?
4 Het jy geweet sy het haar eie rok vir die troue gemaak?
5 Moenie die hond se leiband afhaal nie.
6 Ons het besluit om die kombuis en die eetkamer blou te verf.
7 Hy speel Sondae saam met sy vriende sokker.
8 Uiteindelik skyn die son weer.
9 Het hy naar gevoel omdat hy slegte kos geëet het?
10 Hou jou diere veilig in die huis as julle vuurwerke skiet.
11 Vat jou beursie.
12 Sit jou kitaar op sy plek.

Set 13

1 Waar het jy sulke lekker brood leer bak?
2 Wanneer en waar trou jou tannie?
3 Het jy hierdie week enige oortyd gewerk?
4 Waar's my baadjie?
5 Al die diere is gevoer en in hulle hokke gesit.
6 Hulle het 'n huis gekoop.
7 Toe sy op skool was, het sy in die eerste hokkiespan gespeel.
8 Ek het gehoor dat dit teen Sondag gaan sneeu.
9 Gaan spreek jou dokter.
10 Onthou om betyds vir jou ouma 'n spesiale verjaardagkaartjie te koop.
11 Sy moet nege maande lank minstens honderd rand in daardie rekening hou.
12 Doen aansoek vir die koor, want jy het 'n mooi stem.

Set 14

1 Wat wil jy vir ontbyt hê?
2 Gaan jy nog later jou ouma bel?
3 Waar is jou kantoor?
4 Onthou om bleikmiddel te gebruik wanneer jy daardie hemde was.
5 Honde kan baie stout wees.
6 Ons soek, maar dis moeilik om 'n huis te kry wat bekostigbaar is.
7 Ek hou daarvan om netbalwedstryde by te woon.
8 Reën dit al?
9 Jy moet minder kos met baie vet eet en meer vesel inneem.
10 As jy klavier wil leer speel, moet jy weekliks lesse probeer neem.
11 Probeer om elke week twintig rand van jou salaris te spaar.
12 Die kaartjies vir die konsert is reeds uitverkoop.

Set 15

1 Kook jy elke aand sulke groot etes?
2 Dans jou suster?
3 Sê vir jou baas dat jy hom graag sal help.
4 Stryk al jou langmouhemde.
5 Dit was maklik om vir die papegaai 'n paar woorde te leer.
6 Ons soek 'n losstaande huis in 'n stil buurt.
7 Hou jy van kamp?
8 Lyk dit asof die son binne die volgende paar uur gaan skyn?
9 Onthou om jou medisyne te drink en baie vloeistof in te neem.
10 Moenie weer jou Kersinkopies tot die laaste oomblik uitstel nie
11 Ek betaal elke maand my klererekeninge betyds.
12 Ek hou van sy nuwe liedjie.

Set 16

1. Is jy honger?
2. Probeer om jou ma een keer per week te besoek.
3. Maak die werk vandag klaar.
4. Gee jou ongebruikte klere aan mense wat dit dalk nodiger as jy het.
5. Gedurende die dag hou my twee katte mekaar geselskap.
6. Waar woon jy nou?
7. Kon jy darem verlede week 'n deel van die skaatskompetisie bywoon?
8. Gaan dit al weer die hele week reën of word sonskyn voorspel?
9. Jy moet oefen en 'n dieet volg om gewig te verloor.
10. Daar is altyd 'n groot optog met Nuwejaar.
11. Ek dink ek het genoeg kleingeld.
12. Die sanger en die orkes was goed.

Set 17

1. Maak eers 'n lys voor jy die inkopies gaan doen.
2. Onthou jou pa se verjaarsdag.
3. Jy moet al jou strokies hou wanneer jy op 'n besigheidsreis gaan.
4. Koop vir jou 'n nuwe rok vir die ete.
5. Waar is die hondjie?
6. Verkies jy om in 'n stad of in 'n dorpie te woon?
7. Het jy toe jy op universiteit was, enige bekers vir kompetisies gewen?
8. Sal jy strand toe gaan selfs al is dit nog bewolk?
9. My allergieë raak altyd erger in die somer.
10. Ek geniet altyd die somervakansies.
11. Sy het al weer 'n lening aangegaan.
12. Dis my kitaar.

