ADAPTATION & VALIDATION OF THE NAMES® SPEECH AUDIOMETRY TEST INTO DUTCH

A study submitted in partial fulfilment of the requirements
for the degree of Master of Science of the
University of Hertfordshire

F.A.M. Marinus B.Sc.
Student nr. 16037002

Mary Hare, Newbury
Partnered with University of Hertfordshire, Hatfield

MAY 2019

Table of Contents

List of Tables

List of Figures

Abbreviations

2AFC	Two Alternative Forced Choice
AAST	Adaptive Auditory Speech Test
APD	Auditory Processing Disorder
APE	Auditory Phoneme Evaluation
BELLS	Battery for the Evaluation of Language and Listening Skills
CI	Cochlear Implant
CSV	Comma Separated Value
CVCVC	Consonant Vowel Consonant Vowel Consonant
HI	Hearing Impaired
H-LAD	Heidelberger Laut Differenzierungs test
IPA	International Phonetic Alphabet
IQR	Inter Quartile Range
L_{EQ}	Equivalent continuous sound level
LTASS	Long Term Average Speech Spectrum
MAUS	Münchner AUditiver Screeningtest für Verarbeitungs- und Wahrnehmungsstörungen
mFAST	multi Frequency Animal Sound Test
NAMES®	Name-based Auditory Multilingual Evaluation of Speech
NWRT	Nonsense Word Repetition Test
PB	Phonemically (Phonetically) Balanced
PCM	Pulse Code Modulation
PIT	Phoneme Identification Test
PRS	Phoneme Recognition Score
PTA	Pure Tone Audiometry
RMS	Root Mean Square (effective value)
SL	Sensation Level
SPL	Sound Pressure Level
SPSS	Statistical Package for the Social Sciences
SRT	Speech Recognition Threshold
SD	Standard Deviation

WHO	World Health Organisation
WRS	Word Recognition Score

Acknowledgements

Herewith I want to thank everyone who supported me with this study. First and foremost, I have to thank my two research supervisors/advisors, Dr. Joy Rosenberg and Prof. Dr. Ir. Frans Coninx, who guided me in this research. The never-ending ideas, discussions and critical questioning of Prof. Coninx built on my research skills. He has inspired me since the early 1980's, when we first met.

I thank my colleagues for their patience and support. It was not always easy to combine this study with my job and international travels, but their support and compassion was heart-warming.

Last but not least, I have to thank my family for their encouragement.

Special thanks go to my mother, for her care and moral support and to my wife Pauline, with whom I was able to share thoughts and who helped me tremendously with her corrections on my English.

In dedication to my father.

'Es hört doch jeder nur, was er versteht.'
'Everybody only hears what he understands'
J.W. von Goethe

Abstract

In this research a Dutch version of the NAMES®[1] nonsense word phoneme recognition test was developed and tested. Nonsense words contain minimal reference to language information or other syntactic queues. This makes such a test very sensitive and less dependent of prior linguistic knowledge. Phonotactic rules were used to let the words sound like real words. These appear to influence on-line language processes to segment words from fluent speech. The NAMES® test was originally developed as a tool to use for hearing aid validation and to support speech therapists with diagnosis of their treatment of phonological development. This test was designed so that it also can be used for children. To prevent influences of phonological and phonetic development factors in children, on the evaluation of this test design it was decided not yet to include children in this research. However, factors related to optimizing the test for children were investigated.

A group of 57 adult participants across different age groups with diverse hearing acuity were involved in a first validation of this test. Results indicate that the NAMES® test is performing well, seems to be sensitive and delivers the expected results. Recommendations were made for improvement of the test and suggestions were done for further research.

The NAMES® test is developed for the BELLS® software platform. More information on this can found in the 'Methods' section.

[1] NAMES®: Name-based Auditory Multilingual Evaluation of Speech. NAMES® as well as BELLS®: "Battery for the Evaluation of Language and Listening Skills" are registered trademarks of the Institute for Audiopedagogics, Solingen Germany.

1 INTRODUCTION

Society has become increasingly multicultural and multilingual. Not only in the West, but also in low and middle income countries, where many young people migrate from the rural to urban areas (UNDESA, 2014). In my work in international projects, I often come across challenges of testing multilingual children. It is therefore my personal motivation to explore the process of adapting the NAMES® speech test. The acronym NAMES® stands for "Name-based Auditory Multilingual Evaluation of Speech". NAMES® is a supra-threshold speech screening test, originally developed in the German language by Prof. Coninx from the Institute for Audio Pedagogics (IfAP) in Solingen Germany. In this research, I will investigate the adaptation of NAMES® to my mother tongue, Dutch. This research should provide an insight in adapting this tool to other languages. I have chosen to include only adult participants, to rule out effects of phonological development in this stage of the research. Validation for children will be through follow-up research.

NAMES® is based on CVCVC (C=Consonant, V=Vowel) nonsense words, which are independent of the individual's literacy and education (Cooke *et al.*, 2010) and of the listener's cognition (Akeroyd, 2008). This includes short-term memory and speech processing, which are considered to be the causes of deterioration in speech recognition, particularly in older listeners (Gordon-Salant, 2005). The focus of the NAMES® test is to measure the phonemic identification and differentiation, above an individual's threshold (Nguyen, 2017). A nonsense word speech test is suitable for non-native listeners who have little experience of the language being tested (Paglialonga, Tognola and Grandori, 2014).

Through a literature review, relevant linguistic, phonetic and design parameters for developing a Dutch version of the test were identified. The NAMES® test was compared with existing associated speech tests. In the 'Methods' section, the research design is explained. It presents how the identified parameters were incorporated, and how the words were generated, selected and recorded. The

‘Results and Analysis’ section presents the test data from different groups and outcomes are discussed. The ‘Discussion’ section provides critical notes on the research and recommendations are made for optimization and future research.

2 LITERATURE REVIEW

2.1 Introduction

This section deals with an examination of the existing research in academic literature on speech audiometry and phoneme recognition for children. Nonsense words are used to minimize the effects of the knowledge of language, vocabulary and memory on the test. This review is used to identify parameters and relevant factors for constructing a Dutch version of the NAMES® test and it will identify how this test relates to other existing tests. This literature review is vital in establishing a sound foundation for this research.

2.2 Literature search strategy

A literature review was done with the search terms: “Phonemic Distribution”, “Phoneme frequency of occurrence”, “Nonsense word test”, “Non-Word Repetition Test”, “Phonotactic rules”, “Speech audiometry children” and the Dutch variants of these words. This list was not exhaustive.
Consecutively these terms were used for a search within the electronic library repositories of the University of Hertfordshire. This search included the following international scientific databases, such as: PLOS ONE, PubMed, Google Scholar, SCOPUS, EBSCO host, as well as Studynet, the online library of the University of Hertfordshire. ResearchGate proved to be a good source of information for articles which were not accessible through Studynet. Through the ResearchGate website it is possible to contact authors directly and the site gives recommendations for articles depending on your prior search terms. Furthermore, the bibliography of some of the scientific articles gave useful leads to other articles. Specialized books on Speech Audiometry from Lawson and Peterson (2011) and Martin (1997) provided valuable information on the fundamentals of speech audiometry.

2.3 Definition Speech Audiometry

It is a commonly accepted practice that Speech Audiometry complements pure tone audiometry in differential diagnosis. Speech audiometry is a benchmark of the client’s auditory capacity, because according to Lawson and Peterson (2011) it can tell us how well one hears at different levels and it can give information on how those with hearing disorders tend to respond to a variety of basic measures. The NAMES® test is mainly designed as a speech screening tool for children with a low language level. Meister (2005) observed some aspects for speech audiometry tests for children. They must be age appropriate, fast and efficient. The duration of speech audiometric tests with children are restrained by fatigue. The test materials should match the child's ability of speech perception, use age-specific vocabulary, and consider the child's phonological development. For high objectivity and reliability, it is essential to provide all children with identical instructions. Nonsense words such as used in the NAMES® test overcome the problem of testing at a specific language level.

Meister (2005), in addition, claims that the test should be phonemically balanced. Phonemic balancing can be realized when the different phonemes appear in the test material with the same relative frequency as in every day speech. In the English language, lists with a frequency of occurrence of consonants and vowels from Frye (1947) and Denes (1963) can be found in "Speech Audiometry" (Martin, 1997) pg. 45-46. Dutch phoneme distribution lists are available at the website: http://taalportaal.org, and can be found in tables 2 and 3. A list with the Frequency of Occurrence of phonemes does not exist in every language. It is also possible to phonetically describe parts of texts from newspapers, books or transcripts of radio or television programs. As long as they consist of the contemporary spoken version of that language. The phonemes should be described using the International Phonetic Alphabet (IPA) (Appendix 1) and then afterwards their distribution can be counted.

Lyregaard (1997) notes that the consonants in most cases are fairly defined, but vowels give rise to considerable disagreements, which are to some extent related to dialectical differences. In addition, some phoneme clusters can lead to arguments, because they do not appear in isolation and the question is whether they should be regarded as a single phoneme. For example, /ltʃ/ and /dʒ/. The same applies for diphthongs, such as /ɛi/, /ɑu/ and /œy/. According to Lyregaard (1997) phonemes do not occur as individual units, but in an articulatory or acoustic stream, linked together in such a way that they interact, mainly due to the limitations of the articulatory musculature.

In the Dutch language there are a few lists with the frequency of occurrence of phonemes available: Zuidema (2009) and Oostendorp (2018). This research uses the most recent list from Oostendorp (2018) which will be discussed further in the "Methods" section. The Dutch Language has 34 phonemes, which consist of 18 consonants and 16 vowels.

2.4 Phonotactic rules

The NAMES® test should follow the syntactical and phonotactical rules of the language, in which it is developed. The term phonotactic probability has been used to refer to the frequency with which legal phonological segments and sequences of segments occur in a given language (Jusczyk, Luce and Charles-Luce, 1994). Sensitivity to phonotactic information already occurs very early in life. According to research by Jusczyk (1993), by 9 months of age infants were able to discriminate among sounds that were and were not part of their native language. Jusczyk et al. (1994), also demonstrated that 9 months old infants could discriminate between nonsense words that contained sounds that were more common or less common in their native language. The phonotactic probability impacts how rapidly and accurately adults with normal hearing repeat real words and nonsense words (Vitevitch and Luce, 2005). Phonotactic probability appears to influence several on-line language processes and it is one of several cues that enables infants and adults to segment words from fluent speech (Gaygen, 1997; Pitt and McQueen, 1998). Once a word has been segmented from fluent speech, phonotactic probability also influences how quickly children acquire new words (Storkel, 2001; Storkel, 2003), as well as how quickly normal hearing adults and hearing impaired adults who use cochlear implants recognize spoken words (Vitevitch, 2002a; Vitevitch, 2002b). Hearing loss has a direct impact on the accuracy of word recognition because of missing or distorted information that likely interacts with phonotactic probability of a meaningful word. Phonotactic probability influences the production, in addition to the comprehension of spoken language (Dell *et al.*, 2000; Vitevitch, Armbrüster and Chu, 2004).

2.5 Standards for speech audiometry

Although already defined some decades ago by Watson (1957) and more recently confirmed by Bosman (1995) and Meister (2005), the major criteria for valid speech recognition tests for children are:

1. They should be constructed of monosyllables
2. The words should be within the vocabulary range of the child
3. The lists should be phonemically balanced
4. The lists should be equal in difficulty
5. The responses required must not involve a skill which will cause the subject any difficulty or the tester any uncertainty

Watson (1957) recommended the use of monosyllabic words of the CVC type, because contextual clues are relatively absent in such materials. He noted that nonsense syllables made the test too difficult for children. For a phoneme recognition test like NAMES® this might be a less important factor, because the nonsense words are explained to children as being “names” which do not have any meaning.

In Germany, by 1961 word tests had been standardized in German Standard DIN 45621 (1995). This was based on the research work of Hahlbrock (1970) on the “Freiburger wörtertest”. The current norm is the International standard ISO 8253-3 (2012) which specifies basic methods for speech recognition tests for audiological applications. In order to ensure minimum requirements of precision and comparability between different test procedures including speech recognition tests in different languages, the standard specifies requirements for the composition, validation and evaluation of speech test materials, and the realization of speech recognition tests. This ISO norm does not specify the contents of the speech material because of the variety of languages.

Testing speech perception in noise is a more valid procedure to assess hearing in daily life than is testing in quiet, and therefore it is particularly important for the diagnosis of hearing impairment in children. However, for children with severe hearing impairment, with specific language impairment, and for multilingual children, a test for speech perception in quiet may be indicated.

According to Mancini et al. (2010) attention span plays a big role in testing children. Therefore, they recommend a test design that avoids time consuming tasks that could increase fatigue. NAMES® only consists of 20 words which should be repeated. In general, the duration of this test should be ideal to fit in the attention span of young, even 4 to 5-year-old children.

2.6 Recordings of word lists

Martin (1997) describes conditions and recommendations for recording the speech materials. Although the technology for recording and editing has improved a lot since his publication, the basic principles still apply. Recordings should preferably be done in a professional studio, by a professional speaker. The levels of the individual test words should be adjusted to the same level by means of time weighting. This can be done with audio workstation programs such as Cool Edit Pro 2.1 (2003). Calibration signals should be included. Martin (1997) recommends 125 Hz, 1 kHz and 8 kHz and a speech simulating noise (CCITT 1964). More recent specifications are defined in the norm ISO 8253-3 (2012), which specifies that each copy of the speech test besides the speech test material should contain the following signals:

1. A signal for the calibration of the speech audiometer.
2. Signals for testing the frequency response of the speech audiometer, including the playback equipment and the recording.
3. Signals for testing the harmonic distortion of the speech audiometer.

Nguyen (2017), describes in his PhD thesis how his recordings for the Vietnamese NAMES® test were done. He used a 40-year-old female speaker, who was a native speaker of the South Vietnamese Language. She was asked to pronounce the words with a constant intonation, to avoid an 'asking' intonation, and to maintain a reading speed in a natural pronunciation.
The acoustic stimuli were recorded as a mono signal, with sounds digitized at a 44.1 kHz sample rate into a 24 bits digital signal. The recordings took place in a sound treated room with an ambient noise level of around 25 dB_A.
The syllable durations were adjusted to ensure that all stimuli were balanced in terms of energy, the Root Mean Square (RMS) values of each stimulus (first

and second syllable) were measured with Cool Edit pro 2.1 (2003). Each syllable in each disyllable combination was equalized at a similar total RMS level. Time averaged levelling (L_{EQ}) brings the sound energy of the words to the same level. This technique provides a decrease in variance for responses to words (Dermody, Katsch and Mackie, 1983).

2.7 Phoneme analysis

Phonemes can be arranged in different categories, depending on where and how they are produced, as can be seen on the IPA chart in Appendix 1.
The most common categories are:

- Fricatives
- Affricates
- Vowels
- Diphthongs (glides)
- Stops (plosives)
- Aspirates
- Liquids (e.g. rhotics)
- Nasals

When the different phonemes are scored in the NAMES® test, the software automatically calculates a statistical overview of the type of phonemes the client has problems with. This can for example support the Speech and Language Therapist with phoneme awareness training, which supports early reading and spelling skills (Ball and Blachman, 1991). It can also help the audiologist in adjusting assistive devices (Dreschler, 1989).

2.8 Nonsense Word Repetition Tests

Nonsense Word Repetition Tests (NWRT) aim more at language and memory processing and language impairment, but literature gives a lot of useful information for the design and interpretation of the NAMES® test.
NWRT's can complement traditional language tests, because they are less dependent on language knowledge and tap on basic cognitive underpinnings of language such as phonological processing and short-term memory (Chiat *et al.*, 2015; Gathercole, 2006). A simple phonological complexity of the NWRT

syllable structure and stress pattern can make the test relatively immune to effects of the amount of language exposure. Differences in language experience have more influence on knowledge-based measures of vocabulary and grammar than processing-based NWRT's (Engel, Santos and Gathercole, 2008). Children's performance on NWRT tasks is most commonly interpreted as a reflection of their phonological memory skills, although performance is also influenced by speech perception, lexical knowledge, and motor skills (Coady and Evans, 2008). There is a clear association between NWRT performance and vocabulary size (Gathercole *et al.*, 1999). Children with larger vocabularies perform better on NWR tasks. Gathercole and Baddeley (1989) found a strong relationship between NWRT performance and vocabulary acquisition. There is evidence that NWRT performance can predict new word learning (Gathercole, Hitch and Martin, 1997).

The identification of bilingual children with language impairment is challenging because the delays in language development can arise from impairment but also from external factors such as insufficient exposure to and consequently, limited knowledge of the target language (Kohnert, 2010). Their language skills depend on the amount of bilingual exposure (Thordardottir *et al.*, 2006) and the quality of input (Scheele, Leseman and Mayo, 2010). To minimize the bilingual disadvantage on NWRT, items should be used with a low phonotactic probability or word likeliness in the second language. This can be important when designing a NAMES® version for a group of related languages. Bantu for example is a group of over 440 distinct African languages (Wikipedia, 2019). Nonsense words with simple CVCV and CVCVC structures are relatively universal in terms of syllable structure, whereas nonsense words with consonant clusters (e.g. CCV) are more language specific. Languages differ with respect to many other aspects of lexical phonology, such as word lengths, suprasegmental characteristics and segmental inventories. In NWRT tasks children make more errors with consonants than with vowels (Lyregaard, 1997).

2.9 Other related tests

Several tests are related to the NAMES® test. They were designed for a similar purpose, different target group or they can provide complementary information. A selection of them will be discussed below.

2.9.1 Phonak Phoneme Perception Test 2.1

The Phonak Phoneme Perception Test (PPT) (Phonak, 2014), was designed to improve the client's speech intelligibility. Test results provide information about possible further improvements to a hearing aid's setting. The PPT is NOAH[2] compatible, available in 14 languages and consists of three subtests; Detection, Discrimination and Identification.

2.9.2 IfAP Lingogram

The Lingogram sound identification and detection test is based on the Ling Sounds (Ling, 2002). It is a software application, which runs on the BELLS® platform[3]. It is an adaptive test, in which the client must recognize one of the six Ling sounds, which according to an adaptive method vary in intensity. In this way a rough frequency specific audiogram type of graph can be constructed. It is attractive and fast for young children because it can be presented in the form of a game.

2.9.3 IfAP TiTaTu

TiTaTu (TeeTaaToo) is a syllable identification and discrimination test, which uses disyllables to detect and discriminate between phonemes. The first syllable is to focus the child’s attention while the hearing aid sound processor settles its active processing. Different sets of stimuli can be used, for example:

Set 1 (vowels) titi, tata, tutu, teetee, taitai

Set 2 (plosives) tata, dada, papa, baba, kaka, gaga

Set 3 (fricatives) sasa, shasha, fafa

---

[2] The NOAH software system is designed specifically for the hearing care industry, providing hearing care professionals with a unified system for performing client-related tasks.

[3] Battery for the Evaluation of Language and Listening Skills. BELLS is a software platform hosting several test and rehabilitation applications. See also the Methods section.

It is part of the BELLS® platform (Coninx, 2018a) and can be used from the age of 4 years.

2.9.4 IfAP Multi Frequency Animal Sound Test (mFAST)

mFAST is an adaptive frequency specific threshold measurement, where young children must identify known animal sounds. As a result, the test gives Pure Tone Average values (PTA) over 500Hz, 1kHz, 2kHz and 4kHz. mFAST is part of the BELLS® test battery.

2.9.5 Heidelberger Laut Differenzierungstest (H-LAD)

The Heidelberger Laut Differenzierungstest (H-LAD) is a computerized phoneme discrimination test for the diagnosis of dyslexia, developed in the late 1990's at the University of Heidelberg in Germany (Dierks et al., 1999). In subtest 1b nonsense words for phoneme discrimination are used. With the H-LAD Brunner and Stuhrmann (2013) found a high correlation of phoneme discrimination and spelling ability in the lower school grades.

2.9.6 Münchner AUditiver Screeningtest für Verarbeitungs- und Wahrnehmungsstörungen (MAUS)

The MAUS test is normed for ages 6 to 12 (Nickisch *et al.*, 2006).
This test consists of three parts:

1. Syllable sequence memory
2. Speech understanding in noise (words)
3. Phoneme identification and discrimination

The MAUS can determine to what extent the test results of an individual deviate from those of the normal primary school population. The MAUS can identify children at risk of having an Auditory Processing Disorder (APD) (Nickisch *et al.*, 2006).

2.9.7 Auditory Phoneme Evaluation (APE®)[4]

The Auditory Phoneme Evaluation (APE®) is an audiological evaluation tool that uses strictly defined phonemes as stimulus material for detection, discrimination and identification tests. The APE® was designed as a language-independent

[4] APE®, Registered trademark of Melakos NV, Antwerp, Belgium, www.melakos.be

test to yield supraliminal information on the auditory function with as little cognitive bias as possible. Preverbal infants as young as 7-8 months can be tested. This tool is mainly used for selecting Cochlear Implant candidates (Govaerts, Schauwers and Gillis, 2002).

2.9.8 Digit Triplet test

The digit triplets test uses digit triplets, for example (6-2-8) as speech material (Smits, Kapteyn and Houtgast, 2004; Smits and Houtgast, 2005; Smits, Goverts and Festen, 2013). This test was developed in a way that it can also be used by non-native Dutch speakers and the digits are presented in noise, to determine a person's Speech Reception Threshold (SRT). Digits are among the most frequent used words and therefore are very familiar. The test can easily be repeated, because the risk that people will remember which triplets are used is very low. The test is only meant as a quick screening test, compared to the Dutch standard sentence speech in noise test from Plomp & Mimpen (1979a; 1979b) and the CVC test from Bosman (1995).

2.9.9 The Phoneme Identification Test (PIT)

The Phoneme Identification Test (PIT) was developed to investigate the ability of children to use spectro-temporal cues to perceptually categorize speech sounds based on their rapidly changing formant frequencies (Cameron *et al.*, 2018). The PIT uses an adaptive two-alternative forced-choice (2AFC) procedure whereby the participant identifies a synthesized consonant-vowel (CV) (/ba/ or /da/) syllable. CV syllables differ only in the second formant (F2) frequency along an 11-step continuum (between 0% and 100%-representing an ideal /ba/ and /da/, respectively).

