# The evaluation of internet-based speech-in-noise tests for noise-induced hearing loss screening

Marya Sheikh Rashid

# The evaluation of internet-based speech-in-noise tests for noise-induced hearing loss screening

Marya Sheikh Rashid

This thesis was prepared at the Department of ENT, Clinical and Experimental Audiology, Amsterdam Public Health research institute, at the Academic Medical Center (AMC), University of Amsterdam, Amsterdam, The Netherlands.

ISBN: 978-94-6332-344-4
Cover design and layout by Ferdinand van Nispen, *my*-thesis.nl
Printed by GVO drukkers & vormgevers B.V.

# The evaluation of internet-based speech-in-noise tests for noise-induced hearing loss screening

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 4 juli 2018, te 14:00 uur

door Marya Sheikh Rashid
geboren te Haarlem

## Promotiecommissie

| | | |
|---|---|---|
| Promotor: | Prof. dr. ir. W.A. Dreschler | AMC-UvA |
| Copromotor: | Dr. M.C.J. Leensen | AMC-UvA |
| Overige leden: | Prof. dr. F.G. Dikkers | AMC-UvA |
| | Prof. dr. M.H.W. Frings-Dresen | AMC-UvA |
| | Dr. B. Sorgdrager | AMC-UvA |
| | Prof. dr. ir. A.F.M. Snik | Radboud Universiteit Nijmegen |
| | Prof. dr. A. van Wieringen | KU Leuven |
| | Dr. J.C.M. Smits | VUmc |

Faculteit der Geneeskunde

# Contents

# Chapter 1

## General introduction

# General introduction



Good hearing abilities are essential for auditory communication, and for creating social, educational, and occupational opportunities. Consequently, hearing loss can interfere with academic and work performance, and may reduce overall psychosocial wellbeing (Nachtegaal et al., 2009). Overall, it can have a large impact on an individual's feel of safety and experienced quality of life (Hong et al., 2013, Su and Chan, 2017).

According to the crude global estimates of the World Health Organization (WHO), about 15% of the world's population has some degree of hearing loss, while about 5% (360 million people) have disabling hearing loss (i.e. hearing loss in the better hearing ear greater than 40 dB HL in adults aged 15 years or older, and greater than 30 dB HL in the better hearing ear in children of 0 to 14 years) (WHO, 2012). Approximately 11% of all disabling hearing loss arise from high-income regions, which includes the Netherlands (WHO, 2012).

Exposure to excessive noise is a well-known risk factor for hearing symptoms, such as tinnitus (i.e. ringing, buzzing, or other sounds in the ear or head (Williams and Carter, 2017)), temporary noise-induced threshold shifts (i.e. acute changes in hearing sensitivity that recover over time (Ryan et al., 2016)), and permanent noise-induced threshold shifts that do not recover (Miller, 1974).

## Noise-induced hearing loss

Noise-induced hearing loss (NIHL) is an acquired sensorineural perceptive hearing loss, and a result of repeated and excessive noise exposure. The hearing damage leading to NIHL develops gradually over time and is permanent (Lamoré, 2012). It initially affects the synapses between the auditory sensory inner hair cells and auditory nerve fibers, in the organ of Corti within the cochlea of the inner ear, causing the disconnected fibers to degenerate. This is referred to as cochlear neuropathy or hidden hearing loss, as there is no threshold shift (Plack et al., 2016)2016. This is then followed by the loss of fragile outer hair cells, structures susceptible to noise damage (Le et al., 2017). At a certain level the damage becomes visible in the pure-tone audiogram. It is expressed as a permanent threshold shift, characterized by a typical noise notch around the higher frequencies 3, 4 and 6 kHz (Rabinowitz et al., 2006a), the frequency region at which the ear is the most vulnerable to noise. When the damage becomes more severe, the surrounding frequencies may be affected as well (Hong et al., 2013). Hearing loss may also be a result of sudden and extremely high noise levels of >130 dB SPL, also known as acoustic trauma (Le et al., 2017). One of the initial and main consequences of NIHL is the difficulty experienced in understanding speech against a background of noise, mainly due to a distortion of sounds (Plomp, 1986).

NIHL is an avoidable hearing condition; its development is a gradual process that frequently goes unnoticed at an early stage (Daniel, 2007). Hence, it may take years to discover that one has a hearing impairment, by which time the damage is irreversible. Consequently, prevention is essential.

Certain population subgroups are more frequently exposed to hazardous sounds as compared to the general population, such as individuals who are excessively exposed to noise at work or in their spare time. Moreover, noise exposure is known to be accumulative, and young adults excessively exposed to noise in the workplace, who are additionally exposed to recreational noise, may be more prone to NIHL. Additionally, the hearing function becomes more vulnerable with age, resulting in age-related hearing loss (i.e. presbyacusis), and NIHL may contribute to the eventual high-frequency hearing loss (HFHL) experienced by the individual (Basner et al., 2014).

## Occupational noise-induced hearing loss

Individuals repeatedly exposed to work-related or occupational noise of high intensities are a well-known high-risk population for noise-related hearing problems such as NIHL and tinnitus. These individuals mainly have jobs in the industrial sector requiring long working hours in unpleasant and harmful sound levels resulting from machines and equipment. In the Netherlands, their risk to develop NIHL is 2.6 times higher as compared to those individuals who are not occupationally exposed to noise (Eysink et al., 2007).