Set 18

1. Bêre die brood en botter.
2. Vra jou pa of hy dalk enigiets van die winkel af nodig het.
3. Probeer ontspan voor jy môre vir jou onderhoud gaan.
4. Dis my nuwe rok.
5. Het jy baie leeus en luiperds gesien toe jy op safari was?
6. Gaan jy iemand vra om die woonstel te verf voor jy intrek?
7. Wil jy later saam met ons die krieket wedstryd gaan kyk?
8. Maak die vensters toe ingeval dit begin reën.
9. Rook is sleg vir jou longe.
10. Ons gaan eers volgende week see toe.
11. Neem vyf rand.
12. Watter tipe musiek beplan jy vir die troue se onthaal?

Set 19

1 Haal die aartappels uit die lou-oond en sit dit op die tafel.
2 Onthou om jou suster vanaand lughawe toe te neem.
3 Ek het gister bedank.
4 Ek het 'n baadjie wat pragtig sal lyk by daardie swart romp.
5 Kon jy darem die olifante sien toe jy by die dieretuin was?
6 Waar het jy al die nuwe meubels vir jou woonstel gekoop?
7 Sit al jou nuwe gholfstokke in die sak.
8 Moenie vir die donderweer skrik nie.
9 Ek was nog altyd bang vir inspuitings.
10 Dit reën aanhoudend.
11 Het jy genoeg geld om die reisigerstjeks te koop?
12 Het jy hom hoor sing?

1 **Set 20**

1 Voorverhit eers die oond voor jy die pastei insit.
2 My suster is swanger.
3 Dit vat my soggens twee ure om by die werk te kom.
4 Daardie pak is baie duur, maar die styl is perfek vir jou.
5 Is dit baie moeilik om vir al daardie visse te sorg?
6 Maak die kamer netjies voordat jy televisie kyk.
7 Sê vir my wie gewen het.
8 Neem 'n trui ingeval dit koel word.
9 Ek voel siekerig.
10 Neem jy weer hierdie somer die kinders saam strandhuis toe?
11 Hoeveel is die totale bedrag?
12 Sal jy miskien vanaand tyd hê om die nuwe fliek te gaan kyk?

Set 21

1 Daardie pizza ruik lekker.
2 Dis moeilik om te glo dat sy vier dogters en twee seuns het.
3 Ek sukkel altyd Vrydae om vroeg by die werk weg te kom.
4 Dis aanvaarbaar om informeel aan te trek vir Saterdag se partytjie.
5 Voer die eendjies 'n stukkie van jou brood.
6 Sit die sleutel onder die matjie.
7 Jy moenie so alleen gaan draf nie.
8 Vat 'n sambreel.
9 Gaan jy gereeld vir 'n volledige ondersoek by jou dokter?
10 Kan jy 'n sneeuman bou?
11 Hoe lank neem dit om genoeg geld te spaar vir 'n goeie voertuig?
12 Wil jy môre na 'n musiekkonsert toe gaan?

Set 22

1 Ek was vanoggend so honger dat ek 'n groot ontbyt geëet het.
2 Ons hou elke somer 'n ete saam met my man se familie.
3 Ek is moeg, want ek moet baie reis vir my werk.
4 Het jy jou nuwe skoene by daardie winkel gekoop?
5 Gaan stap 'n entjie met jou hond.
6 Politoer die vloer in die sitkamer.
7 Voltooi die wedloop.
8 Die dae is koel vir hierdie tyd van die jaar.
9 Hy sy al waterpokkies gehad?
10 Het jy al ooit daaraan gedink om van jou stories te laat publiseer?
11 Kan ek moontlik weer 'n banklening kry?
12 Luister na die musiek.