2.9.10 Vietnamese version of NAMES®

Nguyen (2017) developed a Vietnamese version of NAMES®, with a special feature for scoring tonal differences. He divided the participants in six age groups for validation. The normative values were calculated by averaging the PRS scores across the age groups. Nguyen found a deterioration in the fricative scores for the group of older listeners (76-85 years), which implies that the decline in phoneme scores was associated with high-frequency hearing loss by the older listeners (Gelfand, Piper and Silman, 1986; Maniwa, Jongman and

Wade, 2008). Nguyen found disparities in results depending on the way the NAMES® test was scored. The respondent could either respond verbally or in written form. The results showed that the listeners who responded verbally to the NAMES® test had a higher phoneme score (96%) than those who gave written answers (90%). These results revealed that the written response was riskier than the verbal response. This should be considered when interpreting the test results. He also investigated the significance of dialectal effects on phoneme scores of NAMES®. Overall, the non-native listeners scored poorer (roughly 1.5%) than the native listeners on the PRS. Although the difference of 1.5% was negligible, the result suggested a weak effect of dialect on the listeners' phoneme scores.

In the Dutch adaptation of the NAMES® test therefore, the effect of dialects in Dutch is investigated. Some of his participants, who were considered to have a mild or moderate hearing loss (based on their SRT), achieved high phoneme scores in NAMES®. This indicates that the NAMES® test is a very easy task for even those with moderate hearing loss, especially when the test is presented at a supra-threshold level.

2.10 Summary of related tests

Test	Intent	Purpose	Target group
Phonak PPT	Phoneme detection discrimination identification	Hearing aid fine tuning	Adults
IfAP Lingogram	Hearing threshold	Hearing diagnostics	Children
IfAP TiTaTu	Syllable identification	Rehabilitation	Children
IfAP mFAST	Hearing threshold	Hearing diagnostics	Children
H-LAD	Phoneme discrimination	Dyslexia diagnostics	Children and adults
MAUS	Phoneme identification discrimination	Auditory processing	Children and adults
APE®	Phoneme detection, discrimination identification	C.I. selection	Children
Digit Triplet test	Hearing threshold	Screening	Adults
PIT	Phoneme identification	Auditory processing	Children
NAMES®	Phoneme and word identification	Diagnostics and rehabilitation	Children and adults

Table 1 Tests related to NAMES®

Table 1 presents an overview of the main purposes of tests which are related to NAMES®. Phonak PPT is language independent but focusses on speech optimization for hearing aids in the high frequencies. The three subtests are compulsory for optimizing hearing aid fitting. This can be too time consuming for children. Lingogram provides an estimate of hearing thresholds of young children. It is not intended for phoneme identification but can be complementary to NAMES®, with information on the hearing acuity in different frequency regions. Children need training to connect the Ling pictures to a sound. mFAST has a similar purpose as Lingogram but uses animal sounds. It does not need

much explanation because it is assumed that the child already is acquainted with the animal sounds and their pictures. H-LAD, MAUS and PIT are meant for diagnosing processing disorders, in which phoneme identification and discrimination play a role. The PIT is designed to assess the spectral and temporal discrimination skills of children with APD, whereas NAMES® intends optimization of the hearing function. The Digit Triplet test, Bosman CVC lists and Plomp and Mimpen test determine a speech reception threshold (SRT). They provide no information on phoneme identification. The Bosman CVC lists are most commonly used for speech audiometry in the Netherlands, but the words are assumed to be known. That makes the test less suitable for very young children. Plomp and Mimpen sentence tests are too difficult for children, because of their limited phonological memory. The APE® test battery is mainly used for cochlear implant selection and optimizing the mapping by discrimination of phoneme clusters. The multiple-choice phoneme identification part of the test uses representations of drawings of sounding objects (onomatopoeia), like the Ling test. This is suitable for very young children, but the test results are not as detailed as those from NAMES®.
NAMES® has its own place in the battery of available tests. All these 10 tests are relatively language independent. Most of them have a different purpose and some of them are only suitable for adults. What makes NAMES® unique is that it is fast, and it gives specific data for optimizing the fitting of modern hearing devices and for testing the phonological development of children. NAMES® data can be presented as detailed phoneme information as well as quick overall scores.

2.11 Conclusion

Speech test materials should represent every day’s speech. Nonsense words must obey the phonotactical rules of the language for which the test was designed, and they can be made more or less word-like by complying to phonotactic rules as well as following the phonemic frequency of occurrence. For children the test should be short and age appropriate. A clear set of instructions for the client, will make the test more robust. Monosyllabic CVC

words are recommended for children, because they minimally appeal to their knowledge of language and short-term memory. Nonsense words are nowadays quite commonly used in the assessment of language- and auditory processing disorders in children.
Recording of the speech materials should be done professionally by a neutral speaker, with special attention to the intonation. The energy of the test words needs to be levelled and should be related to standardized calibration signals. Recent speech tests for children such as Titatu take time and dynamic processing algorithms of hearing aids into account. NAMES® can provide average scores per phoneme position in the word. Phoneme scores of the second syllable can cater for information about hearing aid dynamics.

In this literature review the researcher also referred to some older articles. That information is still valid, because languages and the fundamentals of phonetics do not change a lot in a few decades. Many scientific articles in the area of phonetics still refer to earlier research like that from Lyregaard (1997). In Dutch speech audiometry major fundamental research was done by Plomp and Mimpen (1979b) whose principles are considered to be a standard for speech testing. Many current articles still refer to their theories.

3 METHODS

3.1 Ethics statement

All methods in this study were approved by the University of Hertfordshire, Social Sciences, Arts and Humanities Faculty's Ethics Committee with Delegated Authority (ECDA) under protocol number: EDU/PGT/CP/03801 (Appendix 7). All participants were informed about the purpose of this research (Appendix 8) and they have signed a written consent (Appendix 9) before participating in this study.

3.2 Introduction

Word or phoneme recognition testing is routinely used by clinical audiologists to aid in the selection and evaluation of appropriate amplification, to determine site of lesion, to assess specific rehabilitative needs, and to assess central auditory function (Bess, 1983). The aim of the NAMES® test was to assess the correct recognition of words and phonemes, as they occur in the Dutch language. In this section the researcher will explain the types of research used, which factors were investigated and how. The researcher will explain how the test was compiled and recorded, how participants were selected and how the data was collected.

3.3 Method design used in this research

This research used a combined experimental and quantitative design. The experimental part maintained control on all the factors that could affect the results (Kombo and Tromp, 2006). The quantitative part of this research presents information about the test/retest conditions, which relate to the reliability of the test. As Babbie (2013) defines it: Quantitative methods emphasize objective measurements and the statistical, mathematical, numerical analysis of collected data using computational techniques. For instance, in this study the factor hearing loss (level) can impact the phoneme recognition. Normally this type of test should be done at a comfortable audible level. To investigate the effects of hearing loss however, instead of presenting the test at a fixed level of 65 dB_{SPL}, which is normal conversational speech level, the researcher choose to present the test also at two lower levels, 50 dB_{SPL} and 35

dB_{SPL}. This made it possible to explore ceiling and floor effects of the test. Experimenting with presentation and scoring methods enabled optimization of the test conditions and it yielded recommendations for future versions and adaptations of the NAMES® test. In the experimental part of the research a causal relationship between hearing loss, different types of hearing losses and the effects on phoneme identification could be scrutinized.

Through this triangulation of different test methods, validation of the obtained data was feasible. For example, a high correlation between poor consonant scores and presbycusis (high frequency) hearing loss is expected. With the experimental design, the researcher was in control of the test parameters, which could be changed individually, such as, selection of participants with or without hearing loss, scoring methods and regional dialects. It helped us to limit alternative explanations and infer direct causal relationships in the study. Quantitative data retrieved from the NAMES® test gave numerical information on the phoneme scores. The combination of this experimental / quantitative approach provided the highest level of evidence for this study. For a full validation of this test a larger quantitative study should be done with a higher number of participants, in different age groups and preferably also with young children.

The main aim of this research was to develop a Dutch version of the test and check the most optimal parameters for presenting and scoring.
The scoring of this test was done by identifying the phonemes which the person being tested repeats verbally, while the tester keyed them into the NAMES® programme. The NAMES® words were presented in a random order. The presented nonsense words were stored in the order in which they were presented in NAMES®, together with the keyed in responses. BELLS® calculated the word scores and the scores of seven pre-defined phoneme categories. For research purposes data in BELLS® was selected and exported in a CSV format. Consecutively the data was imported into Microsoft Excel for statistical analysis and manipulation.

3.4 Phonemic distribution

Words used in the test should equally represent the distribution of phonemes in the language for which the test is used (Martin, Champlin and Perez, 2000). With the limited number of words (n=20) it was not possible to create an exact representation of all Dutch phonemes but at least it matched the distribution as closely as possible. Calculated from the frequencies of the phonemes which are used in this version of the NAMES® test, 93.7% of the Dutch vowels and 86.1% of the Dutch consonants are represented. The /z/ was left out because in the Northwest of the Netherlands the /z/ is voiceless as /s/ and in the South it is a voiced sound.

For this study the token frequency data is used. Email correspondence with the author of the lists (Prof. M. van Oostendorp 2018, personal communication, 19 October) confirmed that it is very unusual to arrange vowels and consonants according to a type frequency. This is what normally is done with words in sentences or texts. In this case the 'token-frequency' is about how often a sound appears in a Dutch text when it has been described phonetically. Type frequency relates to the number of words within the database in which this phoneme occurs. Type and Token here relate to the words in the CELEX[5] database. The respective Dutch phoneme distributions for Vowels and Consonants can be found in tables 2 and 3.

[5] CELEX- Dutch Centre for Lexical Informaton

Dutch Vowel Distribution

Segment	Type frequency (%)	Segment	Token frequency (%)
[ɑ]	14.5	[ə]	23.2
[ɛ]	11	[ɑ]	12.7
[ɔ]	10.7	[ɛ]	11.5
[ɪ]	9.1	[ɛi]	9.1
[a]	7.7	[a]	7.6
[e]	6.7	[ɔ]	7.3
[i]	6	[ɪ]	7
[o]	6	[o]	5.3
[ʏ]	5.9	[e]	5
[u]	5.7	[i]	5
[ɛi]	4.4	[u]	2.4
[œy]	3.1	[ʏ]	1.2
[ə]	2.6	[œy]	1.1
[ø]	2.3	[y]	0.8
[y]	1.9	[ɑu]	0.6
[ɑu]	1.7	[ø]	0.3
[ɛː]	0.7	[ɛː]	<0.1
[ɔː]	0.1	[ɔː]	<0.1
[œː]	0.1	[œː]	<0.1

Table 2 Dutch Vowel distribution (Linke and Oostendorp, 2018b)

Dutch has several diphthongs, as can be seen in figure 1, but only three of them are indisputable phonemic [ɛi], [œy] and [ɑu] (Collins and Mees, 2003). All of them end in a non-syllabic close vowel [i,y,u], but they may begin with a variety of other vowels.

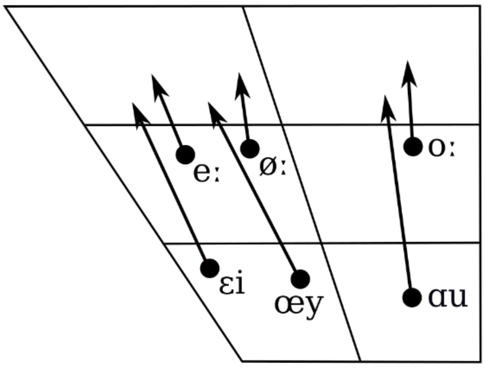


Fig. 1 Diphtongs of Northern Standard Dutch (Gussenhoven, 1999)

Dutch Consonant distribution

Segment	Type frequency (%)	Segment	Token frequency (%)
[s]	12.4	[n]	17.8
[r]	12.3	[t]	14.5
[t]	12.1	[d]	9.4
[l]	9.5	[r]	9.3
[k]	8.7	[z]	5.6
[n]	6.9	[l]	5.4
[p]	6.3	[k]	5
[x]	5	[m]	4.9
[m]	4.7	[v]	4.5
[f]	3.5	[s]	4.2
[b]	3.1	[x]	4.2
[ʋ]	3	[h]	3.7
[d]	2.2	[ʋ]	3.3
[v]	1.8	[p]	3.2
[j]	1.8	[b]	1.4
[h]	1.6	[f]	1.3
[z]	1.6	[j]	1
[ŋ]	1.6	[ŋ]	0.7
[ʃ]	1.1	[χ]	0.3
[χ]	0.4	[ʃ]	<0.1
[g]	0.3	[ʒ]	<0.1
[ʒ]	0.2	[g]	<0.1
[dʒ]	0.1	[dʒ]	<0.1
[c]	<0.1	[c]	<0.1
[ɲ]	<0.1	[ɲ]	<0.1

Table 3 Dutch Consonant distribution (Linke and Oostendorp, 2018a)

3.5 Phonotactic rules

The term *phonotactics* is a composition from the Greek words for “sound” and “arrange” (Booij, 1978). In phonology, phonotactics is the study of the ways in which phonemes are allowed to combine in a particular language. Phonotactic constraints are rules and restrictions concerning the ways in which syllables can be created in a language. Linguist Zsiga (2012) observes that languages “do not allow random sequences of sounds; rather, the sound sequences which a language allows are a systematic and predictable part of its structure”.

The description of the phonotactics of Dutch relies heavily on the concept of the syllable (σ) (Köhnlein and Linke, 2018b). The syllable is assumed to consist of the hierarchically ordered constituents as can be seen in Figure 2.

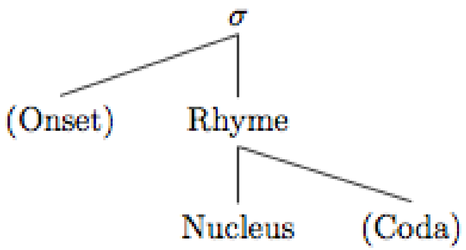


Fig. 2 Constituents of a syllable (Köhnlein and Linke, 2018a)

There are however other factors that influence phonotactics, such as prosodic factors. NAMES® words were recorded with special attention on the pronunciation of the words, so that the prosody is similar in all the words. Köhnlein & Linke (2018b) give a summary of factors that play a role in Dutch phonotactics. Related to the NAMES® test design these are:

3.6 Phonotactics at the syllable level

- The occurrence of consonants in clusters of two or more consonants is more restricted. NAMES® lists use the CVCVC structure, thus these limitations do not play a role.
- The nucleus position of a syllable in Dutch is usually occupied by a vowel. The NAMES® CVCVC string complies with this.
- All consonants of Dutch, except for /h/, can occur in coda position (Booij, 1995).

3.7 Phonotactics at the word level

In words with more than one syllable, the sonority relation at the syllable contact plays a role (Seo, 2011). The Syllable Contact Law (SCL) according to Gouskova (2004) belongs to a class of constraints that require adjacent elements to differ by a certain number of steps of a hierarchy. In the NAMES® CVCVC string the first syllable always ends with a vowel, which according to the sonority hierarchy (Booij, 1995) is stronger than the consonants, as illustrated below.

Sonority hierarchy:
vowels > glides > rhotics > laterals > nasals > fricatives > stops (decreasing sonority) (Booij, 1995).

In the Dutch language, there are several rules for word stress. Köhnlein (2018) mentions that in a large majority of cases, the placement of primary stress is restricted to one of the last three syllables of a word. Phonetically primary stress in Dutch correlates with:

- Pitch movements
- Intensity
- Vowel duration, and
- Vowel quality

During the recording of the NAMES® words, putting stress on either syllable was avoided, to keep the word stress as identical as possible.

3.8 Generation of the NAMES® word lists

A specially designed Microsoft® Excel spreadsheet (Coninx, 2018b) was used to generate 4 lists of 20 words each. When words were generated randomly, some double words or words which resemble other known Dutch, English or German words were unavoidable. Each row of words across the 4 lists was generated using the same phonemes, therefore the words could be interchanged. From these lists, one word from each row was selected for the final test.

The CVCVC word was built as follows:
Consonant 1 - Vowel 1 - Consonant 2 - Vowel 2 - Consonant x3 (Con1-Vow1-Con2-Vow2-Conx3).

In this construct it was important to know whether a consonant (Conx3) could occur in a final position of a Dutch word. In the Excel spreadsheet only the consonants which are allowed at a final position of a word were included in the Conx3 list.

Vowels:

From the list of possible Dutch vowels, 10 (N1=10) vowels were selected for use in the NAMES® lists. In the Dutch language there are 19 possible vowels, but the selected 10 represented 93.7% of the occurrence of vowels in spoken Dutch. N1 did not exceed 9-12 for two reasons:

- The total number of vowels (at phoneme positions 2 and 4, Vow1 and Vow2) is 40. Using 10 different vowels means that on average each vowel will be used 4 times. A lower number would not allow statistical analysis.
- When using free typing input, a screen-based phoneme keyboard for vowels was used. This keyboard did not contain too many keys, because it could complicate and slow down the process of entering patient response data.

Dutch phonotactical rules were observed and special and complicated vowels were avoided. In the Excel sheet a list of 40 vowels was generated, which contains N1, in this case 10 different vowels. The numbers of the vowels were rounded, and their representation was entered in the Excel-generator sheet.

Consonants:

Twelve consonants (Mc135, Con1-Con2-Conx3) with the highest Frequency of Occurrence were selected. In total they represented 86.1% of the consonants in Dutch words. The /z/ which represents 5.6% was left out due to large regional differences in pronunciation. From the remaining 12 consonants /d/, /v/ and /h/ were excluded at position Conx3, as they never occur as a final consonant in Dutch words. In the recordings there was a neutral pronunciation of the /r/ which is also marked by strong regional variations. The number of consonants, in this version was limited to 12 because of the screen-based keyboard. Sixty consonants were selected. The consonant distribution was calculated and rounded, and the consonants were split up in two lists. One of 40 consonants (Mc13) for the positions 1 and 3 in the word and a list of 20 consonants (Mc5) for the final position. All 60 consonants were entered in the Excel word-generator (Coninx, 2018b).

Generating words:

a. The Vowels, Consonants and final Consonants in the word-generator Excel sheet were randomized. This was done through a formula in Excel and the four wordlists (A, B, C and D) were generated accordingly (table 4).
b. The 80 generated words were checked according to the criteria below:
    - From each row the most unknow “word” was selected
    - Words that are close to a meaningful word in Dutch, English or German were avoided
    - Meaningful words were not selected

3.9 Selection and exclusion of words

Nr	List A	List B	List C	List D	extra	List 3	List 4
1	tɛsan	sɛtan	tasɛn	satɛn		tɛsan	satɛn
2	pəxət	xəpət	pəxət	xəpət		pəxət	xəpət
3	rənen	nəren	renən	nerən		nəren	renən
4	dɛtɛx	tɛdɛx	dɛtɛx	tɛdɛx		tɛdɛx	dɛtɛx
5	dɔvip	vɔdip	divɔp	vidɔp		vidɔp	vɔdip
6	halɔt	lahɔt	hɔlat	lɔhat		hɔlat	lahɔt
7	datɑr	tadɑr	datɑr	tadɑr		tadɑr	datɑr
8	tIrɛim	rItɛim	tɛirIm	rɛitIm		tɛirIm	rɛitIm
9	danIk	nadIk	dInɑk	nIdɑk		dInɑk	nadIk
10	rahɛin	harɛin	rɛihɑn	hɛirɑn		rahɛin	hɛirɑn
11	vətes	təves	vetəs	tevəs		vətes	tevəs
12	nədor	dənor	nodər	donər		nodər	nədor
13	nɔmot	mɔnot	nomɔt	monɔt		monɔt	nɔmot
14	mIlɛt	lImɛt	mɛlIt	lɛmIt	mɛlIt	mɛlIt	mɛlIt
15	dədəl	dədəl	dədəl	dədəl		dədəl	dədəl
16	nɛisɑk	sɛinɑk	nasɛik	sanɛik		sanɛik	sɛinɑk
17	nakət	kanət	nəkat	kənat		kənat	nəkat
18	vəgɛn	gəvɛn	vɛgən	gɛvən		vəgɛn	gəvɛn
19	təkar	kətar	takur	katər	ratək	ratək	takər
20	nirɛil	rinɛil	nɛiril	rɛinil		nɛiril	nirɛil

Table 4 Generated NAMES®-NL word lists

The words marked in green were in the final selection for List 3, and the words in blue were selected for list 4. Words marked in yellow had a resemblance with Dutch words and therefore were excluded from the selection. The word "mɛlIt" in row 14 was selected in both lists because of the better pronunciation by the speaker. From row 19, an additional word "ratək" was composed from the available phoneme combinations, to allow the /r/ to appear two times as a start consonant and two times as a final consonant, for a better balancing of the list.

3.10 Audio recording

3.10.1 Recording equipment and specifications

The recordings for the NAMES® word lists were done in a sound treated chamber at an audiological centre. A large membrane 2/3-inch AKG Perception 120 USB microphone was connected to a laptop via a standard USB port. The microphone was mounted on a shock mount with a plop filter on a tripod as in figure 3.

Fig. 3 Audio recording setup

The words were recorded with Audacity® version 2.2.2 software as a mono signal in 24-bit resolution at a 44.1 kHz sample rate. Through the on-body switch on the microphone a bass cut was selected to avoid unwanted low frequency noise. Recordings were stored on the computer in an uncompressed PCM wave (.wav) format.