NIHL is one of the most prevalent occupational condition in industrialized countries (Nelson et al., 2005). In the Dutch construction industry NIHL is one of the most commonly reported occupational diseases, with an incidence of 8125 per 100 000 workers in 2014, showing an increasing trend (van der Molen et al., 2016b). NIHL may impact employees' work safety, while the main occupational consequences of tinnitus are long during work absenteeism and work disability (van der Molen et al., 2016a). Complaints that arise from hindering occupational noise include poor speech understanding, concentration losses, and startle responses (van der Molen et al., 2016a).

Noise exposure can be expressed in terms of equivalent levels for a normal working day of 8 hours, measured in A-weighted decibels (dBA, where 'A' corresponds with the physiology of the human auditory system). The risk that occupational noise exposure may result in NIHL increases when the noise exposure is prolonged and exceeds certain intensity levels. The higher the levels and the longer the exposure duration, the more severe the hearing loss (Dobie, 2007). The International Organization for Standardization (ISO) reported on the dose-response relationship between occupational noise exposure and hearing loss and provided the widely accepted norm ISO 1999 (ISO, 1990) (revised in 2013). Various occupational safety criteria have been established, to protect noise-exposed workers for the risks of developing NIHL. Examples of generally accepted occupational noise exposure limits are the Recommended Exposure Limit (REL) of the US National Institute for Occupational Health and Safety (NIOSH, 1998), and the European Union Directive 2003/10/EC (European Parliament, 2003). Both specify permissible equivalent sound levels of 85 dBA with an exchange rate of 3 dBA; for each 3 dBA above the maximum permissible sound levels, the time spent in noise should be halved in order to be safe.

## Recreational noise-induced hearing loss

Moreover, there is growing concern about the effects of excessive recreational noise exposure, especially on young people's hearing (Henderson et al., 2011, Punch et al., 2011, Basner et al., 2014, Carter et al., 2014). Children, teenagers and young adults are often involved in several leisure noise activities, such as spending much time listening to loud music at concerts, clubs and festivals, as well as through earphones and personal listening devices. There has been a rise in popularity of personal music players and smartphones, and technical improvements have led to frequent listening for extended periods of time (Punch et al., 2011, Henry and Foots, 2012, Muchnik et al., 2012). Moreover, these devices are capable of producing sound levels without distortion up to 120 dBA, exposing young people to high intensities of noise (Fligor and Cox, 2004, Serra et al., 2005, Keith et al., 2008, Portnuff et al., 2011).

Studies exploring young people's attitudes towards leisure noise and perceived risk conclude that there is a lack of knowledge of the hearing-related damage that may occur, and consequently, a lack of concern for personal hearing health in this specific population (Vogel et al., 2008, Gilliver et al., 2015, Hunter, 2018). Moreover, those youngsters who are aware of the risks, do not always take steps to protect their hearing (Beach et al., 2013). Furthermore, youngsters tend to prioritize other public health issues such as smoking (Quintanilla-Dieck Mde et al., 2009).

Several attempts of risk assessment have been performed in order to identify whether teenagers and young adults have indeed an enhanced risk of developing NIHL. As there is a lack of evidence-based damage risk criteria for evaluating recreational or music exposure, occupational damage risk standards are usually applied in order to so (Chung et al., 2005, Levey et al., 2011, Knobel and Lima, 2012, Portnuff et al., 2013, Twardella et al., 2017). According to several studies, specifically those teenagers who attend music venues (i.e. clubs, disco's, music concerts and festivals), and listen to amplified music through personal music players engage largely in risky behavior (Beach et al., 2013, Portnuff et al., 2013, Keppler et al., 2015, Twardella et al., 2017).

The association between recreational noise and hearing loss (Feder et al., 2013), and prevalence of hearing-related problems associated with recreational noise, such as temporary and permanent threshold shifts and tinnitus (Niskar et al., 1998, Niskar et al., 2001b, Shargorodsky et al., 2010) have been studied thoroughly. However, literature regarding the increasing prevalence of NIHL due to recreational noise exposure is inconclusive, largely due to methodological heterogeneity and limitations (Basner et al., 2014). According to a comprehensive review, there is sufficient evidence that some recreational activities may provide hazardous noise levels, however, there is still a lack of consensus on the extent of the risk of developing noise-induced hearing loss (Carter et al., 2014). The majority of the studies that have been performed are mainly cross-sectional and based on self-reported noise exposure, while longitudinal studies, providing stronger evidence for establishing causality, are scarce.

For instance, studies using cross-sectional audiometric data available from the National Health And Nutrition Examination Surveys (NHANES) have reported on the changes in prevalence of (high-frequency) hearing loss in US children and adolescents aged 12 to 19 years from 1988 to 2010 (Niskar et al., 1998, Niskar et al., 2001b, Shargorodsky et al., 2010, Su and Chan, 2017). In the latest NHANES (2009-2010), estimates of HFHL and noise-induced threshold shifts (NITS) significantly decreased from 15-20% to approximately 12%. While reported noise and music exposure increased from 20% to 42% in the latest study, there were no associations with hearing loss (Su and Chan, 2017), which is in contrast with previous conclusions on the relationship between noise exposure and hearing loss in children (Schlauch and Carney, 2011).