Set 23

1 Ons gaan graag stad toe, want die winkels daar verkoop interessante kosse.
2 My oupa se huis is net 'n entjie van hier af.
3 Moet jy baie op reis gaan vir werk?
4 Kan ek jou rooi trui leen?
5 Jou kat se naels is baie skerp.
6 Sluit die deur.
7 Koop vir ons seisoenkaartjies vir hierdie jaar se rugby.
8 Dis vanaand baie warm buite.
9 Voel jy beter vandat jy begin reg eet het en elke dag oefen?
10 Gaan julle in die winter of somer met vakansie?
11 Hou al die kleingeld.
12 Sit asseblief die musiek sagter voor die bure weer daaroor begin kla.

Set 24

1 Ons gaan gewoonlik een keer per week by 'n restaurant eet.
2 Gaan jy môre weer vir jou ouers kuier?
3 Het julle van sy bevordering gehoor?
4 Waar het jy daardie mooi baadjie gekoop?
5 Voer die diere.
6 Ons het al die houtvloere laat skoonmaak en poleer.
7 Hulle hou van diepsee hengel.
8 Daar is 'n droogte, want dit het die afgelope maande nie gereën nie.
9 Het jou been lank geneem om gesond te word?
10 Pos jou kaartjies betyds.
11 Jy moet jou bonus nou belê terwyl die rentekoers nog hoog is.
12 Koop daardie nuwe klavier terwyl die winkel so 'n groot uitverkoping het.

Set 25

1 Wat sal ons vir ete maak wanneer ons bure kom kuier?
2 Is jou broer op skool?
3 Kry julle elke jaar 'n bonus by die werk?
4 Het jou ma daardie mooi lang rok spesiaal vir die ete laat maak?
5 Daardie slang is lank.
6 Hulle het ‘n mooi mat gekoop.
7 My broer versamel lankal graag munte en seëls.
8 Laasnag het die donderweer en blitse ons almal tot laat wakker gehou.
9 Drink jou pille.
10 Onthou om jou swemhanddoek in jou sak te sit.
11 Moenie so baie aan klere spandeer nie.
12 Kry kaartjies vir volgende week se konsert voordat dit alles uitverkoop is.

Set 26

1 Het jy 'n broodresep?
2 Gaan jy môre na jou oom se verjaardag toe?
3 Gaan jou man verlof neem in Maart vir die kinders se skoolvakansie?
4 Koop daardie drafskoene.
5 Die hond blaf vandag baie hard.
6 Volgende lente moet die tuin heeltemal oorgedoen word.
7 Hy kon nie aan die wedloop deelneem nie, want hy was siek.
8 Is dit warm?
9 Moenie vergeet om elke aand jou tande te borsel nie.
10 Laat die kinders wakker bly vir Nuwejaar.
11 Sit jou geld in 'n spesiale rekening wat meer rente sal verdien.
12 Hy speel elke Maandagaand die trompet in 'n blaasorkes.

Set 27

1 Hoe lank neem dit om 'n hoender te braai?
2 Beplan al hulle kinders om volgende week na die reünie toe te kom?
3 Moenie laat kom nie.
4 Vat 'n trui, want dis koud.
5 n Paar klein duifies sit op die draad.
6 Ons wil nuwe rame in ons vensters sit, maar dis te duur.
7 Wie het gewen?
8 Het die weerburo sonskyn of reën vir môre-middag voorspel?
9 Drink jou vitamienes elke oggend ná ontbyt.
10 Maak seker dat al die meisies rokke vir die skoolkonsert het.
11 Mense wat nou in aandele en goud belê, loop 'n risiko.
12 Daardie klavier is baie oud.

APPENDIX 10

TEST ADMINISTRATION OF THE SAV-L

1. Purpose of the test

The main need established for which the SAV-L was developed in the first instance, was to assess and monitor cochlear implant candidates. The purpose of the SAV-L is therefore firstly to compare the performance scores of an individual with a hearing loss with the scores obtained from individuals with normal hearing in order to make a management decision. Data so far suggests that an individual with normal hearing should score 91.97% or higher at 50dBHL with a +5dB SNR on any one of die SAV-L sentence lists. Any score less than this score should be seen as resulting from a challenge to the participant's speech perception abilities. However, in order for the SAV-L to be used to establish auditory device candidacy in the population it is intended for, it needs to be further evaluated, to establish its sensitivity and to develop normative data.