3.10.2 Recording procedure

The word lists were spoken by an experienced female speaker. The two main reasons for this were that firstly the test is mainly intended for use with children, and young children are often more acquainted to the voice of the mother. Secondly, the overall formant frequencies of female speakers are higher than those from male speakers (Pépiot, 2015). Pépiot also found that consonants were proportionally longer in words produced by female speakers than by men, and they are likely to be more important than vowels in oral word recognition (Owren and Cardillo, 2006). Therefore, female speakers tend to produce “clearer” speech. Lyregaard (1997) recommends the use of a common national dialect, preferably a Radio or TV broadcasting speaker, because most people are used to that dialect. During the recordings, each word was preceded by a carrier word and followed by a carrier word. For example: “one–word1–two”, and “one-word2–two”, to avoid prosodic differences towards the end of the word. The words were spoken by a presenter first and were then repeated by the speaker to avoid prosodic differences towards the end of the lists. Words were recorded at a -12dB peak level to have a good amount of headroom to avoid distortion.

3.11 BELLS® platform and NAMES® interface

3.11.1 BELLS® platform

BELLS® is an acronym for “Battery for the Evaluation of Language and Listening Skills” (Coninx, 2018a). It is a software platform developed over the last two decades by Prof. Coninx at IfAP[6]. This platform is a test management system with a client database and test interfaces for several audiological tests and rehabilitation tools. The BELLS® database can be used for detailed inspection of the test results. Batch files can be created for randomisation and the platform is flexible and suitable for research. By using an external microphone, it can also judge and reject test results based on ambient noise and it can record client responses for later evaluation.

3.11.2 NAMES® interface

The NAMES® test offers versions that differ mainly in the way, the response from the patient is registered and entered into the computer.
The scoring methods are:

---

[6] Institut für Audiopädagogik (audiopedagogics) IfAP Solingen Germany
Coninx is Emeritus Professor in Educational Audiology, University of Cologne, Faculty of Human Sciences Department of Special Education and Rehabilitation.

1. The patient repeats what he/she has heard, and the examiner selects the buttons of the phonemes which were correctly repeated.
2. The patient repeats the stimulus and the examiner enters the phonemes into 4 or 5 response buttons through special on-screen simplified keyboards. Table 5 shows the characters on the NAMES® keyboard compared to the IPA descriptors (Appendix 1). The examiner does not see the target word on the screen.

IPA	*NAMES®*
[ə]	*u*
[ɑ]	*a*
[ɛ]	*e*
[ɛi]	*ei*
[a:]	*aa*
[ɔ]	*o*
[ɪ]	*i*
[o:]	*oo*
[e:]	*ee*
[i]	*ie*

Table 5 IPA versus NAMES® keyboard characters.

3. NAMES® software identifies the phonemes in the patient's spoken response automatically, using an automatic speech recognizer software tool operating at the phoneme level. This feature is still under development.
4. There are also options for self-test, where the patient can key in or select the phonemes which he or she recognized.

3.12 Test equipment

Participants in this research were tested in sound treated chambers at audiological centres. All testing was done using a Sennheiser HD 280 Pro closed circumaural headphone with a high ambient noise attenuation (<32 dB). The test words were presented through NAMES® software running on a

Microsoft Windows 10 laptop with a touch screen for collecting responses. The audio signal to the headphone was delivered through an external NuForce uDAC-2 asynchronous 24-bit, USB Digital to Analog Converter – headphone amplifier. The volume control of the uDAC was set and fixed in the mid position. Figure 4 shows a schematic diagram of the test setting.

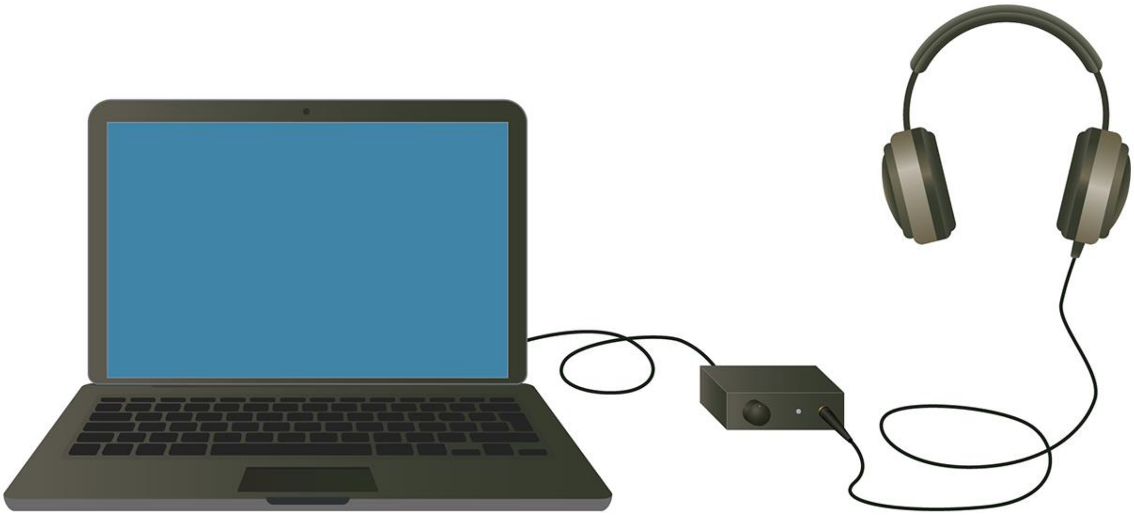


Fig. 4 NAMES® test setup

The combination of the headphone and DAC was calibrated at IfAP using a Grass calibration system. When spoken at conversational level, running speech averages 65 dB_{SPL} with positive peaks of the signal accruing 12 dB above the average level, and the negative peaks occurring at approximately 18 dB below the average level (ANSI, 2009; Skinner, 1988). Thus, the intensity range of average conversational speech is approximately 30 dB (47-77 dB_{SPL}). For the actual test, a presentation level which most likely results in the highest possible score should be used. According to findings of Maroonroge and Diefendorf (1984) this should be done at 30 to 40 dB_{SL} (relative to SRT) for those with normal hearing, and 40 dB_{SL} for hearing impaired clients. In clients with retro cochlear pathology a decrease in speech recognition may be shown when the intensity increases (Jerger and Jerger, 1971; Dirks *et al.*, 1977). This effect is called “roll-over”. Therefore, the words should not be presented too loudly. The intensity of the speech level was determined by measuring the Root Mean Square (RMS) value of the NAMES® words and adjusting the level to the reference CCITT noise which is generated and used for calibration in the BELLS® platform. Figure 5 shows the BELLS® calibration menu.

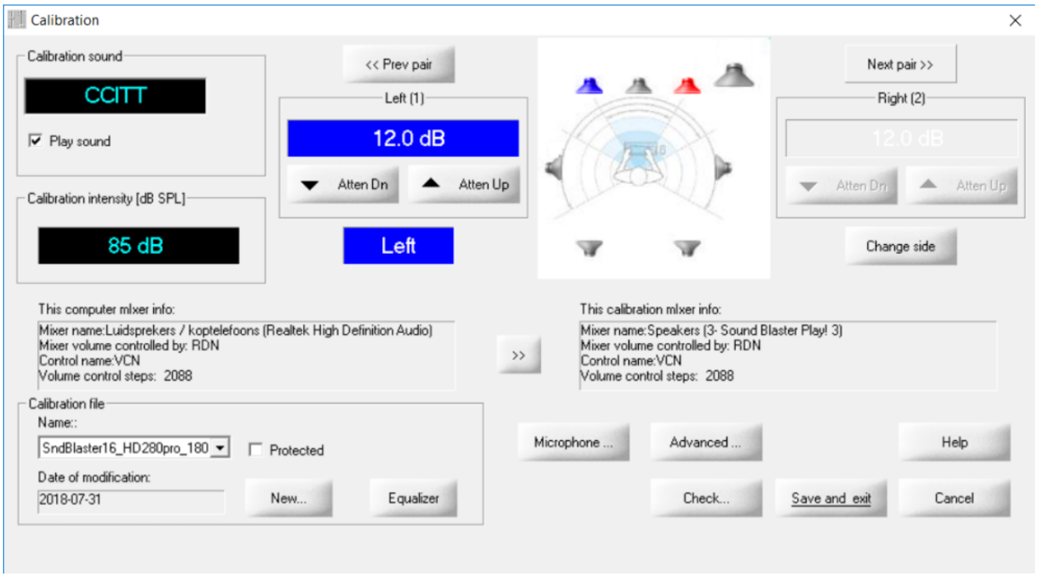


Fig. 5 BELLS® calibration menu

3.13 Selection of the participants

Depending on the age and status of hearing loss of the participant he or she was placed in either group 1 (18-59 years, normal hearing) or in group 2 (>= 60 years or hearing loss). The groups were sub-divided with respect to the region in the Netherlands from which they originate. Group 1 consisted of 35 normal hearing persons. Their hearing acuity was screened using the BELLS® Duotone®[7] test (Appendix 3) with the frequencies 500 Hz and 6 kHz, or by conducting BELLS® Pure Tone threshold audiometry (Appendix 4). Participants in this group did not have a higher threshold than 30 dB_{HL} at one of the octave frequencies between 500Hz and 6 KHz. Group 2 consisted of 24 participants between 60 and 80 years, as well as younger participants with a known hearing loss (PTA >30dB_{HL}). This group was used to determine the effects of hearing loss on the NAMES® test results. Pure Tone air conduction audiometry was done with this group at octave frequencies from 500Hz to 8 kHz. Participants were recruited from three different regions in the Netherlands, to investigate whether there were effects of dialects.

[7] The DuoTone® procedure is patended by Coninx IfAP Solingen.

3.14 Test Procedure

The NAMES® words were presented at both ears asynchronously, monaurally in a random ear and word order. The participants were asked to repeat what they had heard and the test leader keyed in the responses on a dedicated touchscreen keyboard as shown in figure 6. The instructions for participants can be found in Appendix 10. The participant's responses were recorded for later verification by other examiners which is further explained in the "Results" section.

3.15 Scoring method used in this research

Scoring in this research was done by the test leader who keyed in the phonemes spoken by the participant. Figure 6 shows a screenshot of the Vowel keyboard. The interface is designed in such a way that the layout of the keyboard changes for consonants and vowels depending on the phoneme position in the word. This makes the interface clear and efficient. It is also possible to have both keyboards on the same screen, but prior experiences prove that responding takes longer, which increases the test time. The position and the characters of the phonemes on the keyboard can be set individually in the BELLS® software. The arrangement of the keys can also be changed. Furthermore, it is possible to include keys such as "All correct" and "All wrong" as in figure 7, to speed up the entry process. This was not used in this experiment. It is also possible to score on the entire word for screening purposes. In this research we scored all the phonemes individually to have more detailed results. The "NAMES®" word can be presented on the screen, but for this research a blind scoring was preferred by the test leader, to minimize a response bias.

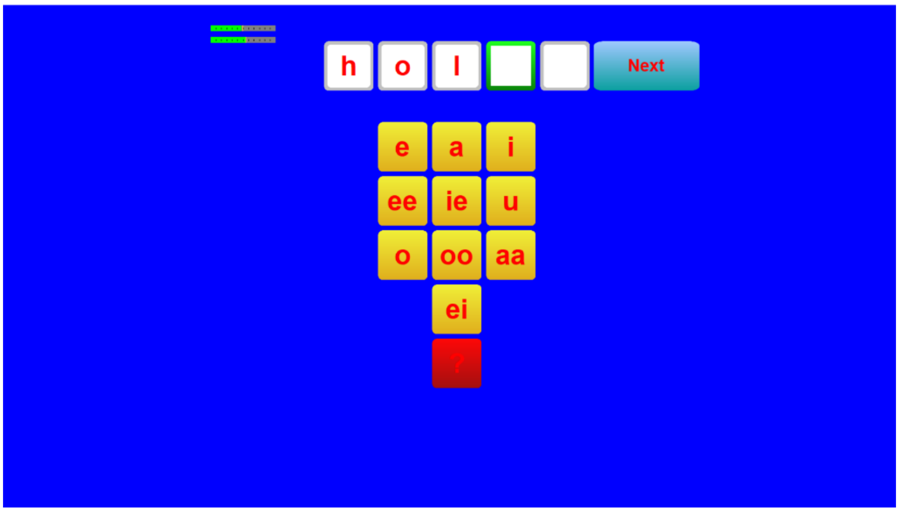


Fig. 6 Screenshot NAMES® vowel score keyboard

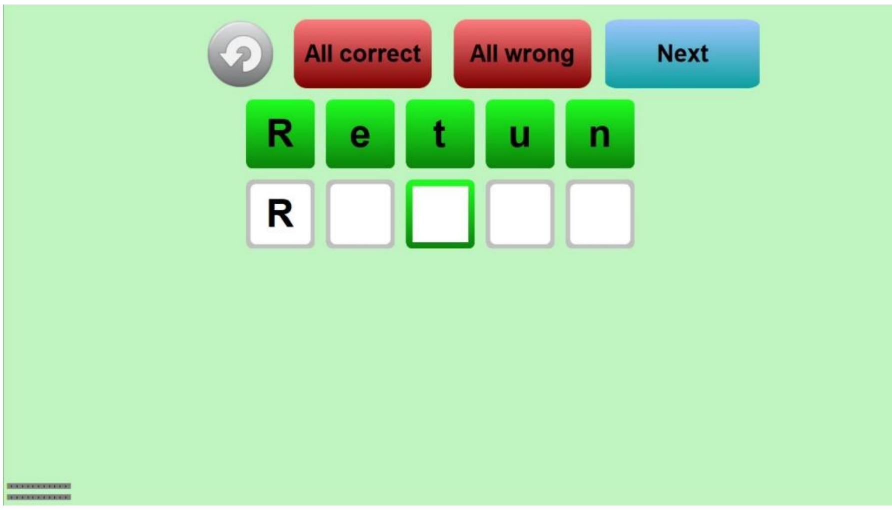


Fig. 7 Screenshot NAMES® score with extended options

3.16 Data collection

The data collection protocol started with presenting the word lists in the following order: 65 dB (R/L) – 65 dB (L/R) – 50 dB (R/L) – 50 dB (L/R) – 35 dB (R/L) – 35 dB (L/R). The test ear to start with was selected randomly and at each consecutive test the side was changed. Soon after the start of the experiment a memory effect was suspected because some participants could literally reproduce the words at the lowest presentation level, despite the use of nonsense words which should be difficult to remember. Therefore, it was decided to reverse the presentation order to minimize that effect: 35 dB (R/L) – 35 dB (L/R) – 50 dB (R/L) – 50 dB (L/R) – 65 dB (R/L) – 65 dB (L/R).

Not all participants could be tested at 35dB. After testing some more participants, it was found that a few of them reported that they could hear the words in both ears. A check of the equipment showed that when the right ear was selected, both ears were stimulated. This fault in the software corrupted the data. Therefore, it was decided to discard the acquired data and first have the software problem fixed.
There still was a suspicion that a few participants could remember some words or part of the words. It was decided to compile another list (List 4) out of the 4 generated lists (A, B, C, D), see table 4. In the two lists which were used, only some phonemes exchanged position, so the lists could be relatively similar. At each level, two different lists were used for the right and left ears and words were still presented in a random order.

Audio recordings were made from most of the tests to investigate the reliability of the scoring by the researcher. This was done with a Tascam DR-05 handheld audio recorder placed near to the respondent.

4 RESULTS AND ANALYSIS

4.1 Introduction

Speech Audiometry can be done for different purposes. It either can be used for differential diagnosis, or for hearing evaluation that is related to treatment of communication problems associated with hearing disorders (Lawson and Peterson, 2011). The NAMES® test focusses more on the latter one. This section will discuss in detail the results of the two groups of participants; normal hearing listeners and the group with hearing loss. Results will be explained with a few examples. Inter-rater reliability of different examiners will be discussed, and other influencing factors and regional differences will be highlighted.

4.2 Score interpretation

NAMES® can present results as Word Recognition Score (WRS), Phoneme Recognition Score (PRS) and it can present scores of predefined phoneme categories. PRS is the number (percentage) of phonemes correctly identified out of 100 phonemes per list. Phoneme error analysis is important in hearing aid assessment and aural rehabilitation. Table 6, by Kramer (2018) shows the commonly used categories to qualify Word Recognition Scores. This kind of qualification is typically used by audiologists to give some type of limited explanation.

WRS Word (Percent Correct)	Degree of Impairment	Word Recognition Ability
100-90	None	Excellent/Normal
89-75	Slight	Good
74-60	Moderate	Fair
59-50	Poor	Poor
<50	Very poor	Very poor

Table 6 Categories to describe results of WRS testing (Kramer 2018)

Madell (2011) however is much more critical towards speech perception scores, especially for children in a regular classroom (see table 7). She argues that, if a child has a speech perception score of 74% and it is described as good or excellent, the assumption will be that the child is doing well and that nothing should be changed in his management. But if it is rated as fair, professionals should try their best to improve speech perception. Test results are critical in planning hearing management, especially for children who are building their phonemic awareness (Miller, Bergeron and Connor, 2008). Well fitted hearing devices and optimized acoustical classroom conditions are imperative to improve learning conditions for hearing impaired children. I concur with Madell, being very critical on presenting the scores. Monitoring tests like NAMES® help professionals with evidence about the child's phonological development.

Qualification	Speech Perception Score
Excellent	90-100%
Good	80-89%
Fair	70-79%
Poor	<70%

Table 7 Speech Perception Qualifiers (Madell et al 2011)

The seven phoneme categories which were defined for the Dutch version of the NAMES® test can be found in the first column of table 8. Column two gives the descriptor used on the NAMES® keyboard and column three shows the IPA descriptor of the phoneme. The last two columns provide phoneme examples in Dutch and British English words to the extent that the phoneme exists in English. For this study, a group of 77 participants was recruited. But due to a software problem at the beginning of the experiment, the first 18 participants were discarded. Two other participants were rejected; one because of wrong test settings and the other one because of profound hearing loss. Participants were selected from the South, North and West of the Netherlands.

Category	Phoneme NAMES®	IPA descriptor	Example Dutch	Example (British) English
Vowel	e	/ɛ/	leg	get
Vowel	a	/ɑ/	pan	arm
Vowel	i	/I/	lip	sit
Vowel	ee	/e:/	zee	
Vowel	ie	/i/	liep	heat
Vowel	u	/ə/	put	
Vowel	o	/ɔ/	zot	not
Vowel	oo	/o:/	boom	
Vowel	aa	/a:/	laat	half
Diphthong	ei	/ɛi/	klein	
Plosive	t	/t/	tak	time
Plosive	d	/d/	dak	do
Plosive	p	/p/	paard	pig
Plosive	k	/k/	kas	kilo
Fricative	ch	/x/	lach	
Fricative	s	/s/	sap	six
Fricative	v	/v/	vuil	very
Nasal/lateral	n	/n/	nat	no
Nasal/lateral	m	/m/	man	milk
Nasal/lateral	l	/l/	lat	live
Rhotic	r	/r/	rat	read
Aspirate	h	/h/	waarheid	home

Table 8 Phonemes and phoneme categories of the Dutch NAMES® test

Whenever possible the researcher tried to conduct the NAMES® test at three different levels and separate for each ear; 35dB$_{SPL}$ - 50dB$_{SPL}$ and 65dB$_{SPL}$. In the group with hearing loss it was not possible to test all participants at 35dB and 50dB. Table 9 presents the distribution of participants according to the two groups and the three regions. The average age of the group 1 participants was 38;6 years (n=35).

Region	Group 1	Group 2	Σ
North	5	8	13
South	23	8	31
West	7	6	13

Table 9 Distribution of participants (n=57)

4.3 Overall results

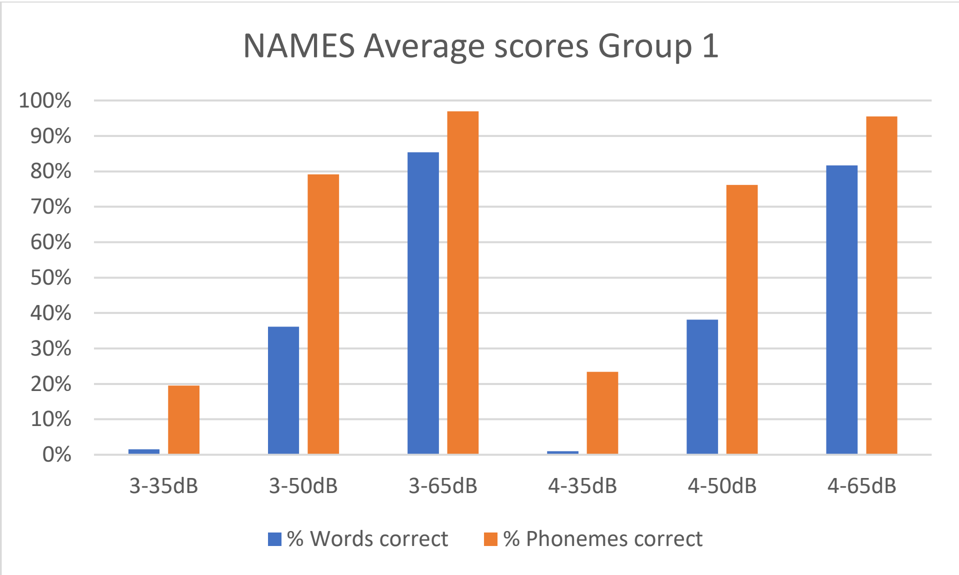


Fig. 8 NAMES® average word scores per list Group 1

Speech test scores for normal hearing participants at conversational level (65 dB_{SPL}) should be relatively high, near to 100% (Nguyen, 2017). The graph in figure 8 shows that the phoneme scores were significantly higher than the word scores, as expected (Markides, 1978). In more difficult listening situations or in case of hearing loss, the differences between word- and phoneme scores become larger (Billings *et al.*, 2016). Minimal differences were observed between the two Dutch wordlists used. The maximum average phoneme scores for this group reached up to 97% for list 3 and 96% for list 4 respectively, which is similar to the findings of Nguyen who found an average 96% PRS (Nguyen, 2017). The use of nonsense words contributes to the fact that scores do not reach the maximum 100% score at normal conversation level.