There is no consensus yet, and more longitudinal studies with long enough surveillance periods and comprehensive noise exposure assessments are required. Nevertheless, there are signs that about 10% of teenagers are at an increased risk of NIHL due to excessive music exposure (Basner et al., 2014, Carter et al., 2014). As it is well known from the occupational field that frequent and long during exposure to high noise levels increase the risk of NIHL, this may in potential apply to the recreational field as well. Even though the teenage population seem to have relatively small elevations in risk, as they are exposed during a shorter period in their teenage years, it is a non-negligible proportion of this wide-spread population that on average run a 10-year risk of NIHL (Vogel, 2009). Moreover, large numbers of people at small risk, may contribute to many cases of NIHL to the whole population (Rose, 2001). Also, high population means of exposure to risk factors may give rise to more individuals with extreme exposure, and consequently more cases (Rose, 2008). Especially those individuals with ears that are more susceptible to noise are more likely to develop hearing loss (Śliwińska-Kowalska et al., 2006). According to WHO worldwide estimates, approximately 1.1 billion adolescents and young adults aged 12 to 35 years, have an increased risk of developing noise-induced hearing loss due to leisure noise exposure (WHO, March 2015). Therefore, it becomes essential to focus on prevention in this population with mild to moderate risk as well.

## Current methods for preventing NIHL in the occupational and teenage population in the Netherlands

Prevention can be categorized in three stages: primary, secondary, and tertiary prevention (Mackenbach and van der Maas, 2005). Primary prevention is to prevent disease before it occurs. Secondary prevention is to detect asymptomatic early or mild cases in order to prevent worsening. Tertiary prevention is to manage already established disease in order to reduce the negative long-term effects in order to improve quality of life, and therefore tertiary prevention strongly relates to medical treatment.

***Preventive measures in the occupational setting***

In the Netherlands, rules and legislation exist in order to provide a healthy and safe working environment for the occupationally noise-exposed working-age population (in Dutch: "Arbeidsomstandighedenwet" (Arbowet), "Arbobesluit" and "Arboregeling"). According to these laws, employers are enforced to make efforts to avoid or reduce the negative effects of noise exposure, based on several action levels depending on the equivalent noise level of an 8-hour working day. These efforts are based on the European Directive 2003/10/EC, and recommend three exposure limits and corresponding actions:

- A lower exposure action level at 80 dBA: Provision of hearing protection devices, information and training of noise-exposed workers.

- An upper exposure action level at 85 dBA: Required hearing protection, performance of noise risk assessments, audiometric assessments once every four years.

- An exposure limit set at 87 dBA measured at the tympanic membrane, and taking hearing protection into account : Direct action required in order to reduce noise levels.

Primary prevention interventions have not been proven to be sufficiently effective (Verbeek et al., 2014, Rabinowitz et al., 2018), and despite the efforts, occupational NIHL rates are still high. Moreover, not all employees are compliant with wearing hearing protection. Therefore, secondary prevention becomes necessary. The current standard for hearing assessment is pure-tone audiometry. This form of testing may not be the most ideal tool for screening (see section on *screening methods*). Hearing screening should be offered regularly, and hearing assessment once every four years may not be sufficient.

Early hearing losses may not be detected in time, creating more room for slowly occurring deteriorations. Furthermore, participation is voluntary, and in practice response rates are not high (Jellema, 2014).

***Preventive measures in children, teenagers and young adults***

In the Netherlands, standardized formal hearing screening in children is performed at birth (i.e. neonatal screening) and until the age of 4 to 6 years (i.e. preschool age) by means of otoacoustic emissions (OAEs) and pure-tone audiometry (NCJ, 2016). Moreover, attention for the primary prevention of acquired NIHL in children, teenagers and young adults is growing. Since 2013, WHO initiated guidelines, set up by the European Committee for Electrotechnical Standardization (CENELEC) and International Electrotechnical Commission (IEC), apply, limiting the maximum sound output levels of commercially available personal music players. Moreover, several non-profit organizations (run by audiologists, ENT physicians and other hearing health promotors, such as the Dutch National Hearing Foundation (in Dutch: "Nationale Hoorstichting")) have been concentrating on the promotion and enforcement of hearing conservation targeted at the teenage population as a whole, including initiatives such as:

- Increasing knowledge and awareness of the risks of developing hearing loss and its consequences, and including educational programs in school curriculums.

- Changing listening behavior: safe use of personal music players, avoiding loud speakers, taking a break during music events allowing the ears to recover from temporary threshold shifts, using hearing protection.

- Creating a safer environment: limiting noise levels and providing hearing protection at music venues.

- Developing smartphone applications in order to assess personal daily noise levels.