The test is to be used as part of a battery of tests, which includes assessment of peripheral auditory status as well as a case history and formal or informal determination of the participant's subjective disability as a result of the hearing loss.

The test can also be used to compare a participant's performance over time. This enables monitoring of adaptation to a certain auditory device or can be used to compare different devices.

2. Population of interest

In view of the linguistic load of the SAV-L, the test is considered suitable for adults with good language skills. This population would include post-linguistically deafened adults or adults with hearing loss that had successful early intervention for spoken language to develop fully. It is not intended for persons with poor language skills, since a baseline effect will rapidly be seen and negatively influence outcome decisions.

3. General set-up

Testing must only take place using the recorded test material as a means of controlling the test conditions, and not be presented through live voice. The prepared sentence lists were standardised in noise only, and should therefore only be presented in noise. A separate development of the sentences in quiet (to be called the SAV-S, i.e. *Sinslyste in Afrikaans vir Volwassenes in Stilte* [Sentence lists in Afrikaans for Adults in Quiet] (SAV-S) remains a future goal. The SAV-L test material is available in compact disc (CD) format. Version 1 contains a 1000Hz calibration tone of 60 seconds duration as Track 1, together with sentence lists 1 to 27.

A two-channel audiometer and a high quality domestic CD-player containing a headphone socket are needed. In addition, a Y cable with a suitably sized jack to plug into the headphone socket of the CD player and phono (RCA) plugs to plug into the audiometer (see figure A9.1) is required.

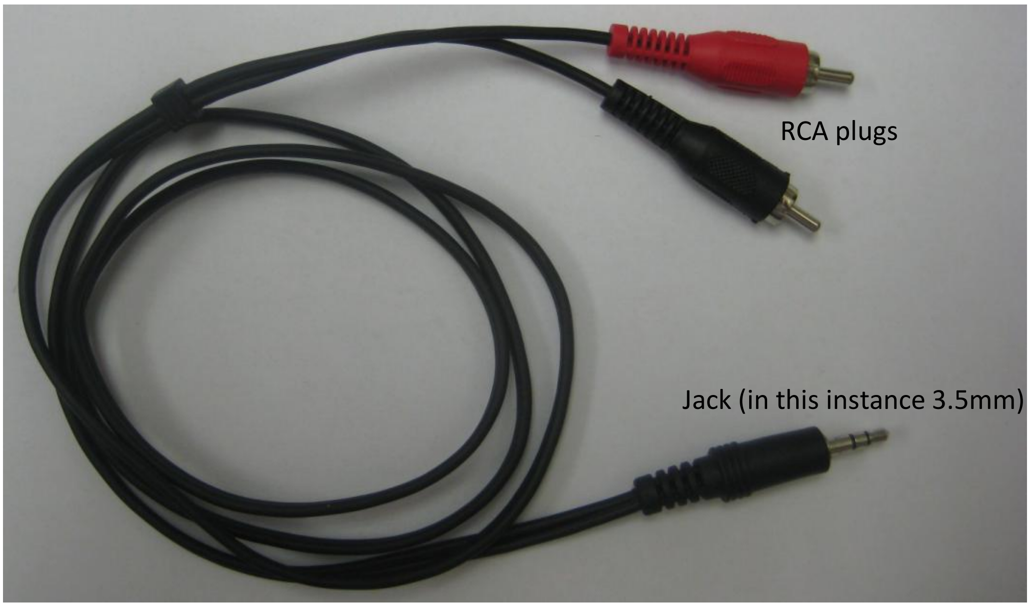


Fig. A10.1: Cable required for test set-up

The jack of the Y cable is plugged into the headphone socket of the CD player and the red and black phono plugs are plugged into the external inputs of the audiometer. The red phono plug is plugged into the line A (sometimes called line 1) level input on the audiometer's jack panel, while the black phono plug is plugged into the line B (sometimes called 2) level input.