After the researcher had tested the first group of participants, memory effects were suspected, because even at lower presentation levels a few participants literally repeated some of the words. In response to that the presentation order of the lists was reversed. The test was started at 35 dB instead of 65 dB and a second word list (list 4) was included. Each ear was tested with a separate list. Both lists consisted of the same phonemes, but with a different order within the words. To investigate whether both lists were comparable, a Long Term Average Speech Spectrum (LTASS) analysis was done with the analysis function of Cool Edit 2.1 software (2003). Byrne et al. (1994) state that the representations of the long-term average spectrum of speech have various acoustical and audiological applications. One of the examples which they give is the use in hearing aid prescription procedures and prescriptive formula. The researcher measured three LTASS spectra which are presented in figure 9.

1. 2 minutes running speech from the female speaker of the NAMES-NL words (blue)
2. NAMES-NL list 3 (green)
3. NAMES-NL list 4 (red)

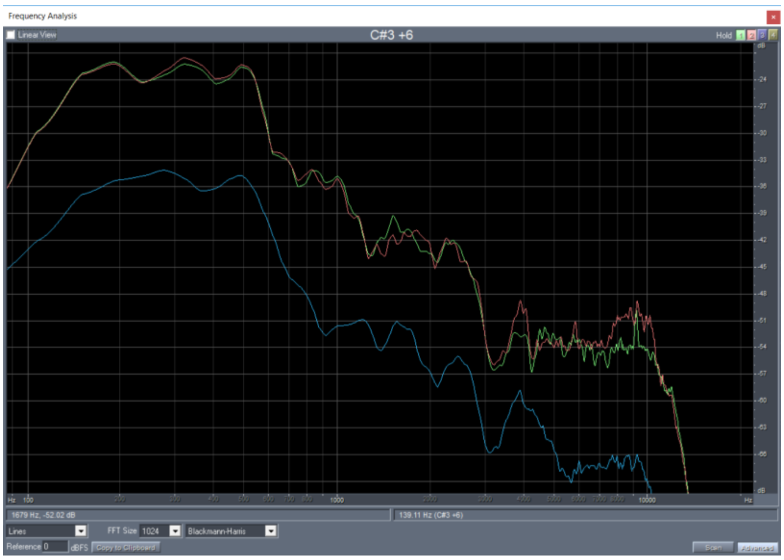


Fig. 9 LTASS spectra

Blue-running speech female NAMES speaker Green–List 3-65 dB Red–List 4-65 dB

The purpose of this measurement was twofold; compare the spectra of the two wordlists used in this experiment, and to investigate how these relate to running Dutch contemporary speech.

The spectra of both lists (green and red curves) in figure 9 are nearly identical. This is expected, because the lists consist of the same phonemes, but only with some minor position changes within the words. Differences in amplitude are less than 3dB. The shapes of the LTASS spectra of the two wordlists are also quite comparable with the LTASS of running speech. Differences between them can be explained by the fact that not all phonemes are represented in the two wordlists, but only the most frequent phonemes. Common in the three graphs is the decrease in amplitude of approximately 30 dB at 6 KHz. This is also what Byrne (1994) found in his study where he compares the LTASS spectra from 12 different languages. The vowels, which are in the 400Hz – 500 Hz region provide the greatest energy.

4.4 Results of group 1

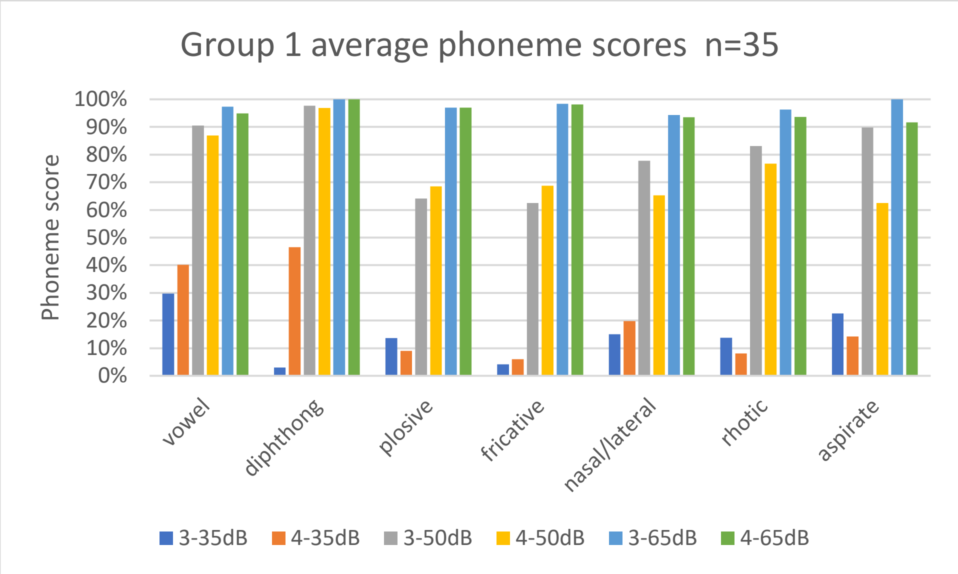


Fig. 10 NAMES® average phoneme scores Group 1

At 65dB presentation level, all scores of Group 1 remain above 90%. The aspirate and diphthong categories led to extreme values because they represent a relatively small number of phonemes and thus, they are statistically less representative.

The 50 dB scores range between 65 and 98%, which reflects the difficulty of phoneme identification between the seven categories. The 35dB presentation level reaches the hearing threshold for most participants, so these results are less predictable. Vowels have higher energy levels and therefore present higher scores.

4.5 Results of group 2

The results in group 2 are much more dependent on the participant's hearing acuity, so generalization of the data is less meaningful. Therefore, results will be discussed based on a few examples. Comparing the average results of group 2 in figure 11 with the average results of group 1 in figure 8 shows that in general the scores are lower and that the differences between word and phoneme scores are larger.

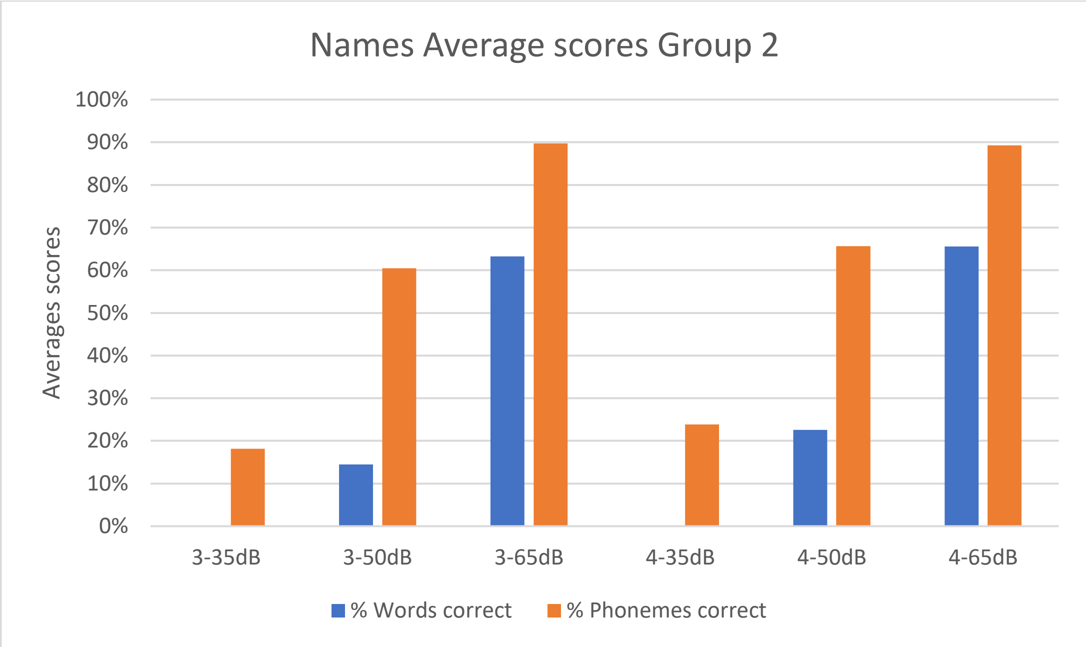


Fig. 11 NAMES® average word scores per list Group 2

Most of the participants in group 2 had a high frequency hearing loss. This impacted especially the fricatives, which represent the higher frequencies. A difference in presentation level of 15 dB, already gave a decrease in fricative scores of more than 50% at 50 dB, as can be seen in figure 12. From the same graph, it is obvious that the diphthong scores in group 2 are completely different with that of group 1 in figure 10. In this test only one type of diphthong the /εi/ was used. In both lists, there were only 4 diphthongs out of the 100 phonemes. Therefore, this category can show extreme values, which are not always statistically relevant. The same applies for the aspirates.

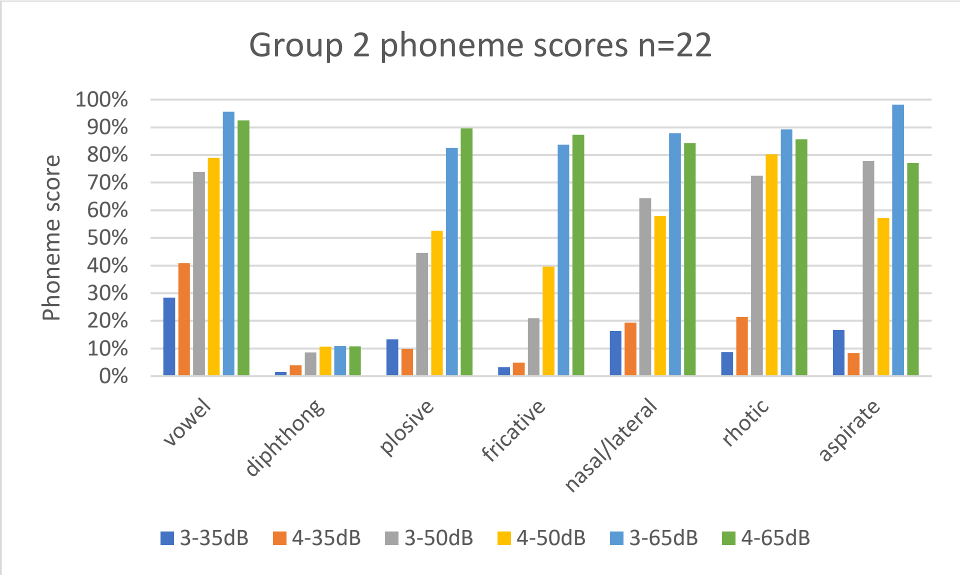


Fig. 12 NAMES® average phoneme scores Group 2

4.6 Examples

In the next pages a few examples explaining the impact of the participant's phoneme scores at different types of hearing loss are discussed.

Participant N57

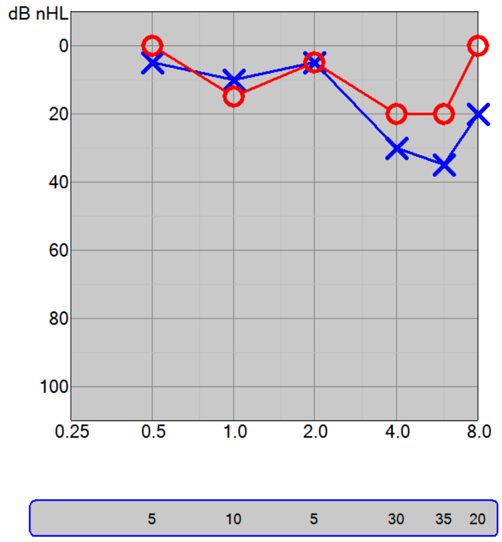


Fig. 13 Pure tone audiogram of participant N57

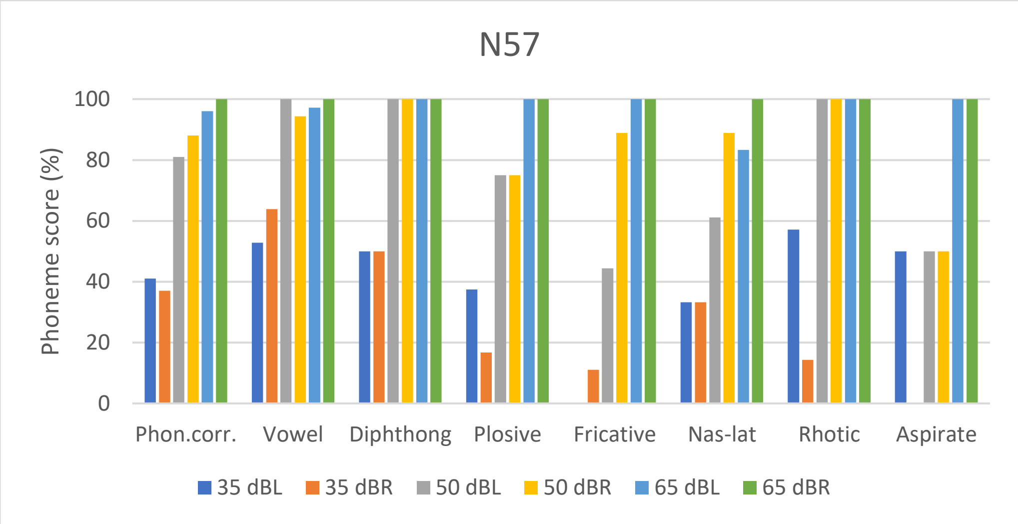


Fig. 14 Phoneme scores per category of participant N57

The audiogram in figure 13 shows that participant N57 had a slight high frequency loss, above 2 kHz. At 6 kHz the loss in the left ear was about 15 dB greater than at the right ear. The fricative scores in figure 14 at 50 dB clearly show the impact of that loss on the phoneme score compared to the right ear. Also, in the nasal-lateral category a lower score is found for the left ear. It is arguable whether this difference would be observed in speech audiometry with existing words.

Participant N73

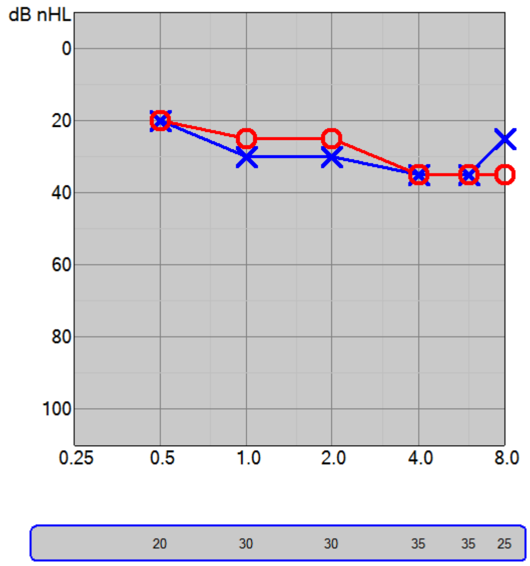


Fig. 15 Pure tone audiogram of participant N73

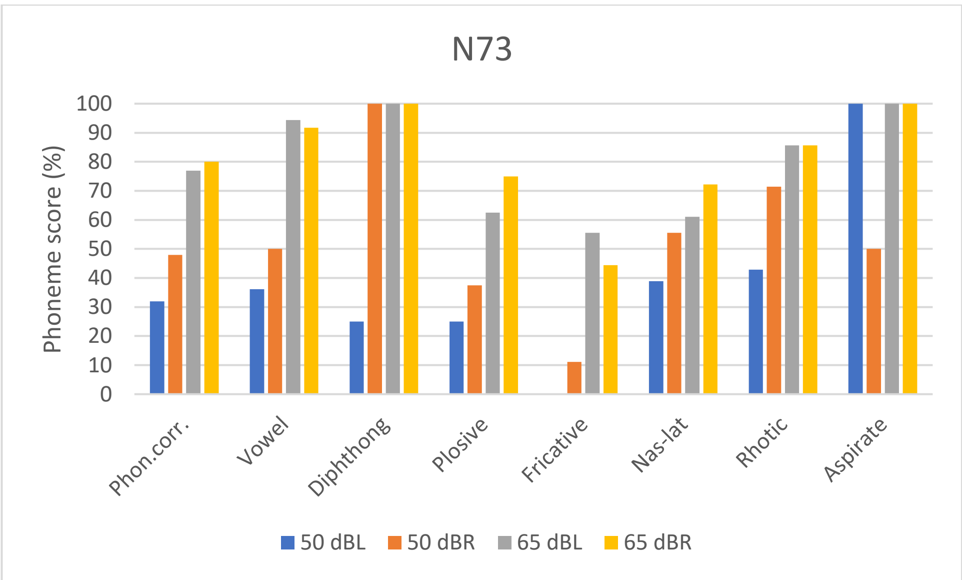


Fig. 16 Phoneme scores per category of participant N73

A relative flat mild hearing loss as in the audiogram in figure 15 impacts all phonemes and results in a lower overall score (figure 16). Measurements at a 35dB were not possible with this participant. WHO (2013) refers to a disabling hearing loss, when it is greater than 40 dB in the better hearing ear in adults and greater than 30 dB in the better hearing ear in children. This marks the importance of this test for children.

Participant N49

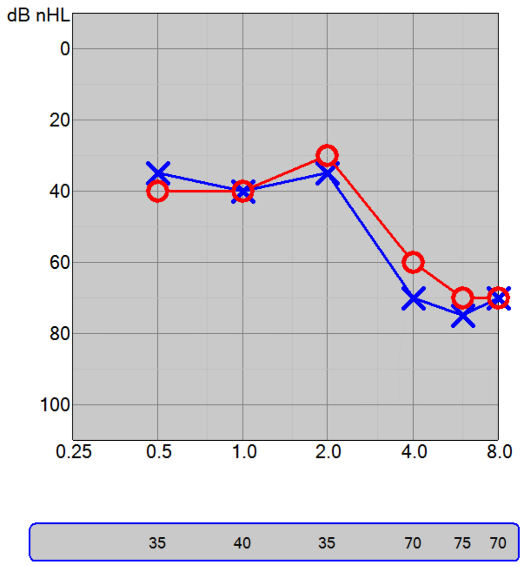


Fig. 17 Pure tone audiogram of participant N49

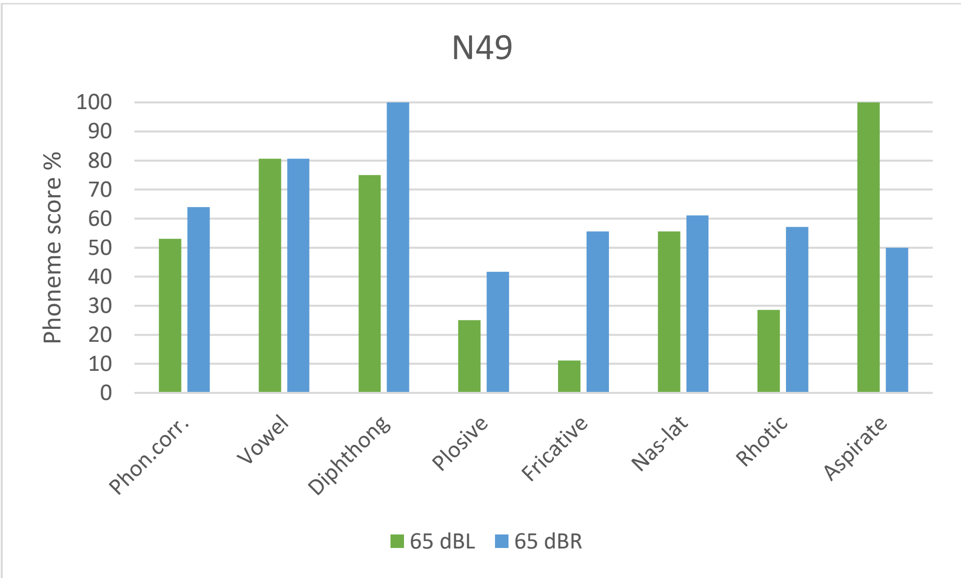


Fig. 18 Phoneme scores per category of participant N49

Participant N49 had a mild hearing loss of 40 dB in the low frequencies up to 2 kHz and a steep sloping moderate high frequency hearing loss up to 75 dB as shown in figure 17. This hearing loss leads to a reduced overall phoneme score of 53% for the left ear and 64% for the right ear, which can be explained by the differences in the audiogram. At conversation level the vowel scores are reduced to 80% (figure 18) which complies with the low frequency loss. The score of the plosives which represent the 500 Hz to 1500 Hz region is also lower. This may be caused by the plosives at the word initial and word medial, which have less energy than the word final plosives, as can be seen in the consonantal speech banana in appendix 2. The high frequency loss in the left ear at 4 kHz is 10 dB greater than at the right ear. This explains the lower fricative score in the left ear and it again reflects the presumed sensitivity of this test.

4.7 Results student sub-group

Group 1 was defined as a group of normal hearing participants with a hearing threshold better than $30dB_{HL}$. This still cannot be considered as a homogenous group. Therefore, a sub-group was selected which comprised of 8 young university students with an average age of 20;4 years. For this group, near to maximum phoneme scores on the NAMES® test would be expected.