Such initiatives have been supported by the Dutch State Secretary for Health, Welfare and Sport and captured in a so-called "Action plan prevention of hearing loss" (in Dutch: "Actieplan preventie gehoorschade") in 2015, in which several important parties are involved (van Rijn, 2017). The action plan is based on three pillars: education and awareness, safe environment, and screening. So far, important steps have been taken to realize this action plan. As far as screening concerns, research has been initiated in order to get an insight into the prevalence of NIHL in the Netherlands. In addition, the guideline for youth hearing screening (age 0 to 18 years) has been revised, and now includes hearing loss due to noise exposure. Voluntary hearing screening at later moments in life performed by general practitioners and hearing aid dispensers is being encouraged. Also, the options of online self-testing are being explored.

## Screening methods

Prevention can be performed by two approaches: the high-risk approach and the mass-population approach (Rose, 2008). Employees working daily in high noise levels can be considered as a well-defined high-risk group, as they have an increased risk of developing NIHL in comparison to the general population. On the other hand, population-wide screening of teenagers who might develop NIHL due to excessive music exposure, might also increase the number of hearing losses identified. For both approaches, it may be necessary to screen relatively large populations on a regular base. Usually, it is logistically not feasible and costly to test a broad public at audiological centers. Consequently, testing may preferably be done in an occupational or school environment. Therefore, it is essential to have an easily accessible valid screening test that can reliably be performed in a remote setting.

Traditionally, pure-tone audiometry has been the clinical standard for diagnosing and screening NIHL worldwide, as in the Netherlands (Fredriksson et al., 2016). Simple tones at several frequencies need to be detected at several intensities in order to establish a hearing threshold. By means of this test the type and degree of the hearing loss can accurately be assessed, and presented in a pure-tone audiogram. However, in order to attain valid and reliable measurements, a soundproof room, regularly calibrated equipment, and a trained test administrator are required. Invalid testing and unreliable screening is undesirable and inefficient, as it may lead to wrong classifications, and as a consequence, to the missing of cases and unnecessary referrals of normal-hearing individuals. The much demanding criteria for proper audiometric assessment makes the test relatively expensive and less feasible for frequent screening in remote settings. Furthermore, especially when identifying early or mild hearing losses, pure-tone testing may be subject to variability due to calibration issues, test protocol, test-retest reliability, test environment, and tester and participant factors (Schlauch and Carney, 2012). Besides, this detection test it is also not a very good indicator of a person's daily communicative ability, especially in noisy circumstances (Jansen et al., 2014a). Since pure-tone audiometry is mainly a tool to measure the presence and extent of NIHL, rather than the functional impact (Le et al., 2017).

Other methods that have been suggested for screening purposes are questionnaires, OAEs, and speech perception in noise tests. Screening by means of questionnaires can be a cost-efficient alternative, however these self-rated hearing assessments are known to be less valid, and in particular, less sensitive for mild hearing losses (Hong et al., 2011, Fredriksson et al., 2016). Furthermore, results of questionnaires are usually influenced by age, which leads to underreporting of the hearing status (Smits, 2005, Mosites et al., 2016). OAEs, measuring the outer hair cell function, have the advantage of being objective, sensitive, and easy to administer for the detection of early hearing loss (Lapsley Miller et al., 2006, Le et al., 2017). This makes them highly suitable and commonly used for neonatal screening. However, OAEs have some limitations concerning equipment and test procedure, such as calibration, probe fitting, and the influence of environmental noise. Moreover, when the target population is an older occupationally exposed population with pre-existing moderate to severe NIHL (or other forms of hearing loss, such as presbyacusis), the recording of OAEs becomes less reliable due to low or absent OAEs (Helleman et al., 2010). Speech perception in noise tests (i.e. speech-in-noise tests) are supra-threshold (i.e. speech is presented above the auditory threshold), and measure the ability to understand speech in a background noise, i.e. the speech-in-noise performance. Speech-in-noise tests are, in contrast to the more traditional pure-tone audiometry detection test, suitable to assess the functional impact of hearing loss (Le et al., 2017). In addition, speech-in-noise tests may create more awareness. Part of the speech signals presented during testing are audible, but cannot be understood. This may be more confronting for the listener than not hearing a soft tone while one is not aware that the tone is being presented. Speech-in-noise tests provide promising possibilities for hearing screening. The following paragraphs elaborate on this type of hearing screening.

## Alternative testing: online assessment of speech perception in noise

The main outcome measure of a speech-in-noise test is the speech-reception threshold (SRT), which is the average signal-to-noise ratio (SNR) of the test responses needed for 50% speech intelligibility. The SRT is highly correlated with the pure-tone average of 2 and 4 kHz (Smoorenburg, 1990, Smoorenburg, 1992). The first Dutch sentence test in noise was developed by Plomp and Mimpen (1979a). Tests using the same principles, involving an adaptive up-down procedure for the presentation of speech in noise, have been used worldwide, and serve as a base for the development of other speech-in-noise tests.

Speech-in-noise tests have several main test features, such as the main purpose of testing (e.g. diagnostics, screening or monitoring) and the target group (e.g. general public or specific subpopulation, normal hearing or hearing impaired), the type of speech material (e.g. sentences, monosyllables, digits, non-words, closed or open set, female or male speaker), the type of background noise (e.g. stationary, fluctuating, filtered, multi-talker, babble), and the applied measurement procedure (e.g. starting level, test length, test environment, presentation device, monaural or binaural, presentation through headphones or speakers). For this reason, different types of speech-in-noise tests have been developed.