On channel 1 of the audiometer 'External B' as the stimulus input should be selected. This channel contains the stimulus sentences. On channel 2 'External A' should be selected, which contains the babble. The appropriate transducer and ear(s) for testing should be selected as required per local protocol.

Testing was standardised and therefore must be undertaken in a well-lit sound-treated room. The tester and participant should be able to see each other. This allows for any unclear responses to be clarified by the tester through lip-reading of the participant's speech, and also to give positive encouragement through a nod or smile from the tester, to ensure continued high motivation levels.

4. Calibration

Prior to testing, the volume control of the CD player needs to be set at approximately its middle range. The 1000Hz calibration tone on track 1 should be played and the VU meter on both channels of the audiometer be adjusted to be on 0dB. A subjective listening check is also recommended to ensure no distortion is heard, which can occur if the volume dial on the input CD player is set too high.

5. Presentation mode

Test conditions can be adapted according to the information required. Monaural versus binaural test conditions, unaided versus aided conditions and different transducers, including supra-aural and insert earphones, as well as loudspeakers can be used, but the conditions need to be clearly specified on the test form. This will ensure accurate comparison of scores over time, whether it is inter- or intra-participant comparison. Although international standards (SANS 8253-3:1996) for loudspeaker placement are available, a consensus between clinics has yet to be established with regards to number of loudspeakers used. Arrangements span from one loudspeaker for both the speech and the noise to up to five loudspeakers (Arlinger & Hagerman, 1997, p 278), and loudspeaker set-up should therefore also be specified on the test form.

When testing in the soundfield, it is appropriate to select 'Right and Left' on both channels of the audiometer unless directionality (BKB-SIN manual, n.d., p. 9) or FM advantage (Madell, 2010) is assessed, in which case presentation of the stimulus and masking from different loudspeakers would be more suitable.

6. **Presentation level**

It must be kept in mind that the variable sentence lengths of the SAV-L results in varying levels of difficulties for the individual sentences. The test can therefore not be used as an adaptive test, but only as a fixed speech-in-noise test, as it was intended to. The proposed test level of 50dBHL is one that is representative everyday listening. A SNR of +5dB in the presence of babble at 45dBHL is proposed, however, the test level can be adjusted as required.

7. **Test order**

A practice list must be presented prior to testing, in order to ensure maximum performance. If a participant finds the task difficult to complete, the SNR of parts of the practice list can be increased, in order to build the participant's confidence. This practice list's score should not be recorded as an absolute result. A second and third list can also be presented as further practice lists, should the examiner feel it necessary, since a number of three practice lists is the amount of lists found to be necessary before performance stabilises (MacLeod & Summerfield, 1990, p. 40). The list numbers should be indicated on the test form, as these can be used again as practice lists in subsequent administrations of the test. This will allow for the majority of the test lists to remain unfamiliar to the participant, so that learning effects do not affect performance. The lists can be presented in any particular order.

During testing, a complete list of 12 sentences must be presented at any one SNR. No stimulus must be repeated, as this can affect scores significantly.

8. Test instructions

"Jy/u gaan stemme hoor wat saamgesels en dan gaan 'n manlike spreker tussen hulle sinne van verskillende lengtes sê. Die man se stem waarna jy/u moet luister, is dieselfde stem as wat aan die begin van die sinlys byvoorbeeld 'Sinslys 7' aankondig. Herhaal asseblief elke sin sodra dit gesê is. Selfs as jy/u nie die hele sin gehoor het nie, herhaal asseblief soveel as moontlik wat jy/u dink gesê is, selfs al is dit net 'n deel van 'n woord. Jy/u is welkom om te raai". This translates as, "You will hear voices talking together and then a male speaker amongst them saying sentences of varying lengths. The man whose voice you need to listen to, is the same voice that announces for instance 'Sentence list 7' at the beginning of each sentence list. Please repeat each sentence as soon as it is said. Even if you do not hear the whole sentence, please repeat as much of the sentence as you think you heard, even when it is just part of a word. You are welcome to guess".