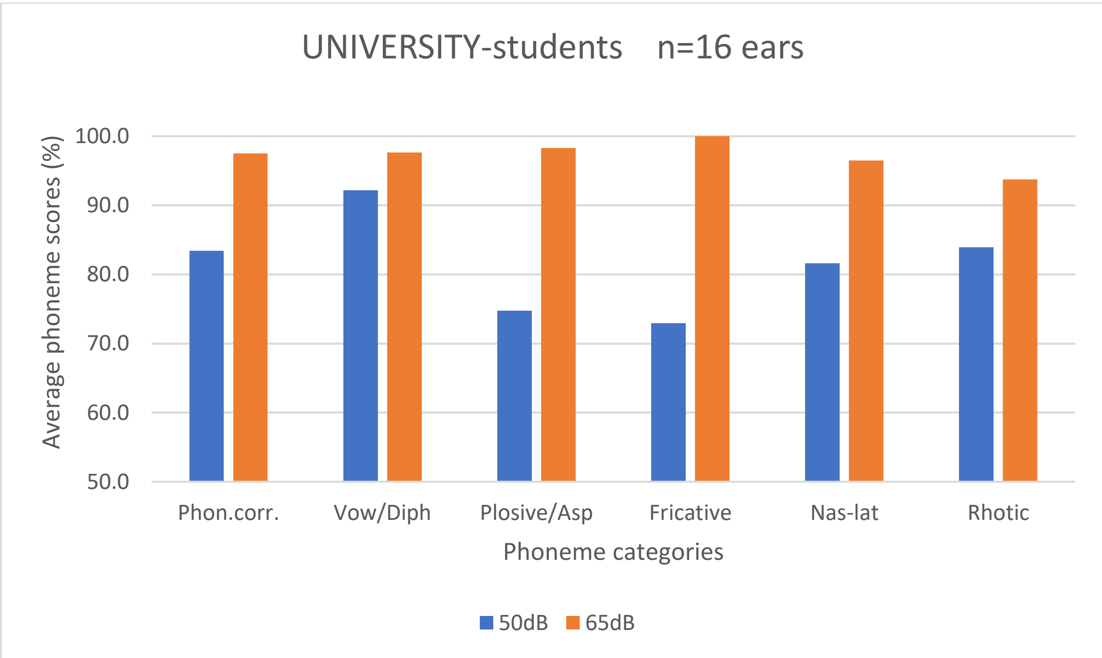


Fig. 19 NAMES® average phoneme scores subgroup of University students

The graph in figure 19 gives the average student's phoneme scores at 50dB and 65dB presentation levels. Only the fricative score at 65dB shows a ceiling effect. Comparing these results with the scores of group 1 in figure 10 shows that this group of students' overall score was better by about 3% at 65dB but shows larger differences at 50dB. This indicates that this test is sensitive, but a larger group of participants will be needed for reliable norm data.

4.8 Examiners reliability

To investigate the inter-rater reliability, audio recordings were made of most of the tests. From these recordings 6 lists at 50 dB level were selected. Three were from male and three from female participants. Three audiology professionals (T2, T3, T4) scored these lists, presented from a computer speaker at normal conversation level. Average score results of the group professionals compared to the scores of the researcher (T1) as presented in figure 20 show that overall the researcher was scoring about 10% more favourable than the professionals. This cannot be explained by the hearing acuity of the researcher. Their audiograms are in appendix 5. Possible explanations are:

- researcher's familiarity with the words which biases the expectation of the participant's response
- the live presentation mode, where the researcher, made use of speech reading
- interpretation on how to score

Bosman (1995) states that nonsense syllables are well suited for analytic testing using either an open or closed-response format, but he indicates that their use as test items require that the examiners should be thoroughly trained, as naïve listeners tend to respond with sense words.

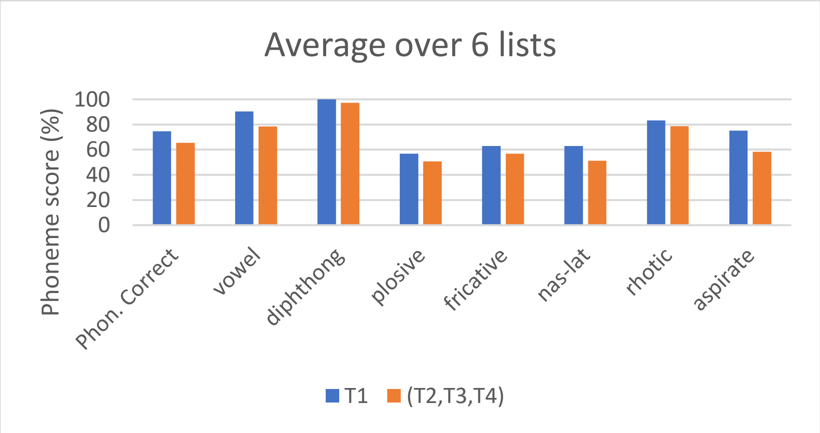


Fig. 20 Comparison phoneme scores researcher with other examiners

As a result of these findings, I recommend drafting of proper instructions for the examiners. It should be noted that presentations of this test at supra-threshold level, are easier to score. At that level participants are more secure in their responses and they are louder with their utterances.

4.9 Dispersion of the PRS scores

PRS	Group 1-65dB	Students – 65dB	Group 1-50dB	Students-50dB
Mean	96,56	97,50	78,46	83,44
Median	97	98	80	85
Quartile 1	95	95,75	73,25	80,5
Quartile 3	99	99	85	88,25
IQR	4	3,25	11,75	7,75
SD	2,81	2,03	12,56	7,72
MIN	86	93	22	65
MAX	100	100	97	97

Table 10 Average and dispersion values of the PRS scores

All participants of Group 1 met the criterion of having a better hearing threshold than $30dB_{HL}$ but they did not have the same hearing acuity. This results in a dispersion of their phoneme recognition scores. As shown in table 10. Interquartile range (IQR) and standard deviations (SD) have been calculated for group 1 and for the sub-group of university students. In general SD should be smaller than IQR. This is not the case in Group 1 at 50 dB. This probably can be explained by a few outliers. IQR in this case will be a better indicator for the dispersion of the data, because it will be less affected by the outliers.

Due to the homogeneity of the group, the dispersion of the phoneme scores of the university students as shown in figures 24 and 26 was slightly less than that of Group 1. However, the difference between the IQR's of both groups gets larger when the presentation level is lower. This clearly indicates an influence of hearing loss on the dispersion of the scores. When collecting age related norm data, it therefore will be recommended to divide the group in several age categories, for example in 10-year age groups.

At both presentation levels there was one outlier, participant N52, marked by the red circles in figures 23 and 25. The audiogram of this participant is shown in figure 21.

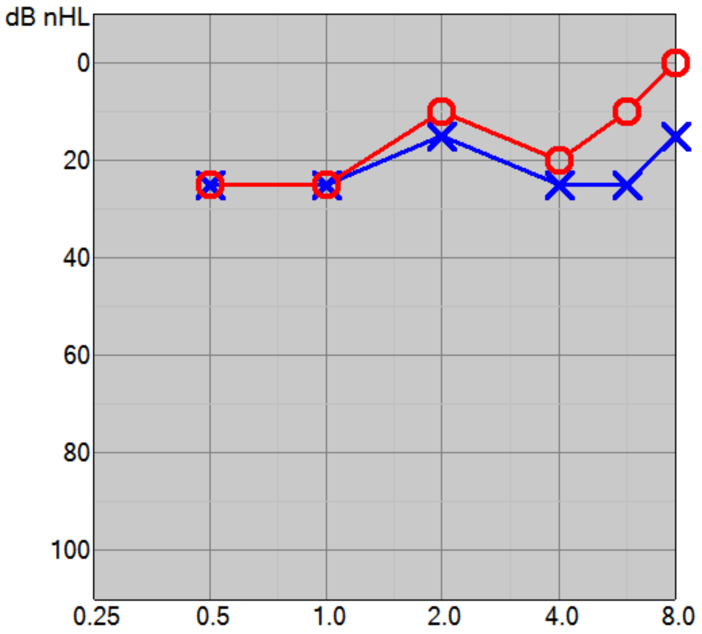


Fig. 21 Audiogram of participant N52

This outlier cannot be explained by the audiogram. In group 1 there were several participants with similar losses who scored better. The presentation order was, first the right ear and then the left ear. The hearing level of this participant's right ear in the higher frequencies is between 5 and 15 dB better than in the left ear. At the 65dB presentation level however the score for the right ear is worse than the left ear. On the other hand, at the 50dB presentation level the score for the right ear is better. A possible explanation for this inconsistency and poor results might be an auditory processing disorder. This is the only participant in this research where this observation occurred, and therefore we must be a bit careful with conclusions.

In figure 25 another outlier, marked by the green circle for participant N48 was noted. The participant first scored 44% for his right ear and then 74% for his left ear. This participant has a normal hearing, (see figure 22), and therefore this outlier cannot be justified. The scores from this participant at 65dB were within normal range, as was the score for 50dB for his left ear. The most likely cause for this outlier is a dip in the attention of the participant.

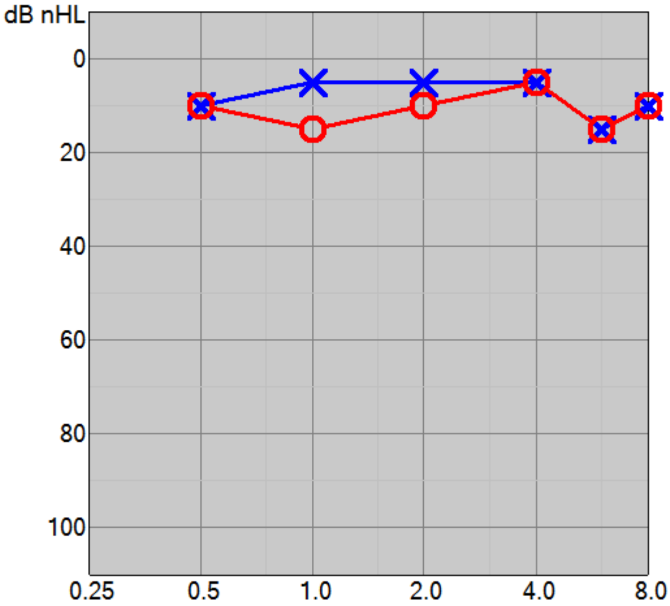


Fig. 22 Audiogram of participant N48

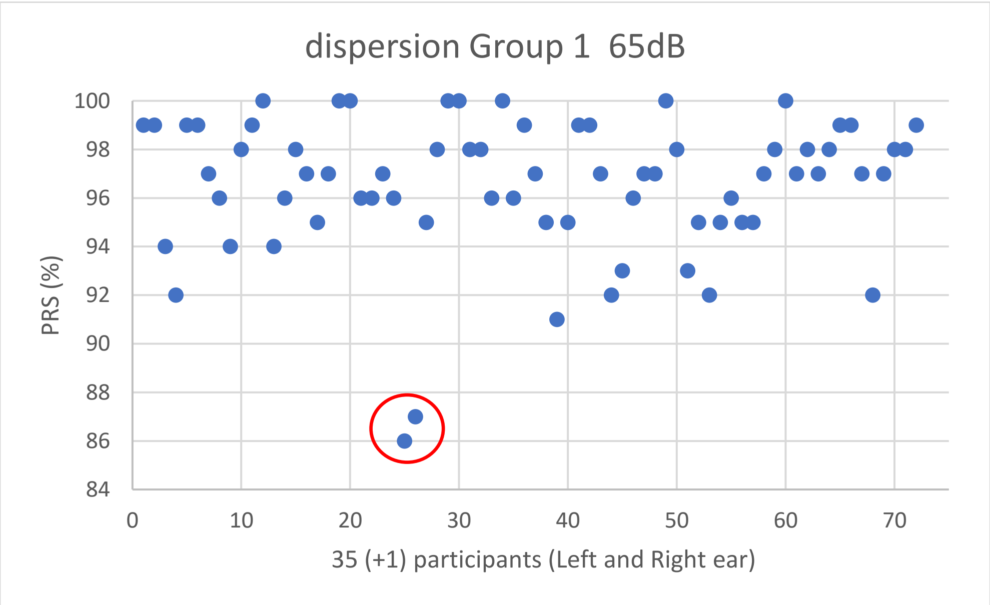


Fig. 23 Dispersion of Phoneme Scores Group 1 at 65dB

When list 4 was introduced in this experiment, results of one participant were used to cross check the outcomes with list 3. Scores were within 2% and this confirmed that we could proceed with that list. It explains the two extra data points in the graph of figure 23.

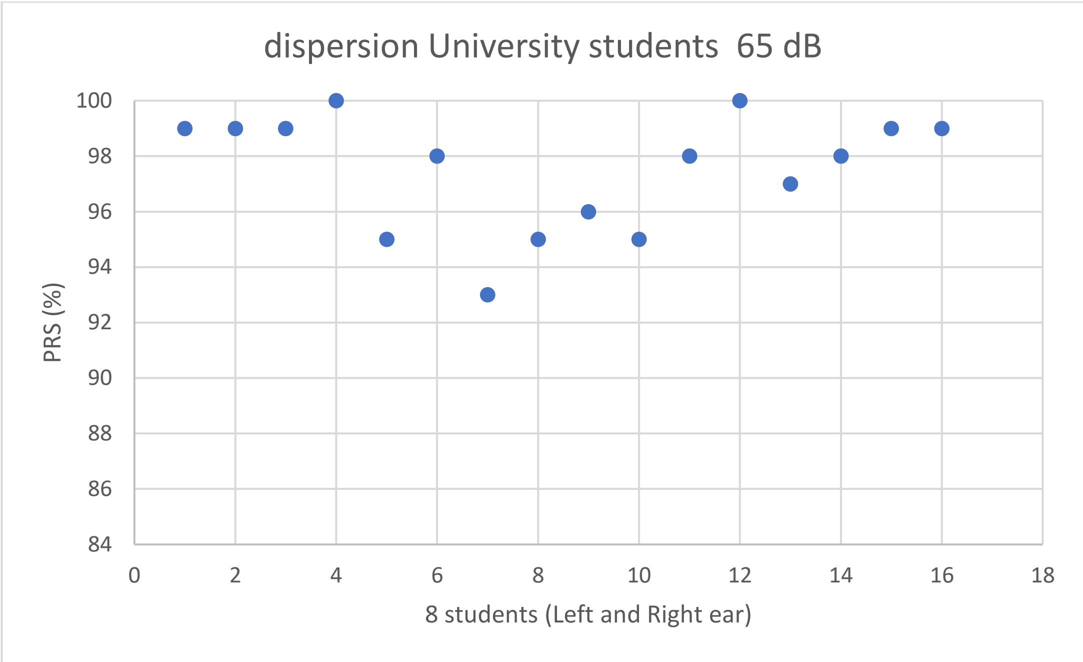


Fig. 24 Dispersion of Phoneme Scores University students at 65dB

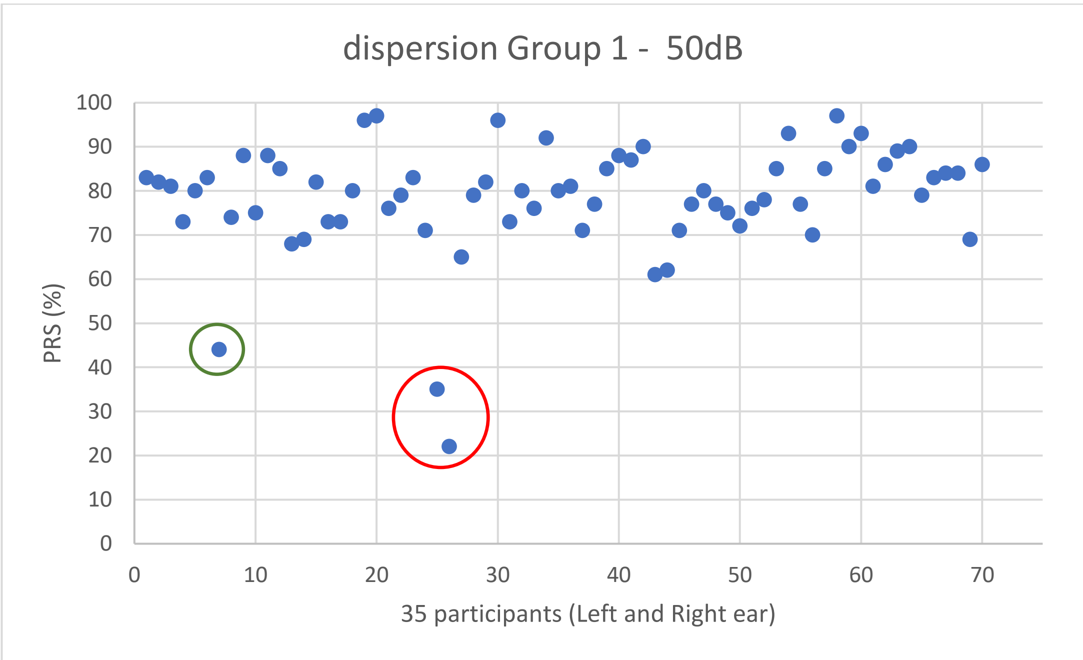


Fig. 25 Dispersion of Phoneme Scores Group 1 at 50dB

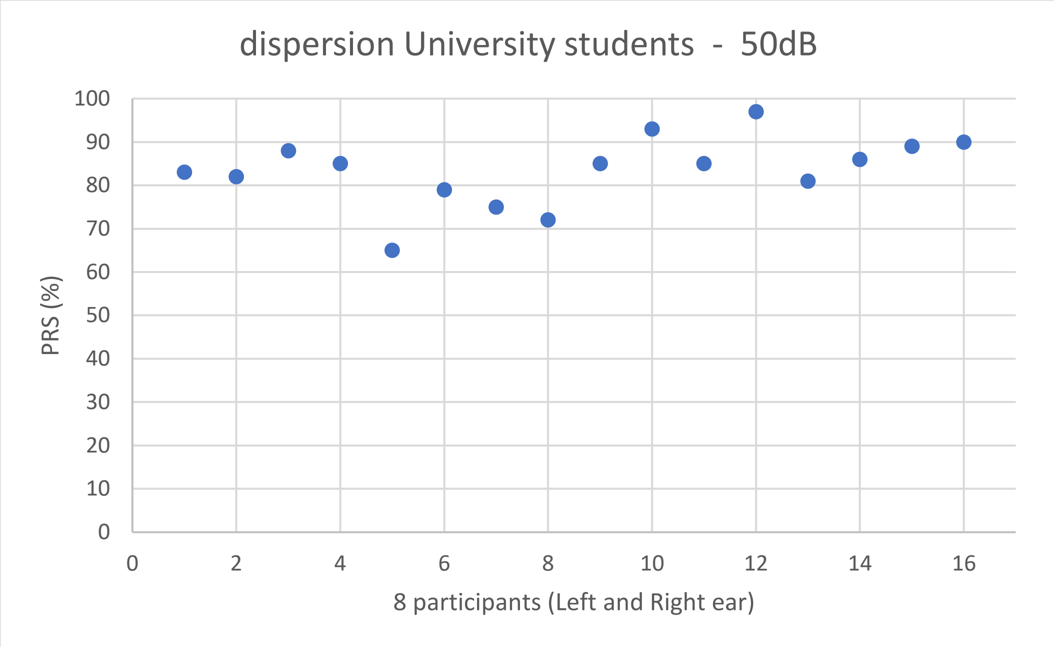


Fig. 26 Dispersion of Phoneme Scores University students at 50dB

4.10 Regional effects

When results from the three regions were compared, no evidence was found for serious effects of regional dialects. A suspicion that the word “monɔt” could be linked to the German word “monat” (month), could not be proven. There even could be a bias towards the second syllable “nat” which means ‘wet’ in Dutch. The phonemes /ɔ/ and /ɑ/ are very close and it is arguable whether this small difference is relevant to the test.

One important observation was that the phoneme /v/ in the West of the Netherlands is pronounced as an /f/. This can lead to different interpretations in the scoring. In the Netherlands it is common practice in speech tests to score a /v/ correct when /f/ is spoken, especially for participants from the Western regions. Nguyen (2017) found in the Vietnamese version of NAMES® a small dialectical effect (roughly 1,5 %) in the phoneme scores of non-native listeners from another Vietnamese region. According to Nguyen the Northern and Southern Vietnamese dialects are different because of the history and geographic separation. The distance from North to South Vietnam is like that from Northern to Southern Europe (Copenhagen to Rome). In the Netherlands we also have regional dialects. But in this research care should be taken when interpreting the results between the three regions, because of the relatively small number of participants. Also, here the effect appears to be small at the $65dB_{SPL}$ presentation level. Participants from the North scored about 1.9 % worse, from the South 0.4% better and from the West 0.5% worse than average. It is an indication that if at all there are regional dialect effects these probably have a small influence, of less than 2% on the PRS.

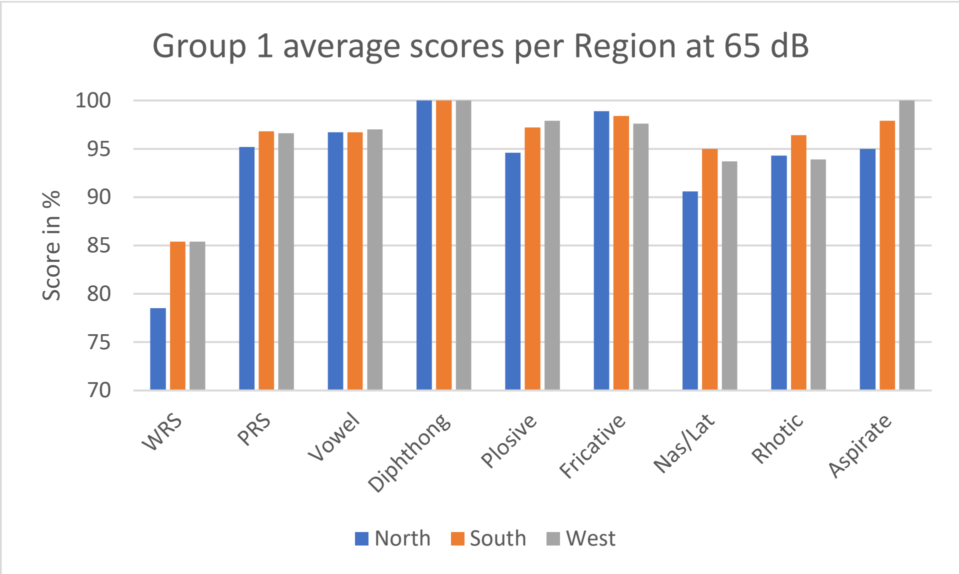


Fig. 27 Average scores per region

The scores across the two regions South and West are quite similar (figure 27). The slightly lower scores for the group from the North can probably be explained by their small number (n=5) in this group. One or two poor scores can have a large effect.

4.11 Conclusion

The Dutch version of NAMES® seems to be a sensitive, quick and reliable test to check phoneme identification. The overall word and phoneme scores are useful indicators for screening. No major unforeseen regional effects of dialects were observed. Recommendations should be made towards the instruction of examiners. For statistical relevance phoneme categories with less than 10 phonemes should be combined. In this Dutch version, I recommend combining the diphthongs with the vowels and the aspirates with the plosives.