Over the past years, speech-in-noise tests have been taking a new direction, namely by presentation through the internet. The use of computer and mobile technologies and wireless networks in providing health information, education, and healthcare services is becoming mainstream (i.e. e-Health and m-Health), and plays an important role in audiology as well (Eikelboom and Swanepoel, 2016, Paglialonga et al., 2017). Practical so-called tele-audiology applications facilitate the remote delivery of efficient and sustainable public and hearing health care services (Laplante-Levesque et al., 2016). Moreover, a speech-in-noise test is a supra-threshold test, and measures the SNR, making it a relatively robust test and insensitive to environmental noise (Smits et al., 2004, Culling et al., 2005). This is an important starting point for self-administered testing through the internet, with the main advantage of a sound-isolated booth becoming unnecessary for reliable testing. A quiet room with ambient levels up to 40 dBA appears to be sufficient (Jansen, 2013). Internet-based speech-in-noise testing has some other advantages. This type of testing does not involve expensive and frequently calibrated equipment and the test can be fully self-administered, making experienced test leaders unrequired. Furthermore, test results are – within limits - not influenced by the absolute presentation levels in stationary noise (Plomp and Mimpen, 1979b). In addition, speech-in-noise tests are insensitive to conductive hearing loss, hearing loss caused by damage of the external or middle ear (Plomp, 1986). Overall, online speech in-noise tests are easily accessible, quick and low-cost, offering possibilities for screening purposes.

## Internet based speech-in-noise tests: current state of the art

In recent years, several online speech-in-noise tests for screening hearing disability have been developed and evaluated.

### *Digits-in-noise test*

The first automated speech-in-noise screening test was the Dutch "National Hearing Test", presenting digit triplets in a stationary noise, developed by Smits et al (2004). The test was at first presented through telephone, but shortly followed by the implementation online, in collaboration with the Dutch National Hearing Foundation (Smits et al., 2006). Later on, this test was improved and adapted to the digits-in-noise (DIN) test, also known as the digit triplet test (DTT) (Smits et al., 2013), for which the original telephone limited bandwidth of the stimuli were adapted to broadband stimuli, making the test more suitable for the detection of HFHL. Due to its presentation of digits in noise, the DIN test is cognitively low demanding and has a low contextual redundancy, making it a reliable test (Jansen, 2013, Smits et al., 2013). Overall, high sensitivity and specificity values and low measurement errors have been reported for the DIN test.

Within the framework of Hearcom (Hearing in the Communication Society), a European multicenter project, the DIN test has been developed and evaluated in several other languages, such as French, American-English, German, Polish, Russian, Turkish and Spanish (Jansen et al., 2010, Watson et al., 2012, Zokoll et al., 2013, Akeroyd et al., 2015, Kollmeier et al., 2016, Smits et al., 2016). The DIN test has been used for several purposes, such as screening in the general and elderly population (Smits et al., 2006, Koole et al., 2016), occupational screening (Jansen et al., 2013), and screening in children (Koopmans and Smits, 2015). Moreover, the test has been introduced as a clinical and screening instrument in underserved communities and underdeveloped areas in South-Africa (Potgieter et al., 2015). The DIN test has also been applied as a monitoring tool in large longitudinal cohort studies, measuring intra-individual changes in hearing over time (Stam et al., 2015). Moreover, the test has been adapted to words-in-noise tests (Jansen et al., 2014b, Vlaming et al., 2014), and has been presented in other types of background noises (Vercammen et al., 2017). Currently, the test is also being administered through smartphones (Potgieter et al., 2015, Potgieter et al., 2017).

***Other online speech-in-noise screening tests***

Several other online speech-in-noise screening tests have been developed, such as the words-in-noise test in Swedish (Molander et al., 2013), and the Italian adult hearing screening test 'Speech Understanding in Noise' (SUN) test. The SUN test is available in multiple languages, including Brazilian Portuguese, Spanish and Mandarin Chinese (Vaez et al., 2014, Paglialonga et al., 2014). Furthermore, an Australian telephone-based speech-in-noise test 'Telscreen' has been developed, which was modelled on the DTT (Dillon et al., 2016).

For the Dutch language, three internet-based self-screening tests have been developed in collaboration with the Dutch National Hearing Foundation and Leiden University Medical Center, during the period of 2004-2007. The main purpose was to offer tools for awareness and self-administered screening through the internet, for children, teenagers and adult workers exposed to recreational and occupational noise, possibly at risk of developing NIHL.

***Earcheck***

First of all, "Earcheck" (EC, in Dutch: "Oorcheck"), was specifically targeted at teenagers and young adults at risk of recreational NIHL, and designed as a screening and awareness tool (Albrecht et al., 2005). The original test included monosyllables, CVCs (consonant-vowel-consonant) with high-frequency consonants instead of digit triplets in a broadband noise, accompanied by picture responses.