9. Test forms

The same test forms can be used for the SAV-L and SAV-S, providing the test conditions are indicated on the test form. An example of a test form can be found in Appendix 12. In order to compare scores over time or when different amplification strategies are being compared, test conditions need to be replicable. To therefore ensure good record-keeping of test conditions, the options for test conditions are indicated on the test forms and need to be completed in full.

During the test's development, it was found that inter-rater reliability increased after indicating the scoring units on the test forms by underlining the first letter of each scoring unit. To ensure accurate recording of test results, it is therefore recommended that the test forms be used during testing.

10. Scoring

The SAV-L was developed to be scored at syllable level. In instances where a word has two syllables that are often pronounced as one syllable, the syllables were marked on the scoring sheets according to pronunciation. An example of this is the word *dae* (*days*) which has two syllables *da* and *e* but is pronounced as /da:/ and was therefore marked as one scoring unit. It is therefore important to carefully consider the units identified on the test forms during scoring.

When it is clear that the test participant has heard the word correctly but pronounced it differently due to dialectal differences, the participant should be given full credit for the word. An example of this would be the word *watter* (*which)*, which is sometimes pronounced as *watse* and should be given its full credit of two points.

11. Interpretation

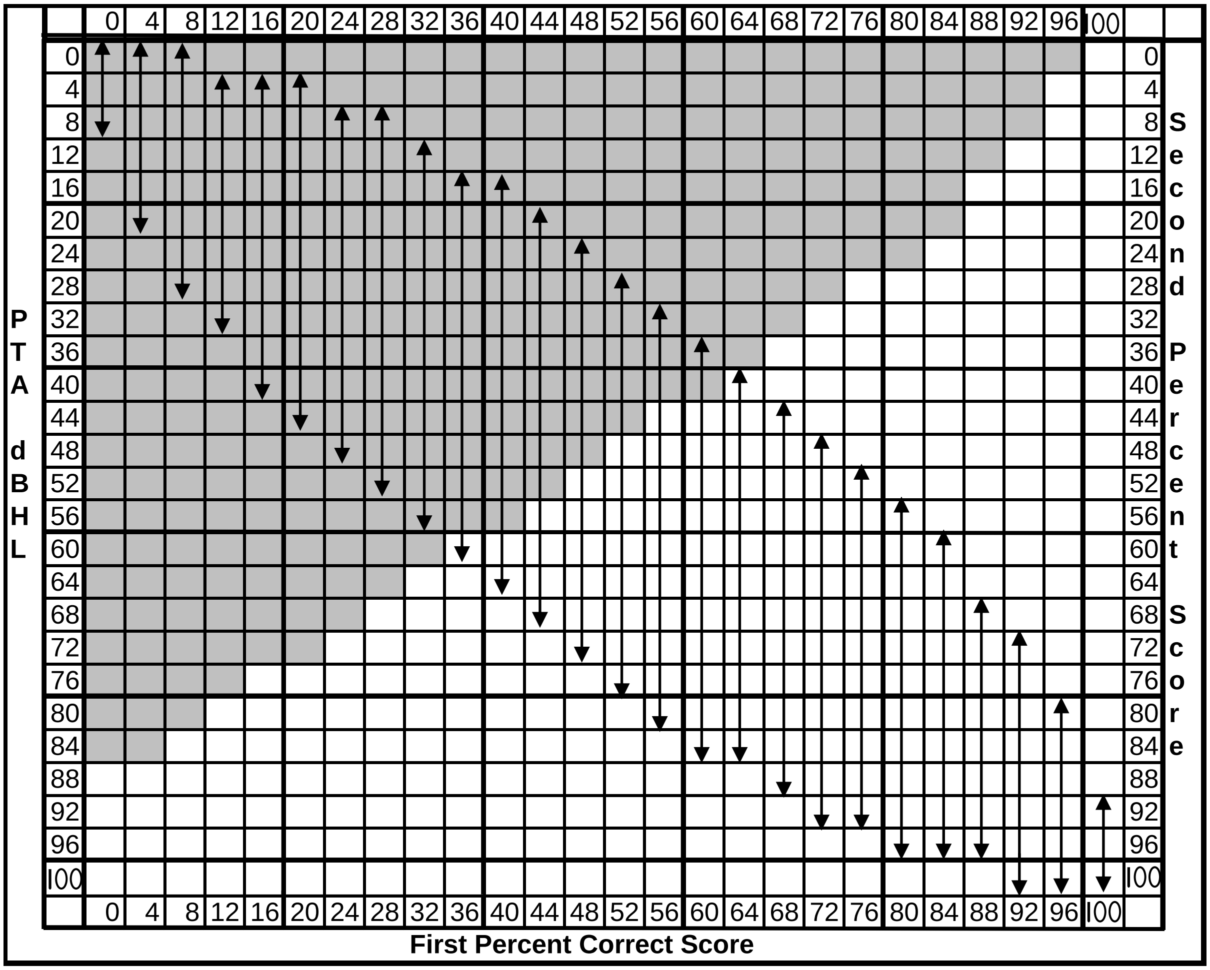
Currently up to 13.87% variability has been found to be present amongst the averages of the sentence lists of the SAV-L. When a participant shows an improvement of this amount or more, a positive improvement in scores can be said to be present.