5 DISCUSSION

5.1 Introduction

In this section the connection will be made between the research questions, the development of the test and the first experiences. Recommendations will be made for improvements of the test and suggestions will be done for further research.

5.2 Project purpose

The main purpose of this project was adapting the NAMES® phoneme identification test for the Dutch language. The research had three main objectives:

- Identify relevant parameters for the test
- Compose and record the test
- Evaluate the test

5.3 Parameters

Through a literature review, relevant linguistic, phonetic and design parameters for developing a Dutch version of the test were identified, which included:

- Phonemic balancing
- Phonotactic rules
- Standards, calibration and word balancing
- Requirements for testing children
- Recording method
- Scoring options

In most of the literature (Meister, 2005; Bosman and Smoorenburg, 1995; Martin, 1997) it is emphasized that a speech test should be phonemically balanced, to optimally represent the current spoken language. Bosman (1995) argues that phonemic balancing increases the validity of the test for the prediction of speech perception in real-life conditions. Gelfand (2001) however, found that phonemic balancing had little impact on the outcome of speech recognition tests. He questions its clinical relevance. In my opinion it depends on the purpose of the test. Screening might not require an exact representation if the major phonemes and phoneme categories are included. But for assessment of hearing in real-life situations a near to correct representation of the language is needed. In that case I support Bosman's (1995) view. In a NAMES® wordlist with only 20 words, and a total of 100 phonemes it was not possible to include all phonemes. At least most of the phoneme categories are represented by a selection of phonemes with the highest frequency of occurrence. Phonemic balancing, in combination with application of phonotactic rules made the nonsense words sound like natural words.

NAMES® was designed as a speech screening tool for testing children. Theoretically nonsense words should be independent of the child's language level, but for young children we have to consider the stage of their phonemic awareness development and their phonological memory (Anthony and Francis, 2005) and the child should be able to reproduce the sound (Oller, Oller and Badon, 2014). Young children were not included in this research on purpose, so that these factors do not play a role in this evaluation of the design. In the literature review however, there was a special focus on requirements of speech tests for children. Several arguments were found (Jusczyk *et al.*, 1993; Jusczyk, Luce and Charles-Luce, 1994; Gaygen, 1997; Pitt and McQueen, 1998) why phonotactic rules are important for speech recognition in children, even when using nonsense words. This, together with the importance of natural sounding test-words were reasons to investigate the Dutch phonotactic rules.
The literature review was also used to relate the NAMES® test to other existing linked tests. This revealed that NAMES® has its unique purpose and position and can complement other tests.

5.4 Scoring and score reliability

Even though that in this research all testing was done under quiet conditions and that all participants were tested by the same examiner, there were several factors that could have influenced the scoring:

- The hearing acuity of the examiner
- Bias in scoring towards know test words
- Unclear articulation of the participants at lower presentation levels

- Keyboard confusions among vowels and long vowels e.g. /I/ and /i/, /ɔ/ and /o:/, and between /ɛ/ and /ə/
- Regional pronunciation of /v/ as /f/

NAMES® words are presented as CVCVC combinations. If clients respond with other sequences it was not always obvious on how to score. Clear instructions and training of the examiners is necessary, and an unambiguous scoring instruction must be made. It would be worthwhile to investigate whether there is a large bias effect when the word is presented on the screen and the examiner just selects the correct phonemes. That method will limit the response possibilities but might improve the reliability. Scoring reliability is a critical factor for comparing results. Scoring can be made more independent of an examiner by using phoneme recognition algorithms. Prof. Strik from Radboud University Nijmegen (H. Strik 2018, personal communication 2 October) affirmed that in the last decade speech to text recognition software has made an enormous progress, however detection of individual phonemes is still very difficult because they relate very much to the position in the word and the combination with other phonemes in the syllable, as the researcher also found in his literature review (Booij, 1995; Köhnlein and Linke, 2018b). According to Strik no useful software is currently available, but promising research is going on. He referred to one of his earlier articles with experiments in this area (Wester *et al.*, 2001).

The interactive keyboard, which only shows the consonants or vowels depending on the score position is very efficient. It is user friendly towards examiners who are not acquainted with the IPA symbols, though it needs explanation. An optional IPA keyboard for professional examiners should be considered. A combined score key for phonemes ‘v’ and ‘f’ will solve the score confusion between these two phonemes. This can be allowed because both phonemes are in the same frequency/intensity region as can be seen in Keen’s consonantal speech banana in Appendix 2.

It is observed that elderly people take more time to respond and Atcherson (2015) confirms that even older adults with normal or near to normal hearing sensitivity may exhibit age-related central auditory processing deficits. One older participant in this research missed the first word presented in this list because she did not expect a female speaker. In the regular version of the NAMES® test, this is overcome by including a continuous 'repeating' word before the test starts. NAMES® is a supra-threshold test, where in the regular version the client gets the opportunity to change the intensity to a comfortable level in a range of plus or minus 5 dB around $65dB_{SPL}$. It is advisable to extend this adjustment range in 5 dB steps up to 15 dB (80 dB_{SPL}), although roll-over effects in case of hearing loss might occur.

5.5 NAMES® for children

The NAMES® test is most probably suitable for children, but this has not yet been systematically evaluated. The diversity of factors influencing score results of children was reason not to include them in this research. Literature revealed that there is a clear association between vocabulary size and NWRT performance (Gathercole *et al.*, 1999). The vocabulary of young children is developing, as is their auditory memory. Dawson et al. (2002), found that children with Cochlear Implants perform less on short term memory skills than their hearing peers, when the stimuli are verbally coded. Probably with the current earlier age of implants this has changed. Testing children would also involve assessing their receptive language and nonverbal intelligence. This was beyond the scope of this study but future research with children is recommended. An advantage of NAMES® is that it is fast and can provide specific information about several phoneme categories. It is not yet known how children from different ages and with a different phonological awareness, for example bilingual children respond to the test. More research needs to be done with this group. It will be worthwhile to investigate whether NAMES® can also be used to screen for auditory processing disorders (refer to the discussion on participant N52 in the analysis section). A suggestion to make the NAMES® test less abstract for children is to add one or two carrier words. For example: Mister 'neiriel', Missis 'noomot' or Mister John 'neiriel', Missis Sarah 'noomot'. An extra benefit will be that this will give the hearing aid or cochlear implant (CI) more time to settle its dynamic algorithms. Noise suppression algorithms in hearing aids are typically based on the temporal behaviour of the signal (Kates, 2008). According to Plyler (2005a; 2005b) fast time constants can attenuate the noise between syllables, but can also generate “burbling” artefacts. Slow time constants reduce this effect but slow down the system response; the onset of a speech sound may be attenuated by the noise suppression gain set for the preceding noise level. It may be expected that the use of disyllable words as in NAMES® and the use of carrier words may reduce this effect and provide more realistic outcomes. NAMES® can also be presented in a Free Field setting for hearing aid and CI users. The BELLS® platform can provide the necessary calibration signals.

5.6 Test-retest reliability

Data from 22 Group 1 participants (Appendix 6) with near to symmetrical audiograms was selected to check the test-retest reliability. Their phoneme scores for the first and second presentations (right ear-left ear) at 65dB level were averaged. At 65 dB, presentation 1 resulted in a PRS average of 96.3% and presentation 2 resulted in a score of 96.7%. This 0.4% difference indicates that overall the test is quite repeatable. The absolute differences between two presentations ranged between 0 and 5% in PRS scores, with an average deviation of 1.68%.

From the same group, Cronbachs α was calculated (Appendix 6) for the presentations at both ears for 65 dB and 50 dB intensity levels, to check the correlation between the two measurements. This resulted in: $\alpha 1(65dB) = 0.7825$ and $\alpha 2(50dB) = 0.7159$, which can be considered acceptable values. Here we should note that the data was compared from different ears which were considered equal. Comparing only two measurements per participant also results in a Cronbach α value which is on the lower side.

5.7 Dialects

Speech tests are prone to regional dialects (Lyregaard, 1997). From literature (Nerbonne *et al.*, 1996) and experiences in the Netherlands it is acknowledged that some phonemes like /r/, /z/ and /v/ are pronounced differently depending on the region. This was a factor which had to be considered in the design of the test. It was possible that word scores would be influenced by a regional dialect. That however could not be confirmed by our data. This suggests that the design is solid and that this version of NAMES® can be used across the Netherlands.

5.8 General conclusions

A comparative analysis of the results from participants with hearing loss indicates that the NAMES test is very sensitive. Small differences in hearing loss between both ears already results in differences in phoneme scores, and the scores in phoneme categories relate to the frequency regions of the loss. For example, high frequency loss correlates with lower phoneme scores in the fricative category, low frequency losses correlate with lower vowel scores.

The scores of the 3 examiners in the inter-rater reliability check, differed slightly with those from the principal investigator. A bias towards the correct phonemes is more apparent with the principal investigator who was more acquainted with the test. This had a small impact on the score totals, but relative differences of outcomes between categories remained. Therefore, it can be concluded that absolute outcomes can be slightly different, but the relative outcomes appear still to be useful, especially for screening. Clear instructions on how to score must be defined to get more identical results. The feature in BELLS® to record the client's responses for a check afterwards can be useful, especially when testing multilingual clients. It is expected that automatic scoring, which is under development, will provide more stable results. Additional research in that area therefore is advised.

An added value of the NAMES® design is that the test words consist of two syllables, which makes the test suitable for testing with modern hearing aids, cochlear implants and Bone Conduction Devices (BCD's). The first syllable of the NAMES® word can make the hearing aid settle and then the second syllable can provide the correct scores. This area also needs further exploration. NAMES® is flexible in providing scores at different phoneme positions, and that information can be used to provide suggestions for hearing aid adjustment. Data from different tests in BELLS® can be combined for extensive analysis and age appropriate tests can be selected, to optimize the results.

NAMES®, as part of the BELLS® test battery provides fast and adequate data for validation of modern hearing aid validation and phoneme diagnostics for speech and language therapy. Besides this there are reasons to investigate the use of NAMES® in screening for Auditory Processing Disorders.

6 REFERENCES

Akeroyd, M. A. (2008) 'Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults', *International journal of audiology,* 47(sup2), pp. S53-S71.

ANSI (2009) 'S3. 22-2009, American National Standard Specification of Hearing Aid Characteristics', *New York: American National Standards Institute*.

Anthony, J. L. and Francis, D. J. (2005) 'Development of Phonological Awareness', *Current Directions in Psychological Science,* 14(5), pp. 255-259.

Atcherson, S. R., Nagaraj, N. K., Kennett, S. E. W. and Levisee, M. (2015) 'Overview of Central Auditory Processing Deficits in Older Adults', *Seminars in hearing,* 36(3), pp. 150-161.

Babbie, E. R. (2013) *The basics of social research.* Boston MA,: Cengage learning.

Ball, E. W. and Blachman, B. A. (1991) 'Does Phoneme Awareness Training in Kindergarten Make a Difference in Early Word Recognition and Developmental Spelling?', *Reading Research Quarterly,* 26(1), pp. 49-66.

Bess, F. (1983) 'Clinical assessment of speech recognition.', in Konkle, D.F. and Rintelmann, W.F. (eds.) *Principles of Speech Audiometry*. Baltimore: Univ Park Press, pp. 127-202.

Billings, C. J., Penman, T. M., Ellis, E. M., Baltzell, L. S. and McMillan, G. P. (2016) 'Phoneme and Word Scoring in Speech-in-Noise Audiometry', *American journal of audiology,* 25(1), pp. 75-83.

Booij, G. (1978) *Fonotactische restricties in de generatieve fonologie.* Spektator.

Booij, G. (1995) *The Phonology of Dutch. The Phonology of the World's Languages.* New York: Clarendon Press.

Bosman, A. J. and Smoorenburg, G. F. (1995) 'Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment', *Audiology,* 34(5), pp. 260-284.

Brunner, M. and Stuhrmann, N. (2013) 'Phonemdiskrimination und RechtschreibschwächePhoneme discrimination and dyslexia', *HNO,* 61(1), pp. 65-70.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J. and Lui, C. (1994) 'An international comparison of long-term average speech spectra', *The journal of the acoustical society of America,* 96(4), pp. 2108-2120.

Cameron, S., Chong-White, N., Mealings, K., Beechey, T., Dillon, H. and Young, T. (2018) 'The Phoneme Identification Test for Assessment of Spectral and Temporal Discrimination Skills in Children: Development, Normative Data, and Test-Retest Reliability Studies', *J Am Acad Audiol,* 29(2), pp. 135-150.

Chiat, S., Armon-Lotem, S., de Jong, J. and Meir, N. (2015) 'Nonword repetition', *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*, pp. 125-150.

Coady, J. A. and Evans, J. L. (2008) 'Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI)', *International Journal of Language & Communication Disorders,* 43(1), pp. 1-40.

Collins, B. and Mees, I. (2003) *The phonetics of English and Dutch.* Leiden: Brill.

Coninx, F. 2018a. Battery for the Evaluation of Language and Listening Skills (BELLS) Version 6.6.5. Solingen: Institute for Audio Pedagogics (IfAP).

Coninx, F. 2018b. Excel NAMES2 generator. Solingen: Institut für Audiopaedagogik.

Cooke, M., Lecumberri, M. L. G., Scharenborg, O. and Van Dommelen, W. A. (2010) 'Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages', *Speech Communication,* 52(11-12), pp. 954-967.

Cool Edit Pro 2.1 2003. Digital audio workstation. Syntrillium Software.

Dawson P.W., Busby P. A., McKay C. M. and M., C. G. (2002) 'Short-Term Auditory Memory in Children Using Cochlear Implants and Its Relevance to Receptive Language', *Journal of Speech, Language, and Hearing Research,* 45(4), pp. 789-801.

Dell, G. S., Reed, K. D., Adams, D. R. and Meyer, A. S. (2000) 'Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 26(6), pp. 1355.

Dermody, P., Katsch, R. and Mackie, K. 1983. Amplitude normalisation techniques for speech intelligibility testing. *Proceedings of the 11th International Congress of Acoustics.*

Dierks, A., Seibert, A., Brunner, M., Körkel, B., Haffner, J., Strehlow, U., Parzer, P. and Resch, F. (1999) 'Testkonstruktion, -analyse und Erprobung des Heidelberger Lautdifferenzierungstests zur auditiv-kinästhetischen Wahrnehmungstrennschärfe (HD-LT)', *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie,* 27(1), pp. 29-36.

DIN 45621-1 1995. Word lists for recognition tests - Part 1: Monosyllabic and polysyllabic words. Berlin: German Institute for Standardization.

Dirks, D. D., Kamm, C., Bower, D. and Betsworth, A. (1977) 'Use of performance-intensity functions for diagnosis', *Journal of Speech and Hearing Disorders,* 42(3), pp. 408-415.

Dreschler, W. A. (1989) 'Phoneme perception via hearing aids with and without compression and the role of temporal resolution', *Audiology,* 28(1), pp. 49-60.

Engel, P. M. J., Santos, F. H. and Gathercole, S. E. (2008) 'Are working memory measures free of socioeconomic influence?', *Journal of Speech, Language, and Hearing Research,* 51(6), pp. 1580-1587.

Gathercole, S. E. (2006) 'Nonword repetition and word learning: The nature of the relationship', *Applied Psycholinguistics,* 27(4), pp. 513-543.

Gathercole, S. E. and Baddeley, A. D. (1989) 'Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study', *Journal of memory and language,* 28(2), pp. 200-213.

Gathercole, S. E., Hitch, G. J. and Martin, A. J. (1997) 'Phonological short-term memory and new word learning in children', *Developmental psychology,* 33(6), pp. 966.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M. and Martin, A. J. (1999) 'Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship', *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition,* 13(1), pp. 65-77.

Gaygen, D. (1997) 'The effects of probabilistic phonotactics on the segmentation of continuous speech', *Unpublished doctoral dissertation, University at Buffalo, Buffalo, NY*.

Gelfand, S. A. (2001) *Essentials of audiology.* 2nd edn. New York: Thieme.

Gelfand, S. A., Piper, N. and Silman, S. (1986) 'Consonant recognition in quiet and in noise with aging among normal hearing listeners', *The Journal of the Acoustical Society of America,* 80(6), pp. 1589-1598.

Gordon-Salant, S. (2005) 'Hearing loss and aging: new research findings and clinical implications', *Journal of Rehabilitation Research & Development,* 42.

Gouskova, M. (2004) 'Relational hierarchies in Optimality Theory: the case of syllable contact', *Phonology,* 21(2), pp. 201-250.

Govaerts, P. J., Schauwers, K. and Gillis, S. (2002) 'Language acquisition in very young children with a cochlear implant: Introduction', *Antwerp papers in Linguistics*.

Gussenhoven, C. (1999) *Illustrations of the IPA: Dutch. Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* Cambridge: Cambridge University Press, p. 74-77.

Hahlbrock, K.-H. (1970) *Sprachaudiometrie.* 2nd edition edn. Stuttgart: Georg Thieme Verlag.

International Organization for Standardization (ISO) 2012. ISO 8253-3: Acoustics: Audiometric test methods. *Part 3: Speech Audiometry.* Geneva: ISO.

Jerger, J. and Jerger, S. (1971) 'Diagnostic significance of PB word functions', *Archives of Otolaryngology,* 93(6), pp. 573-580.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y. and Jusczyk, A. M. (1993) 'Infants' sensitivity to the sound patterns of native language words', *Journal of memory and language,* 32(3), pp. 402.

Jusczyk, P. W., Luce, P. A. and Charles-Luce, J. (1994) 'Infants' sensitivity to phonotactic patterns in the native language', *Journal of Memory and Language,* 33(5), pp. 630.

Kates, J. M. (2008) *Digital hearing aids.* San Diego CA,: Plural publishing.

Keen, P. 2014. New Consonantal Speech Banana. peter.keenhearing@btinternet.com.

Kohnert, K. (2010) 'Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions', *Journal of communication disorders,* 43(6), pp. 456-473.

Kombo, D. K. and Tromp, D. L. (2006) *Proposal and thesis writing: An introduction.* Nairobi: Paulines Publications Africa.

Kramer, S. and Brown, D. K. (2018) *Audiology: science to practice.* Plural Publishing.

Köhnlein, B. (2018) *Primary stress in simplex words.* Available at: http://taalportaal.org/taalportaal/topic/pid/topic-14020545841607206 (Accessed: 1 October 2018).

Köhnlein, B. and Linke, K. (2018a) *fig_syllable.png*: taalportaal.org. Available at: http://taalportaal.org/taalportaal/topic/pid/topic-13998813314549851 (Accessed: 30 September 2018).

Köhnlein, B. and Linke, K. (2018b) *Phonotactics*: Taalportaal. Available at: http://taalportaal.org/taalportaal/topic/pid/topic-13998813314549851 (Accessed: 30 September 2018).

Lawson, G. and Peterson, M. (2011) *Speech audiometry. Core clinical concepts in audiology* San Diego: Plural Pub.

Ling, D. 'The Ling six-sound test'. *Proceeding of the 2002 Alexander Graham Bell Convention*.

Linke, K. and Oostendorp, M. v. (2018a) *Segment frequency of consonants in Dutch.*: Taalportaal. Available at: http://taalportaal.org/taalportaal/topic/pid/topic-14071525824219919. (Accessed: 1 October 2018).

Linke, K. and Oostendorp, M. v. (2018b) *Segment frequency of vowels in Dutch.*: Taalportaal. Available at: http://taalportaal.org/taalportaal/topic/pid/topic-14071527307009001. (Accessed: 1 October 2018).

Lyregaard, P. (1997) 'Towards a theory of speech audiometry tests', in Martin, M. (ed.) *Speech audiometry.* 2nd ed. London ; New York: Whurr Publishers, pp. xii, 368 p.

Madell, J. R., Klemp, E. J., Batheja, R. and Hoffman, R. (2011) 'Evaluating speech perception performance', *Audiology Today,* 23(5), pp. 52-57.

Mancini, P., Bosco, E., D'Agosta, L., Traisci, G., Nicastri, M., Giusti, L. and Musacchio, A. (2010) 'Testing auditory skills in children CI users: is phonemic discrimination related to acoustic variables only?', *Cochlear Implants Int,* 11 Suppl 1, pp. 332-5.

Maniwa, K., Jongman, A. and Wade, T. (2008) 'Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners', *The Journal of the Acoustical Society of America,* 123(2), pp. 1114-1125.

Markides, A. (1978) *Whole-Word Scoring Versus Phoneme Scoring in Speech Audiometry.*

Maroonroge, S. and Diefendorf, A. O. (1984) 'Comparing normal hearing and hearing-impaired subject's performance on the Northwestern Auditory Test Number 6, California Consonant Test, and Pascoe's High-Frequency Word Test', *Ear and hearing,* 5(6), pp. 356-360.

Martin, F. N., Champlin, C. A. and Perez, D. D. (2000) 'The question of phonetic balance in word recognition testing', *Journal-american academy of audiology,* 11(9), pp. 509-513.

Martin, M. (1997) *Speech audiometry.* 2nd edn. San Diego, Calif.: Singular Pub. Group.

Meister, H. 'Kindertests in der klinik. [Tests for children in clinical practice - An overview of speech audiometric procedures]'. *8th Annual Conference of the Deutsche Gesellschaft für Audiologie, Goetingen February 24-26, 2005.*

Miller, E. M., Bergeron, J. P. and Connor, C. M. (2008) 'Emergent literacy skills during early childhood in children with hearing loss: Strengths and weaknesses', *Volta,* 108(2), pp. 91-114.

Nerbonne, J., Heeringa, W., Van den Hout, E., Van der Kooi, P., Otten, S. and Van de Vis, W. 'Phonetic distance between Dutch dialects'. *CLIN VI: proceedings of the sixth CLIN meeting*, 185-202.

Nguyen, Q.-D. (2017) *Designs of Speech Audiometric Tests in Vietnamese – The Issues of Normative Values, Dialectal Effects, and Tonal Patterns.* PhD Text, University of Cologne, Cologne [Online] Available at: https://kups.ub.uni-koeln.de/7735/ (Accessed: 15-7-2018).