An evaluation study revealed that, although reliable, the original EC was not sensitive enough for NIHL, as listeners with a beginning noise notch benefited from their preserved hearing ability for the low and mid-frequencies (Leensen et al., 2011a). Therefore, the speech and noise material of the test were improved, resulting in speech stimuli that were adjusted in level to achieve equal perceptual difficulty, which is important for an accurate test. Also, a low-pass filtered noise was introduced, stimulating the use of high-frequency speech information, which is advantageous for normal-hearing listeners, improving the discriminative power of the test (Leensen et al., 2011b). Then, test sensitivity for uncontrolled parameters in domestic usage were investigated, including the presentation level, and transducer type (Leensen and Dreschler, 2013b). EC presented in a stationary noise is relatively insensitive for influences of these parameters, however, when presented in a low-pas filtered noise, there are some limitations for domestic testing, probably due to highly set presentation levels by hearing-impaired listeners and variations in test equipment. Finally, the applicability of EC in occupational hearing screening was studied in 250 noise-exposed construction workers, revealing a moderate sensitivity (68%) and specificity (71%) of the test when performed in a domestic setting (Leensen and Dreschler, 2013a).

***Occupational Earcheck***

Another test, "Occupational Earcheck" (OEC, in Dutch: "Bedrijfsoorcheck") was developed for screening and monitoring of occupational NIHL in occupationally noise exposed employees (Ellis et al., 2006). Similar to EC, the original test included CVCs with high-frequency consonants in a broadband noise, accompanied by picture responses.

***Hearing screening test for children***

Finally, in 2007, a hearing screening test for school-age children (in Dutch: "Kinderhoortest") was developed, with the aim of evaluating the hearing of children in an easy and accessible way at a remote setting, such as the school environment.

## Thesis objectives

Building on upon earlier EC research, the goal was to further evaluate the three Dutch online tests, OEC, EC, and the hearing screening test for children, for screening applications. For EC and the hearing screening test for children the purpose and the target group were clear, and the initial speech and noise material and testing procedure were established. However, both tests were not validated prior to online implementation, and although the tests were already available on the internet, primarily as awareness and research tools, further research was needed before adequately applying them for screening purposes. As far as OEC concerns, the hypothesis was that a more reliable test could be achieved compared with EC, mainly due to the use of a monaural presentation, specific high-frequency speech material, and individual starting level and longer run-up before the actual SRT calculation.

The main objective of this thesis is to evaluate existing internet-based speech-in-noise hearing screening tests, in order to improve the quality of those tests. The focus is on the optimization, validation, application and adjustment of the tests in order to obtain valid and reliable screening tests that can be used for secondary prevention of NIHL. The ultimate goal is to develop effective and successful hearing screening tests that will benefit overall quality of secondary prevention of hearing health on the population level, by increasing the number of identified and treated early stage hearing losses due to noise exposure.

Part I of the thesis evaluates the OEC. The main approach was to incorporate the main recommendations of previous research, and from thereon evaluate and modify the test in order to attain a valid and reliable test that can be used for occupational hearing screening. Part II of the thesis focusses on the suitability of EC and the hearing screening test for children for screening in school-age children and teenagers. The main approach was to explore online test results, focusing on test-retest reliability, and the effects of age and presentation type.

## Methodology: test research

This thesis mainly involved test research. The main rationale for evaluating a medical test in clinical research is to reduce its uncertainties, and to assess its benefits, in order to help practitioners and policymakers in sound informed decision making (Leeflang, 2008). It helps to determine whether or not a certain test should be used in clinical or screening practice. When evaluating the quality of an existing or new test, it is essential to assess its characteristics. The main test characteristics are the validity and the reliability; whether the test measures what it is intended to measure (e.g. the presence or absence of hearing loss), and whether the test outcomes are consistent and reproducible.

The main objective of test research is to assess whether a single test (i.e. the index test) adequately discriminates between the presence or absence of a particular hearing loss (i.e. the target condition) (Moons et al., 2004). This is called the (diagnostic) test accuracy. In order to evaluate a newly developed medical screening test, it is important to validate the results of the index test against the results of a gold standard, or reference standard. The reference standard is the best available method to establish the presence or absence of the target condition (Leeflang, 2008). The use of the term 'reference' is more accurate than the term 'gold', since even the best available methods are rarely perfectly accurate. The comparison of the results of both tests are commonly summarized in a two-by-two contingency table, from which important test measures can be calculated (Table 1.1). Based on the definition maintained for the target condition according to the reference standard, and the pass or refer outcomes of the index test (i.e. two or more categories), individuals may be classified into four groups, true positives, true negatives, false positives, or false negatives. The terms positive and negative refer to the presence or absence of the target condition (Altman and Bland, 1994c). The most important test characteristics that can be derived from the table are the sensitivity and the specificity and positive and negative predictive values (NPV and PPV). These are expressed as followed, according to definitions provided by Altman and Bland (1994c, 1994a):

- Sensitivity: the proportion of true positives correctly identified by the test.

- Specificity: the proportion of true negatives correctly identified by the test.

- PPV: the proportion of patients with positive test results correctly diagnosed.

- NPV: the proportion of patients with negative results correctly diagnosed.