Because the performance of hearing impaired persons have not been established with this test yet, existing cochlear implant criteria, such as that of the FDA (Kirk & Choi, 2009, p. 192) and commercial cochlear implant providers (Zwolan, 2009, p. 923) should be interpreted with caution. However, since the SAV-L is based on the CUNY topic-related sentences that are already frequently used as sentence material (Zwolan, 2009, p. 922), the use of the SAV-L in parallel with other sentence tests is encouraged to obtain data of its performance.

APPENDIX 11

SPRINT CHART

SPRINT CHART for 25-WORD LISTS[1]



95% Confidence Limit for PBmax on NU6 25-word list.Plot score according to PTA on left ordinate and percent correct score on the abscissa. If it falls in the shaded area, it is considered disproportionately low. (Adapted from Dubno et al.,1995)

95% Critical differences for 25-word list. Plot first and second score according to the abscissa and right ordinate. If it falls within the arrow, the two scores are not significantly different (Adapted from Thornton & Raffin, 1978)

1. Note - From www.utdallas.edu/~thib. Copyright 2010 by L.M. Thibodeau. Reprinted with permission

APPENDIX 12

EXAMPLE OF A SAV-L & SAV-S TEST FORM

SAV-S & SAV-L TOETSVORM

Naam: ______________ Geb.: ______________ ID: ______________

Toetsdatum: __________ Toetskondisie: Stilte / Lawaai Toetslysnommer(s): __________

Tranducer ☐ Klankveld: Aantal luidsprekers _____ Stimulus ° _____ Maskering ° _____

☐ Oorfone: Supra-ouraal / Insteek Monouraal R / Monouraal L / Binouraal

Toetsvlak Stimulusvlak (dBGP) __________ Maskeringsvlak (dBGP) __________

Versterking R: Onversterk / Versterk L: Onversterk / Versterk

Detail: ______________ Detail: ______________

______________ ______________

SINSLYS 1 (CD 1, SNIT 2)

		Telling
1	Het jy al geëet?	/5
2	Hoeveel van jou broers bly nog by jou ouers?	/11
3	Hoeveel jaar se opleiding het dit geneem om 'n verpleegster te word?	/18
4	Waar kan ek my klere laat was?	/8
5	Katte is maklike troeteldiere, want hulle kom en gaan soos hulle wil.	/20
6	Sy het in 'n drieslaapkamerwoonstel ingetrek.	/13
7	Ek wil graag tennis speel.	/6
8	Ek hoop dit reën, want my tuin is redelik droog.	/12
9	Doen wat die dokter gesê het, anders gaan jy sieker word.	/15
10	Sny die koek.	/3
11	Betaal die tjek voor môre in.	/8
12	Onthou asseblief om vir 'n uur te oefen voor jou volgende klavierles.	/20
	TOTAAL	**/139**
	=	**%**