Nickisch, A., Heuckmann, C., Burger, T. and Massinger, C. (2006) 'Münchner Auditiver Screeningtest für Verarbeitungs- und Wahrnehmungsstörungen (MAUS)', *Laryngo-Rhino-Otol,* 85(04), pp. 253-259.

Oller, J. W., Oller, S. D. and Badon, L. C. (2014) *Milestones : normal speech and language development across the life span.* Second Edition. edn. San Diego: Plural Publishing, Inc.

Oostendorp, M. (2018) *Taalportaal: Segment inventory*: Degemination. Available at: http://taalportaal.org/taalportaal/topic/pid/topic-14020545827588509 (Accessed: 1 October 2018).

Owren, M. J. and Cardillo, G. C. (2006) 'The relative roles of vowels and consonants in discriminating talker identity versus word meaning', *The Journal of the Acoustical Society of America,* 119(3), pp. 1727-1739.

Paglialonga, A., Tognola, G. and Grandori, F. (2014) 'A user-operated test of suprathreshold acuity in noise for adult hearing screening: the SUN (Speech Understanding in Noise) test', *Computers in biology and medicine,* 52, pp. 66-72.

Phonak, A. G. 2014. Phoneme Perception Test 2.1. Switserland: Phonak A.G.

Pitt, M. A. and McQueen, J. M. (1998) 'Is compensation for coarticulation mediated by the lexicon?', *Journal of Memory and Language,* 39(3), pp. 347-370.

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

Plomp, R. and Mimpen, A. (1979b) 'Speech-reception threshold for sentences as a function of age and noise level', *The Journal of the Acoustical Society of America,* 66(5), pp. 1333-1342.

Plyler, P. N., Hill, A. B. and Trine, T. D. (2005a) 'The effects of expansion on the objective and subjective performance of hearing instrument users', *Journal of the American Academy of Audiology,* 16(2), pp. 101-113.

Plyler, P. N., Hill, A. B. and Trine, T. D. (2005b) 'The effects of expansion time constants on the objective performance of hearing instrument users', *Journal of the American Academy of Audiology,* 16(8), pp. 614-621.

Pépiot, E. (2015) 'Voice, speech and gender:. male-female acoustic differences and cross-language variation in english and french speakers', *Corela. Cognition, représentation, langage,* (HS-16).

Scheele, A. F., Leseman, P. P. and Mayo, A. Y. (2010) 'The home language environment of monolingual and bilingual children and their language proficiency', *Applied Psycholinguistics,* 31(1), pp. 117-140.

Seo, M. (2011) *Syllable contact. The Blackwell companion to phonology* Hoboken NJ: Wiley-Blackwell.

Skinner, M. W. (1988) *Hearing aid evaluation.* Upper Sadlle River NJ: Prentice Hall.

Smits, C., Goverts, T. and Festen, J. M. (2013) 'The digits-in-noise test: assessing auditory speech recognition abilities in noise', *J Acoust Soc Am,* 133(3), pp. 1693-706.

Smits, C. and Houtgast, T. (2005) 'Results from the Dutch speech-in-noise screening test by telephone', *Ear and hearing,* 26(1), pp. 89-95.

Smits, C., Kapteyn, T. S. and Houtgast, T. (2004) 'Development and validation of an automatic speech-in-noise screening test by telephone', *International journal of audiology,* 43(1), pp. 15-28.

Storkel, H. L. (2001) 'Learning new words: Phonotactic probability in language development', *Journal of Speech, Language, and Hearing Research,* 44(6), pp. 1321-1337.

Storkel, H. L. (2003) 'Learning new words II: Phonotactic probability in verb learning', *Journal of Speech, Language, and Hearing Research,* 46(6), pp. 1312-1323.

Thordardottir, E., Rothenberg, A., Rivard, M.-E. and Naves, R. (2006) 'Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages?', *Journal of Multilingual Communication Disorders,* 4(1), pp. 1-21.

UNDESA (2014) 'World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs', *Population Division, United Nations,* 32.

Vitevitch, M. S. (2002a) 'Influence of onset density on spoken-word recognition', *Journal of Experimental Psychology: Human Perception and Performance,* 28(2), pp. 270.

Vitevitch, M. S. (2002b) 'The influence of phonological similarity neighborhoods on speech production', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 28(4), pp. 735.

Vitevitch, M. S., Armbrüster, J. and Chu, S. (2004) 'Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 30(2), pp. 514.

Vitevitch, M. S. and Luce, P. A. (2005) 'Increases in phonotactic probability facilitate spoken nonword repetition', *Journal of memory and language,* 52(2), pp. 193-204.

Watson, T. J. (1957) 'Speech Audiometry in children.', in Ewing, A.W.G. (ed.) *Educational guidance and the deaf child*. Manchester: Manchester university press.

Wester, M., Kessens, J. M., Cucchiarini, C. and Strik, H. (2001) 'Obtaining phonetic transcriptions: A comparison between expert listeners and a continuous speech recognizer', *Language and Speech,* 44(3), pp. 377-403.

WHO (2013) *10 Facts on Deafness*. Geneva: World Health Organization. Available at: http://www.who.int/features/factfiles/deafness/facts/en/index6.html (Accessed: 14 September 2018).

Wikimedia Commons (2018) *Extended IPA chart 2005*: Wikimedia Commons, the free media repository. Available at: https://commons.wikimedia.org/w/index.php?title=File:Extended_IPA_chart_2005.svg&oldid=301955262 (Accessed: 24 March 2019).

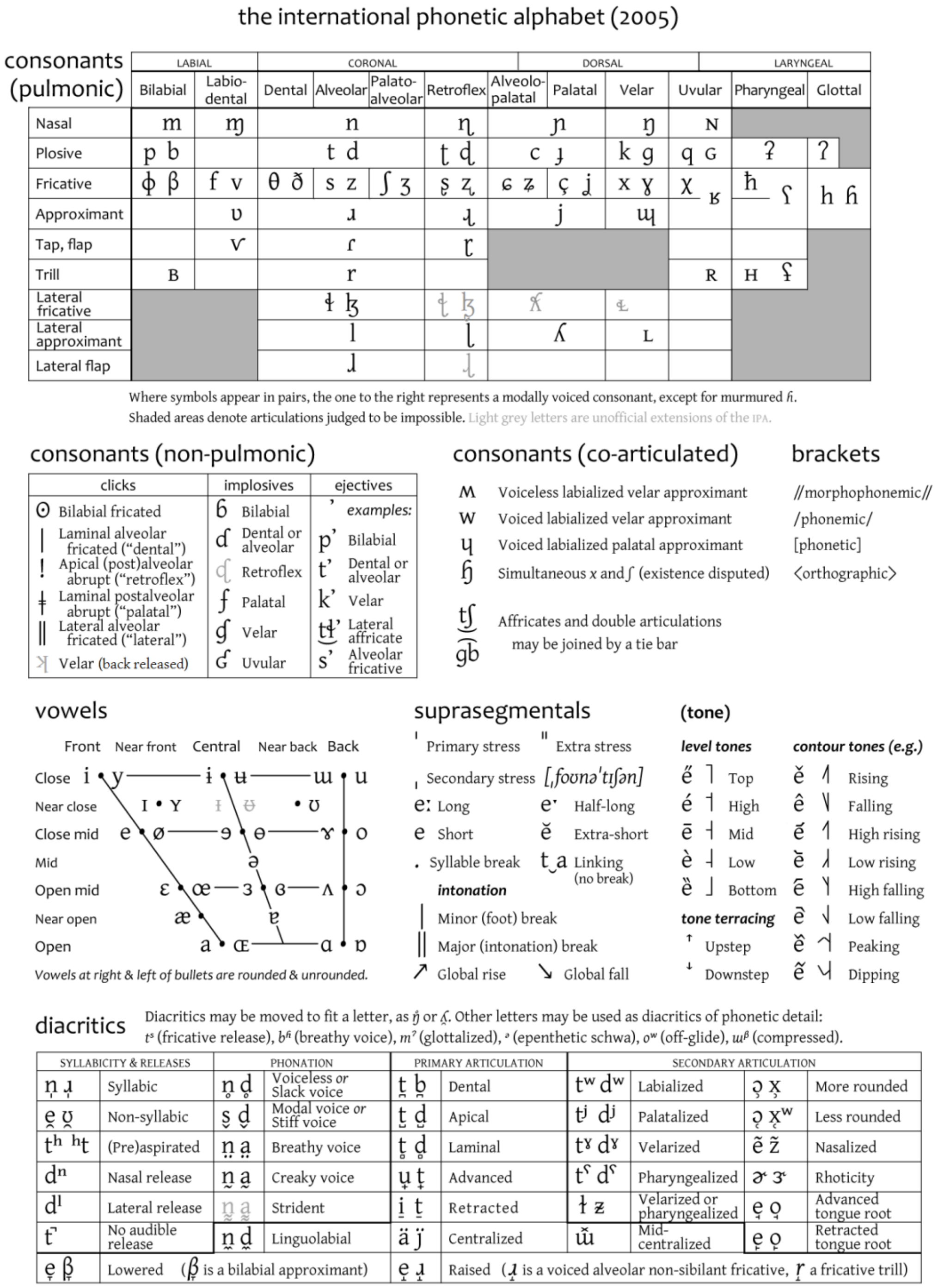
Wikipedia (2019) *Bantu languages --- {Wikipedia}, The Free Encyclopedia*. Available at: https://en.wikipedia.org/wiki/Bantu_languages (Accessed: 14 April 2019).

Zsiga, E. C. (2012) *The sounds of language: An introduction to phonetics and phonology.* Hoboken NJ: John Wiley & Sons.

Zuidema, W. (2009) *A syllable frequency list for Dutch*. Amsterdam: Universiteit van Amsterdam. Available at: https://eprints.illc.uva.nl/379/1/PP-2009-50.text.pdf (Accessed: 1 October 2018).

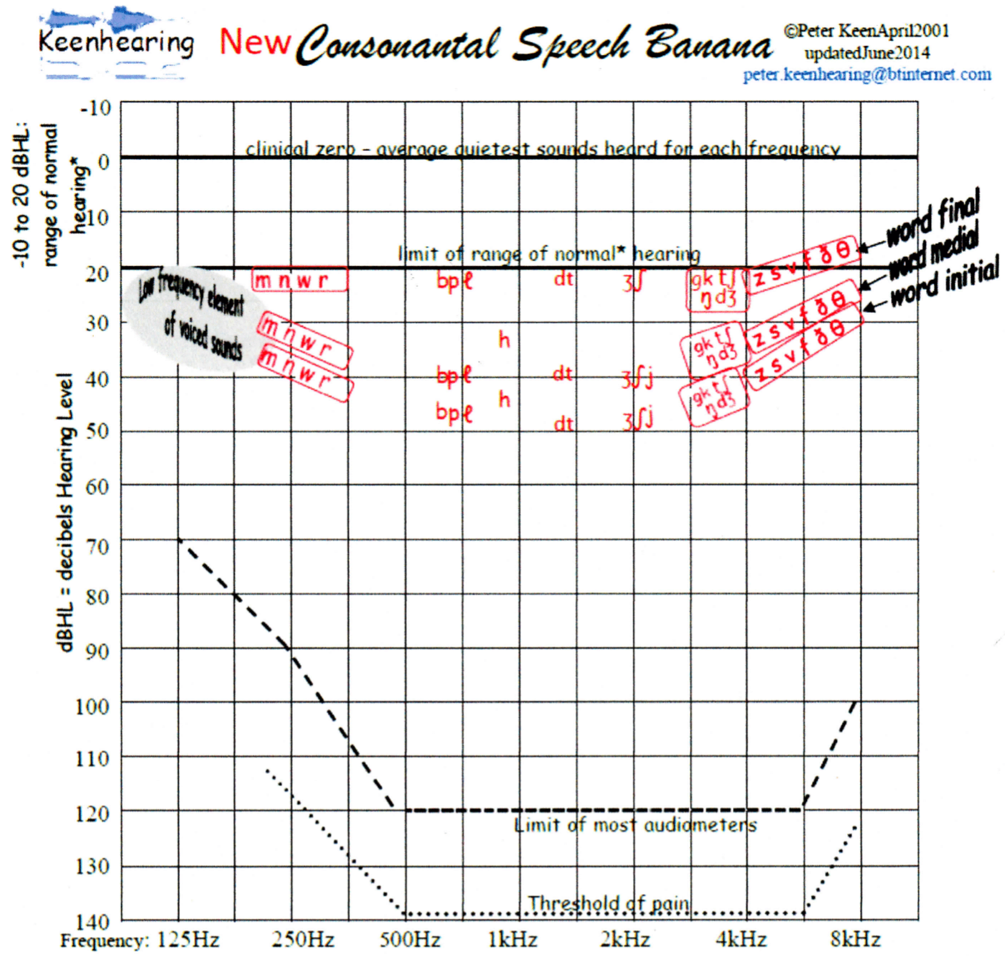
7 APPENDICES

Appendix 1 IPA Chart 2005



The international Phonetic Alphabet (Wikimedia Commons, 2018)

Appendix 2 Consonantal Speech Banana Peter Keen



This speech banana has been updated to include new research for the Count-the-Dot** audiogram (2010). It shows key frequency areas needed by the ear/brain for each consonantal phoneme. Phonemes have other frequency information not key to recognition. **Word Initial** phonemes: loudest, **Word Final:** quietest, **Medials** are quieter than Initial (2nd formant transition both sides helps HF losses perceive Medials better than Initials). The banana shows 'normal' voice level in a quiet room, the ear 1 metre from the speaker. New placement of phonemes allows for the 10 to 12 dB above pure tone threshold needed to perceive the phoneme. Phonemes in boxes cover all area. The levels are dBHL (not dBA or dBSPL). Raising voice or shouting increases low frequencies, not high. Normal voice level moved closer will increase sound level for all phonemes equally, giving better amplified speech perception. P Keen has separate sheet for vowels.

Non-alphabet symbols:

- ŋ as in sing
- j as in you
- ʃ as in ship
- tʃ as in chip
- ʒ as in beige
- dʒ as in just
- θ as in think
- ð as in the

Voiceless phonemes:	p	t	ʃ	k	tʃ	s	f	θ
used with Low frequency element of voiced sounds	↓	↓	↓	↓	↓	↓	↓	↓
become voiced phonemes:	b	d	ʒ	g	dʒ	z	v	ð

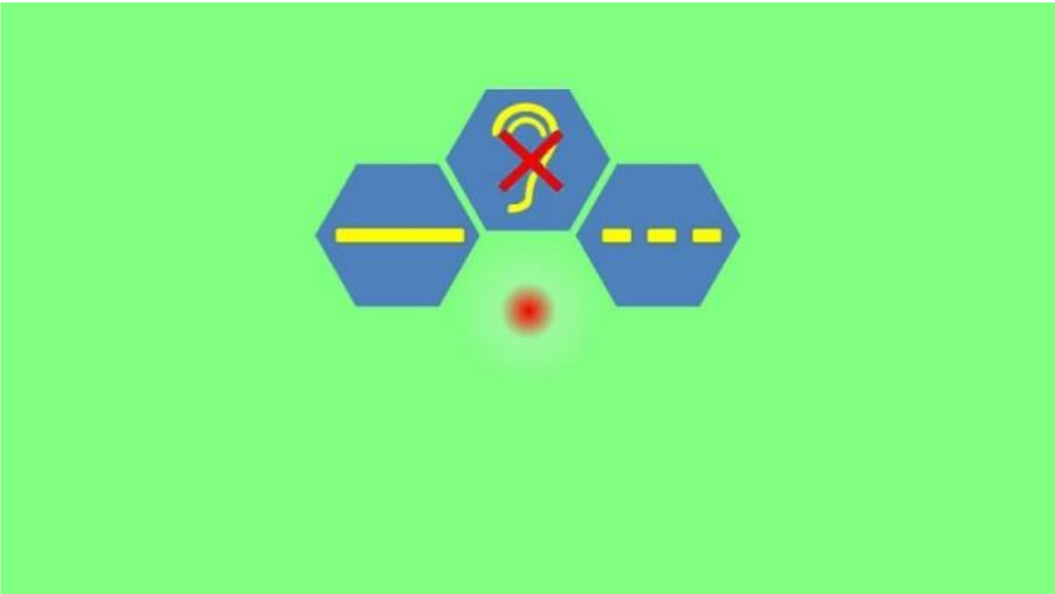
BOTH the voiceless phoneme AND the low frequency element must be heard to identify the voiced consonantal phoneme.

*Normal hearing: hearing allowing perception of 95% - 100% of speech phonemes at 1 m: ≦20dBHL SNR
**Meuller & Killion: A New Count-The-Dots Method, The Hearing Journal, January 2010 Vol 63 No 1

New Consonantal Speech Banana (Keen, 2014)

Appendix 3 DuoTone test

The BELLS DuoTone® screening test as seen in the screenshot below, is based on the patented DuoTone® procedure. In an adaptive procedure two pure-tone stimuli with different frequencies are presented to the user. One stimulus (A) consists of one long tone with a low frequency and *the other* stimulus (B) contains three short tones with the higher frequency. A third stimulus (C) does not contain a signal and represents the “silent” stimulus. One of these three stimuli is randomly selected and presented to the person under test. At the end of the tests, the threshold values are available, one for each tested frequency (50% score values).

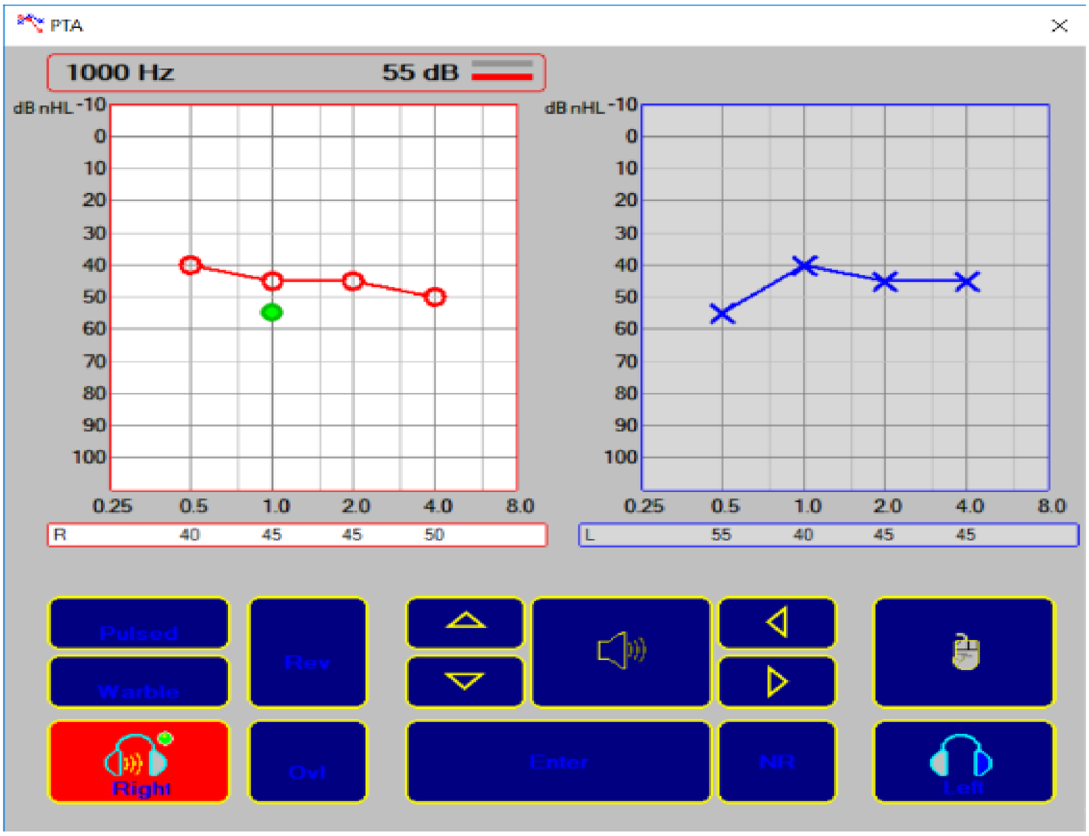


Screenshot Duotone test

Appendix 4 BELLS PTA

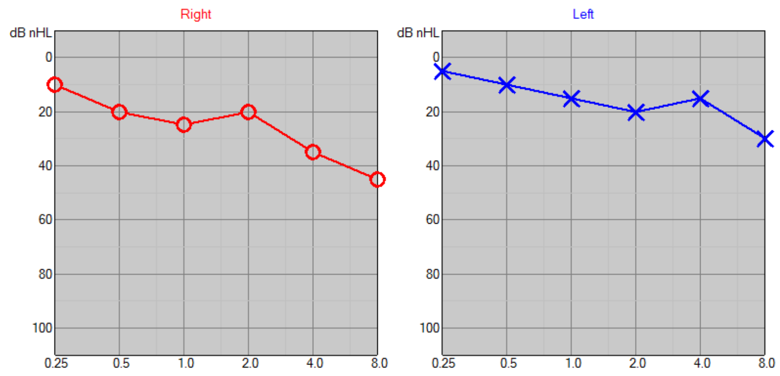
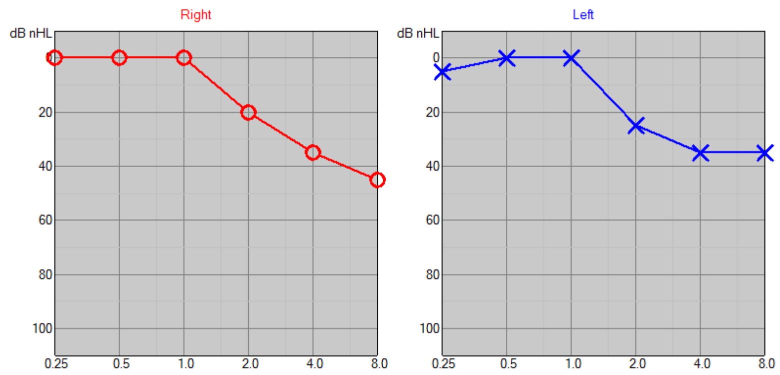
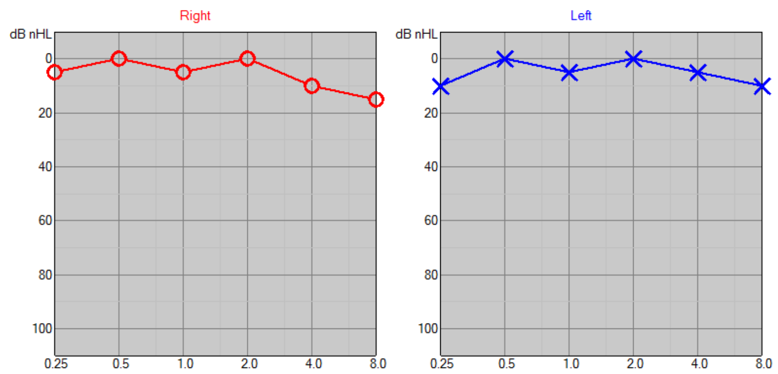
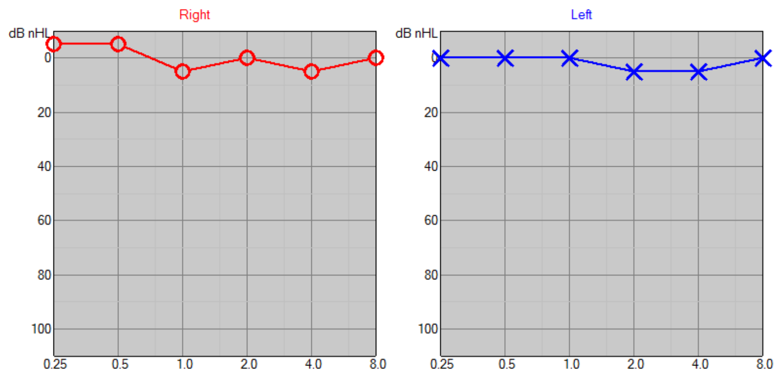
With the BELLS® PTA module as can be seen below, Pure Tone Audiometry can be conducted in the octave frequencies from 250 Hz to 8 KHz. This module is calibrated for the HD 280 PRO headphone in combination with the uDAC-2 external Digital to Analog converter. Other headphones and DA converters may be used. Data is stored in the client file in the BELLS® software.