**Table 1.1.** Example of a two-by-two contingency table.

| Index test | | Reference standard | | |
|---|---|---|---|---|
| | | Target condition present | Target condition absent | Total |
| | Positive test + | *True positive (a)* | *False positive (b)* | *a+b* |
| | Negative test - | *False negative (c)* | *True negative (d)* | *c+d* |
| | Total | *a+c* | *b+d* | *a+b+c+d* |

*Sensitivity=a/(a+c). Specificity=d/(b+d). PPV=a/(a+b). NPV=d/(c+d).*

For clinical application, a high sensitivity suggests that the test is able to accurately detect the target condition. Which means that if the test result is negative, one can be certain that the disease is absent, i.e. a highly sensitive test is able to rule out the disease (also known as the 'SNout' rule). On the other hand, a high specificity suggests that the test is able to rule in disease when the test result is positive (also known as the 'SPin' rule); when the test result is positive, one can be certain that the target condition is present (Parikh et al., 2009). Generally, acceptable values for sensitivity and specificity for screening are set at 80%. However, this is an arbitrarily set rule of thumb, and proper sensitivity and specificity levels mainly depend on the prevalence and severity of the target condition and the consequences of screening (or no screening), and accompanying costs for that specific situation. Moreover, sensitivity and specificity are not fixed test characteristics, and depend on population characteristics and test circumstances (Moons et al., 1997). Likewise, predictive values depend on the prevalence of the target condition, and may differ according to the population being studied (Altman and Bland, 1994a).

In this thesis, the evaluation of OEC is done according to a phased approach; in a controlled setting, and in more realistic occupational settings in unselected noise-exposed populations. The online-speech in noise test is the index test with the SRT as a continuous outcome measure, that is compared to the reference standard pure-tone audiometry, with a dichotomous (i.e. NIHL present or absent) test result based on the pure-tone average (PTA) of the higher frequencies. A ROC (Receiver Operating Curve) plot is used to select a cut-off value for the SRT, in order to distinguish HFHL from no HFHL (Altman and Bland, 1994b). SRT results are also correlated to the PTA of the higher frequencies.

Test-retest reliability is assessed by means of test-retest reliability coefficients, such as the intra-class correlation efficient, and by calculating measurement errors. Additionally, in order to assess the reliability of the speech-in-noise test, the steepness of the psychometric curves, or performance intensity functions, is assessed, with a steep curve corresponding to a precise test. EC has been investigated in a comparable way. In this thesis, though, the emphasis lies on the applicability of the test by relating the SRT to other variables, such as age and measurement procedure, and to assess the effects of training.

**Thesis outline**

The first part of this thesis focusses on the evaluation of an internet-based speech-in-noise test for screening of occupational noise-induced hearing loss in adults exposed to occupational noise, the Occupational Earcheck (OEC). Chapter 2 optimizes and evaluates OEC in a laboratory-based case-control test accuracy study, i.e. by means of a two-gate design, in a group of normal-hearing and hearing-impaired subjects. Alternative low-pass filtered masking noise conditions are investigated, in order to decide whether the test is conceptually right, and in what type of back ground noise the speech material should be presented. Chapter 3 evaluates the accuracy of OEC in a representative population of noise-exposed workers. A cut-off value for a pass or fail outcome is calculated with a proper trade-off between sensitivity and specificity, and a transition is made from ear level to individual level. In chapter 4 OEC is applied in another representative noise-exposed population in order to assess the sensitivity and specificity for the test, incorporating automatic conditional rescreening; the benefit of a rescreen is investigated. Furthermore, the sensitivity and specificity of the test is investigated for different age groups, and for different degrees of HFHL.

The second part of this thesis focusses on the application of internet-based screening tests in young people. Chapter 5 describes the results of five years of online speech-in-noise testing in teenagers and young adults by means of Earcheck (EC). Chapter 6 describes the evaluation of EC in teenagers, investigating the effects of age, gender, education level and test repetition on SRT performance in a high school setting. Additionally in Chapter 7, the suitability of online and smartphone speech-in-noise testing at schools in school-age children is investigated by means of the hearing screening test for children. The main objectives are to assess age effects and to evaluate test-retest reliability.

In the final chapter the main findings of the research presented in this thesis are summarized, and discussed. It reflects on the concrete value of the evaluated tests, methodological aspects, and on the implications for future research and practice.

# Part I

# The evaluation of an internet-based speech-in-noise test for occupational screening purposes.

# Chapter 2

# Laboratory evaluation of an optimised internet-based speech-in-noise test for occupational high-frequency hearing loss screening: Occupational Earcheck.

M. Sheikh Rashid[1]
M.C.J. Leensen[1]
J.A.P.M. de Laat[2]
W.A. Dreschler[1]

[1] Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands
[2] Department of Audiology, Leiden University Medical Center, Leiden, The Netherlands

International Journal of Audiology, 2017, 56 (11), 844-853.

## Abstract

**Objective.** The 'Occupational Earcheck' (OEC) is a Dutch online self-screening speech-in-noise test developed for the detection of occupational high-frequency hearing loss (HFHL). This study evaluates an optimised version of the test and determines the most appropriate masking noise.

**Design.** The original OEC was improved by homogenisation of the speech material, and shortening the test. A laboratory-based cross-sectional study was performed in which the optimised OEC in five alternative masking noise conditions was evaluated.