The BELLS® platform offers possibilities for reporting and printing.



BELLS® PTA screen

Appendix 5 Audiograms of test examiners

Examiner 1 principal investigator	
Examiner 2 T2	
Examiner 3 T3	
Examiner 4 T4	

Appendix 6 PRS scores for test-retest reliability check

Participant	P1-65dB PRS %	P2-65dB PRS %	δ-65 dB	P1-50dB PRS %	P2-50dB PRS %	δ-50 dB	Σ-65dB	Σ-50dB
N21	95	98	3	65	79	14	193	144
N22	97	98	1	81	86	5	195	167
N23	99	100	1	88	85	3	199	173
N24	98	100	2	85	97	12	198	182
N25	96	95	1	85	93	8	191	178
N26	99	99	0	83	82	1	198	165
N27	93	95	2	75	72	3	188	147
N28	99	99	0	89	90	1	198	179
N30	100	100	0	82	96	14	200	178
N41	92	95	3	76	78	2	187	154
N44	98	98	0	73	80	7	196	153
N45	90	92	2	72	69	3	182	141
N47	95	97	2	73	80	7	192	153
N50	96	99	3	76	92	16	195	168
N53	97	92	5	79	83	4	189	162
N54	97	92	5	87	90	3	189	177
N59	97	97	0	71	77	6	194	148
N63	97	96	1	83	71	12	193	154
N67	94	92	2	81	73	8	186	154
N68	97	98	1	90	93	3	195	183
N71	94	96	2	68	69	1	190	137
N72	98	99	1	69	86	17	197	155
Average	**96,3**	**96,7**	**1,68**	**78,68**	**82,77**	**6,82**	**192,95**	**161,45**
SD	**2,43**	**2,69**		**7,14**	**8,58**			
Variance	**5,93**	**7,22**		**50,94**	**73,54**		**21,59**	**193,88**

$$Cronbach\ \alpha = \frac{N}{N-1} \cdot \left(1 - \frac{\sum var(Y_i)}{var(X)}\right) = \frac{N}{N-1} \cdot \left(1 - \frac{\sum_{i=1}^{N} S_{yi}^2}{S_x^2}\right)$$

Cronbach's alpha	Internal consistency
α ≥ 0.9	Excellent
0.9 > α ≥ 0.8	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 > α	Unacceptable

$$\alpha 1(65dB) = \frac{2}{2-1} \cdot \left(1 - \frac{5{,}93 + 7{,}22}{21{,}59}\right) = 0.7825$$

$$\alpha 2(50dB) = \frac{2}{2-1} \cdot \left(1 - \frac{50{,}94 + 73{,}5393}{193{,}8843}\right) = 0.7159$$

Appendix 7 Ethics approval

University of Hertfordshire UH

SOCIAL SCIENCES, ARTS AND HUMANITIES ECDA

ETHICS APPROVAL NOTIFICATION

TO: Ferdinandus A.M. Marinus

CC: Dr Joy Rosenberg

FROM: Dr Timothy H Parke, Social Sciences, Arts and Humanities ECDA Chairman

DATE: 08/10/18

Protocol number: EDU/PGT/CP/03801

Title of study: Adaptation and validation of the NAMES speech test for the Dutch language.

Your application for ethics approval has been accepted and approved by the ECDA for your School and includes work undertaken for this study by the named additional workers below:

This approval is valid:

From: 08/10/18

To: 01/01/19

Additional workers: **no additional workers named**

Please note:

If your research involves invasive procedures you are required to complete and submit an EC7 Protocol Monitoring Form, and your completed consent paperwork to this ECDA once your study is complete. You are also required to complete and submit an EC7 Protocol Monitoring Form if you are a member of staff. This form is available via the Ethics Approval StudyNet Site via the 'Application Forms' page http://www.studynet1.herts.ac.uk/ptl/common/ethics.nsf/Teaching+Documents?Openview&count=9999&restricttocategory=Application+Forms

Any necessary permissions for the use of premises/location and accessing participants for your study must be obtained in writing prior to any data collection commencing. Failure to obtain adequate permissions may be considered a breach of this protocol.

Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1A. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval (if you are a student) and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1A may need to be completed prior to the study being undertaken.

Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be

reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.

Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.

Students must include this Approval Notification with their submission.

Appendix 8 Participant information sheet EC6 (DUTCH)

UNIVERSITY OF HERTFORDSHIRE

ETHICS COMMITTEE FOR STUDIES INVOLVING THE USE OF HUMAN PARTICIPANTS ('ETHICS COMMITTEE')

FORMULIER EC6: DEELNEMER INFORMATIE BLAD

1. Titel van de studie

Aanpassing en validatie van de NAMES® spraakaudiometrie test voor de Nederlandse taal.

2. Introductie

U bent uitgenodigd om deel te nemen aan dit onderzoek. Voordat u beslist of u dit wil doen is het belangrijk dat u begrijpt wat de studie inhoudt en wat daarin uw bijdrage zal zijn. Neemt u alstublieft de tijd om de volgende informatie goed door te nemen en zo nodig met anderen te overleggen. Aarzel niet om het aan ons te vragen als voor u iets niet duidelijk is, of als u meer informatie nodig heeft om te kunnen beslissen. Neem alstublieft de tijd om te beslissen of u aan dit onderzoek wil deelnemen. De (Engelstalige) regeling van de Universiteit over het omgaan met studies waarbij personen zijn betrokken is te vinden via de volgende link:

http://sitem.herts.ac.uk/secreg/upr/RE01.htm

Dank voor het lezen.

3. Wat is het doel van deze studie?

Deze studie is opgezet voor het aanpassen en valideren van de NAMES® foneem discriminatie test voor de Nederlandse taal. Het acronym NAMES staat voor: Name-based Auditory Multilingual Evaluation of Speech". De test wordt gepresenteerd in gesproken vorm, waarbij de gesproken woorden, zoals vaak ook veel namen geen betekenis hebben. De test maakt gebruik van nonsens woorden. Hierdoor wordt de test niet beïnvloed door taalkennis, ondanks het feit dat het wel de fonetische regels van de Nederlandse taal volgt.
Deze test is complementair aan andere spraaktesten, welke vaak wél meer afhankelijk zijn van en beïnvloed worden door taalkennis. Deze test levert ook (frequentie) specifieke informatie op die kan worden gebruikt voor hoortoestel aanpassing en spraak- en taaltherapie.
Voor deze studie nodigen we een groep deelnemers uit met een normaal gehoor, en we vragen hen om exact na te spreken wat ze gehoord hebben, of door op een computerscherm de geluiden te selecteren die ze gehoord hebben. Met deze verzamelde informatie zullen we deze test gaan valideren voor de groep normaal horende luisteraars. In een latere studie zal vergelijkbare data verzameld worden van deelnemers met verschillende soorten gehoorverliezen.
Ook zullen we in dit onderzoek verschillende manieren testen, waarmee de deelnemer kan antwoorden. Én door deelnemers uit verschillende regio's te kiezen kunnen we onderzoeken of lokale dialecten nog van invloed zijn op de test.

4. Moet ik deelnemen?
Het is geheel aan u of u beslist om wél of niet deel te nemen aan dit onderzoek. Als u deelneemt aan deze studie betekent dit dat u ermee instemt dat de data verkregen uit

dit onderzoek door de researcher gebruikt mogen worden voor analyse en validatie van de test.
Als u besluit om deel te nemen krijgt u een kopie van dit informatieblad om te behouden en aan u zal gevraagd worden om een toestemmingsformulier te ondertekenen. Uw instemming om deel te nemen aan dit onderzoek wil niet zeggen dat u verplicht bent om dit onderzoek volledig af te maken. U bent vrij om op elk door u gewenst moment dit onderzoek te stoppen, zonder opgave van reden. Een besluit om op een willekeurig moment te stoppen, of een besluit om helemaal niet deel te nemen zal voor u verder geen gevolgen hebben.

5. Zijn er leeftijds- of andere beperkingen die u weerhouden van deelname?

Om te kunnen deelnemen moet u tenminste 18 jaar oud zijn. Omdat we ook leeftijdseffecten op deze test willen onderzoeken zullen we deelnemers zoeken van verschillende leeftijden boven de 17, en we zoeken naar een gelijkmatige verdeling in leeftijdsgroepen. Voor de validatie zullen alleen gegevens gebruikt worden van deelnemers met een normaal gehoor (<30dB_{HL} bij 500 Hz en bij 6 KHz). De data van deelnemers met gehoorverliezen groter dan 30 dBHL zal gebruikt worden om de effecten van gehoorverlies op de NAMES scores te onderzoeken.

6. Hoeveel tijd kost mijn deelname aan dit onderzoek?

Als u besluit om deel te nemen aan dit onderzoek, zullen we met u een afspraak maken om deze spraakaudiometrie test te doen op één van de Kentalis Audiologische Centra. De test, inclusief korte gehoorscreening duurt niet langer dan 20 minuten. We zullen enkel uw antwoorden verzamelen, en deze gegevens zullen later voor de analyse van de test gebruikt gaan worden. Dit onderzoek eindigt in April 2019.

7. Wat gebeurt er met me als ik deelneem?

Door deel te nemen aan dit onderzoek, verleent u de researcher het recht om de tijdens de NAMES test verzamelde gegevens te gebruiken voor analyse en validatie van de test. U hoeft zelf verder niets te doen.

8. Wat zijn bij deelname mogelijke nadelige effecten, risico's of neveneffecten?

Er zijn geen risico's of neveneffecten verbonden aan deelname aan deze studie.

9. Wat zijn mogelijke voordelen van deelname?

Door in te stemmen met het gebruik van de in dit onderzoek verzamelde gegevens, maakt u het mede mogelijk dat deze test in Nederland beschikbaar komt. Dit zal bijdragen in de diagnostiek van kinderen en volwassenen met gehoorverlies en taalontwikkeling stoornissen. De resultaten van deze test kunnen audiologen helpen met het aanmeten van hoortoestellen en het helpt logopedisten met de identificatie van de fonologische problemen van hun cliënten.

10, Hoe wordt mijn deelname aan dit onderzoek vertrouwelijk gehouden?

Alle gegevens in dit onderzoek worden anoniem opgeslagen. De gegevens worden op een dusdanige manier opgeslagen dat enkel de onderzoeker toegang heeft tot de gegevens. De NAMES® data en audio opnamen worden opgeslagen op een met password en encryptie beveiligde hard disk. Na afloop van de studie zullen de

opnames verwijderd worden en de geanonimiseerde NAMES® data zal nog 7 jaar bewaard worden voor vergelijkende studies in andere taalversies.
De ingevulde en ondertekende toestemmingsformulieren zullen worden gescand en opgeslagen op de persoonlijke externe beveiligde harddisk van de onderzoeker.

11. Audio-visueel materiaal

Gedurende sommige testen zal de NAMES software ook audio registraties maken van uw responses, zodat die later geanalyseerd kunnen worden. De opnames worden op een computer hard disk opgeslagen in de met een password beveiligde BELLS software omgeving en de audio bestanden worden na dit onderzoek gewist.

12. Wat gebeurt er met de in dit onderzoek verzamelde data?

1. De in de NAMES® test verzamelde data zal voor maximaal 7 jaar elektronisch opgeslagen worden op een gecodeerde en met een wachtwoord beveiligde harddisk, waarna de data op een betrouwbare manier gewist zal worden. Alleen de onderzoeker heeft toegang tot deze data.
2. Het bestand met de audio opnames van de deelnemers aan deze test zal door de onderzoeker worden opgeslagen op een geëncrypte en met wachtwoord beveiligde externe harddisk, waartoe enkel de onderzoeker toegang heeft.
3. Voor opslag wordt de data geanonimiseerd.

13. Wordt de data nog gebruikt voor verdere studies?

U geeft toestemming voor mogelijk hergebruik of toekomstige analyse van de NAMES® data in een toekomstig ethisch goedgekeurde studie; de data die wordt hergebruikt zal geanonimiseerd zijn.

14. Wie heeft deze studie beoordeelt?

Deze studie is beoordeelt door:

The University of Hertfortshire Social Sciences, Arts and Humanities Ethics Committee with Delegated Authority.

Het UH protocol nummer is EDU/PGT/CP/03801.

15. Risicofactoren voor anderen

Merk op, dat indien zich gedurende de studie, medische condities of niet-medische condities zoals onwettige activiteiten duidelijk worden, die mogelijk anderen in gevaar hebben gebracht of kunnen brengen, de Universiteit de zaak kan verwijzen naar de daarvoor aangewezen autoriteiten.

16. Met wie kan ik contact opnemen in geval van vragen?

Indien u verdere informatie wenst, of zaken persoonlijk wenst te overleggen, neem dan a.u.b. schriftelijk, per telefoon of per email contact met mij op:

Fred Marinus - F.Marinus@kentalis.nl - 06-51391635

Natuurlijk hopen dat het niet het geval is, maar mocht u klachten of bedenkingen hebben over welk onderdeel dan ook, of over de manier waarop u benaderd of behandeld bent in de loop van dit onderzoek, schrijf dan naar het secretariaat en de registrar van de Universiteit.

Hartelijk dank voor het lezen van deze informatie en het in overweging nemen om mee te doen aan dit onderzoek.

Appendix 9 Ethics Consent form EC3 (DUTCH)

UNIVERSITY OF HERTFORDSHIRE
ETHICS COMMITTEE FOR STUDIES INVOLVING THE USE OF HUMAN PARTICIPANTS
('ETHICS COMMITTEE')

FORM EC3
CONSENT FORM FOR STUDIES INVOLVING HUMAN PARTICIPANTS
(Toestemmingsformulier voor onderzoeken waarin personen betrokken zijn)

Ik, ondergetekende *[geef hier a.u.b. uw naam in BLOKLETTERS]*

…….
contact gegevens: *[geef hier a.u.b. uw contact gegevens, voldoende voor de onderzoeker om met u in contact te komen, zoals post of email adres]*

…….
verklaar hierbij vrijwillig deel te nemen aan het onderzoek:

Adaptatie en validatie van de Nederlandse versie van de NAMES® spraak audiometrie test.

(UH Protocol nummer EDU/PGT/CP/03801)

1 Ik bevestig dat ik een Deelnemer Informatie Blad heb ontvangen (waarvan een kopie is aangehecht aan dit formulier) welke de details van het onderzoek weergeeft, inclusief de doelen, methodes en ontwerp, de namen en contact gegevens van de voornaamste personen en indien nodig de risico's en mogelijke voordelen van het onderzoek, hoe de vergaarde informatie wordt opgeslagen en voor hoe lang en dat ik ben voorgelicht over follow-up studies waarin een beroep op deelnemers kan worden gedaan. Ik ben ook geïnformeerd over hoe mijn persoonlijke informatie op dit formulier wordt opgeslagen en voor hoe lang. Ik ben op de hoogte gebracht over mijn rol en betrokkenheid in dit onderzoek. Er is mij verteld dat in het geval van een grondige wijziging van de doelen van dit onderzoek ontwerp ik op de hoogte zal worden gesteld en dat mij dan opnieuw toestemming gevraagd wordt voor deelname.

2 Men heeft mij verzekerd dat ik mij te allen tijde mag terug trekken uit dit onderzoek zonder enige nadelige gevolgen en ook zonder opgave van redenen.

3 Door mijn toestemming voor deelname aan dit onderzoek, geef ik aan te begrijpen dat er van het onderzoek audio opnames worden gemaakt, ik ben op de hoogte hoe deze opnames worden gebruikt.

4 Er is mij verteld hoe er met mijn gegevens (data verkregen door het onderzoek en persoonlijke gegevens) wordt omgegaan: hoe deze veilig worden opgeslagen, wie toegang heeft tot deze gegevens en hoe deze worden of zouden kunnen worden gebruikt.

5 Ik begrijp dat dit onderzoek gegevens aan het licht zou kunnen brengen die mogelijk een medisch advies noodzakelijk maken. In dat geval zal ik geïnformeerd worden en zal mij aangeraden worden contact op te nemen met mijn huisarts. Als er gedurende dit onderzoek enig bewijs aan het licht komt dat ik een bestaand medisch probleem heb dat een risico vormt voor anderen, begrijp ik dat de Universiteit mij zal doorverwijzen naar de verantwoordelijke autoriteiten en dat ik verder uitgesloten wordt voor deelname aan dit onderzoek.

6 Ik begrijp dat als er zich een onrechtmatige activiteit voordoet of er een indicatie is van niet-medische omstandigheden die een risico vormen voor anderen, de Universiteit die zaak kan doorverwijzen aan de daarvoor bevoegde autoriteiten.

7 Er is mij verteld dat het mogelijk is dat er na enige tijd weer contact met mij gezocht kan worden in verband met deze of een andere studie.

Handtekening van de deelnemer…………………………………………Datum…………………………

Handtekening van de (hoofd) onderzoeker Datum 20 July 2018
Naam van de (hoofd) onderzoeker: F.A.M. MARINUS

Form EC3-1 January 2016

Appendix 10 Participant instructions for the NAMES® test.

Instruction group 1 (18-59) NL

U krijgt zo meteen een hoofdtelefoon opgezet en we laten u een lijst horen met 20 vreemde “namen”. Ik vraag u deze “namen” letterlijk na te spreken zoals u die hoort, ook als u deze niet helemaal verstaan heeft. De testleider zal de klanken van de “naam” intoetsen zoals u die uitspreekt. Het uiteindelijk resultaat geeft ons een beeld welke klanken u wél of niet goed hebt kunnen verstaan.
We beginnen met uw linker/rechter oor. Na twintig “namen” laten we de “namen” horen op uw andere oor. Dit doen we op drie luidheidsniveaus, dus in totaal hoort u de lijst met “namen” in willekeurige volgorde 6 keer, 3 keer op elk oor. Bij de laatste test als de geluiden erg zacht zijn zult u zeker niet alles meer precies kunnen horen. Herhaal alleen wat u gehoord heeft.

Instruction group 2 (>=60) NL

U krijgt zo meteen een hoofdtelefoon opgezet en we laten u twee keer een lijst horen met 20 vreemde “namen”. Ik vraag u deze “namen” letterlijk na te spreken zoals u die hoort, ook als u het niet helemaal verstaan heeft. De testleider zal de klanken van de “naam” intoetsen zoals u die uitspreekt. Het uiteindelijk resultaat geeft ons een beeld welke klanken u wél of niet goed hebt kunnen verstaan. We beginnen met uw linker/rechter oor. Na twintig “namen” laten we de “namen” horen op uw andere oor, ook weer in een willekeurige volgorde. Herhaal alleen wat u gehoord heeft.

Instruction group 1 (18-59) ENG

In a minute you will get some headphones over your ears and we will present to you a list with 20 foreign “names”. I will ask you to repeat these “names” to me literally the way you hear them, even if you cannot hear everything correctly. I will key in the “name” the way you pronounce it. The result shows us which sounds you perceived well, and which sound are missing or wrong.
We will start with your left/right ear. After twenty “names” the “names” will be presented to your other ear. We will do this at three different intensity levels, so

in total you will hear this list 6 times with the words in a random order. 3 lists at each ear. At the last test, when the sounds are very faint you will not be able to repeat everything exactly. Just repeat what you have heard.

Instruction group 2 (>=60) ENG

In a minute you will get some headphones over your ears and we will present to you a list with 20 foreign "names". I will ask you to repeat these "names" to me literally the way you hear them, even if you cannot hear everything correctly. I will key in the "name" the way you pronounce it. The result shows us which sounds you perceived well, and which sound are missing or wrong.

We will start with your left/right ear. After twenty "names" the "names" will be presented to your other ear again in a random order. Only repeat what you have heard.