**Study sample.** The study was conducted on 18 normal-hearing (NH) adults, and 15 middle-aged listeners with HFHL.

**Results.** The OEC in a low-pass (LP) filtered stationary background noise (test version LP 3: with a cut-off frequency of 1.6 kHz, and a noise floor of -12 dB) was the most accurate version tested. The test showed a reasonable sensitivity (93%), and specificity (94%), and test reliability (intra-class correlation coefficient: 0.84, mean within-subject standard deviation: 1.5 dB SNR, slope of psychometric function: 13.1%/dB SNR).

**Conclusions.** The improved OEC, with homogenous word material in a LP filtered noise, appears to be suitable for the discrimination between younger NH listeners and older listeners with HFHL. The appropriateness of the OEC for screening purposes in an occupational setting will be studied further.

## Introduction

High-frequency hearing loss (HFHL) caused by occupational noise, also known as occupational noise-induced hearing loss (NIHL) is an important worldwide public health problem (May, 2000). In the Netherlands, NIHL is one of the most commonly reported occupational illnesses (van der Molen et al., 2014). NIHL is an acquired sensorineural hearing loss with noise as an avoidable cause, and is therefore preventable. The damage that develops over the years due to noise exposure is permanent. NIHL develops gradually and is often unnoticed until the damage becomes substantial. It initially affects the higher frequency region of 3 to 6 kHz, the region most susceptible to noise. This shows as a characteristic notch in the audiogram at 4 kHz (Brookhouser, 1994, May, 2000, Flamme et al., 2014). The notch broadens as noise exposure continues (Hsu et al., 2013). One of the first consequences of hearing loss due to noise is difficulty in understanding speech in daily situations when background noise is present (Kramer et al., 1998). This specific hearing disability can be accurately measured by means of a speech-in-noise test (Smoorenburg, 1992). Such a test measures the ability to understand speech in noise by varying the ratio between speech and noise levels, the signal-to-noise ratio (SNR). The outcome measure is the speech reception threshold (SRT), the average SNR at which a particular percentage (e.g. 50%) of the speech material is correctly identified.

Over the past few years several telephone- and internet-based speech-in-noise self-tests have been developed in various languages, with different purposes, and aimed at various populations (Smits et al., 2004, Smits et al., 2006, Jansen et al., 2010, Leensen et al., 2011a, Watson et al., 2012, Molander et al., 2013, Paglialonga et al., 2014, Vlaming et al., 2014, Williams-Sanchez et al., 2014). These tests differ in important test characteristics such as speech stimuli, type of background noise and test procedure. Speech-in-noise tests have the right properties for use as self-administered internet-based hearing screening tests (Smoorenburg, 1992, Smits et al., 2004, Culling et al., 2005, Smits et al., 2006, Jansen et al., 2010, Leensen et al., 2011a, Smits et al., 2013). The test can be performed quickly with minimal instructions, and its online application makes it easily accessible. The test is relatively independent from the absolute presentation level, as the ratio of speech intensity and level of masking noise is measured (Plomp, 1986, Smits et al., 2004, Wagener and Brand, 2005).

Furthermore, the test is relatively robust against variations in background noise and test equipment (Smits et al., 2004, Culling et al., 2005, Jansen et al., 2010). Such a test may facilitate audiometric hearing evaluation of noise-exposed employees in the workplace, as a trained audiometrist, a soundproof room, and specialised and costly technical equipment are not required (Stenfelt et al., 2011, Leensen and Dreschler, 2013b).

This study concentrates on the Occupational Earcheck (OEC), a Dutch online speech-in-noise test, developed at the Department of Audiology of the Leiden University Medical Center, commissioned by the Dutch National Hearing Foundation (Ellis et al., 2006). It is a test specifically designed to detect HFHL within a few minutes. The OEC is presented via headphones, which allows testing of both ears separately. The OEC was evaluated by Leensen et al (2011a), and shown to be reliable in laboratory conditions (with a standard error of measurement of 1.3 dB, and an intra-class correlation coefficient (ICC) of 0.68), but lacked discriminative power (with a sensitivity of 92% and specificity of 49%). Test precision was assessed by means of the steepness of the slope of the psychometric function (slope=11.0%/dB SNR). The OEC was significantly correlated with pure-tone average (PTA) of the frequencies 0.5, 1, 2, 4, and 3, 4, 6 kHz, and with the Dutch sentence SRT test (r=0.69, r=0.66, and r=0.77, respectively).

Adaptations involving the speech material, and the masking noise could improve the accuracy of the OEC in detecting HFHL. Possible adaptions include adjusting the root mean square levels of the words to achieve equal intelligibility, and filtering of the masking noise. Previous work suggested that a test with a spectrally filtered masking noise better distinguishes between normal-hearing (NH) and hearing-impaired (HI) listeners (Leensen et al., 2011b, Jansen et al., 2014b, Vlaming et al., 2014). A stationary low-pass (LP) filtered masker stimulates the use of high-frequency speech information, which is advantageous for NH listeners. This consequently increases the discriminative power of the test.

The aim of this study was to evaluate the 'OEC' after optimising its speech and noise material, and test procedure. A laboratory-based cross-sectional study was carried out on NH adults, and HI subjects with a HFHL, most probably related to noise exposure. The discriminative power of the optimised test in five different masking noise conditions was assessed. Furthermore, test validity was assessed by comparing pure-tone thresholds to test versions in different masking conditions. Finally, test reliability was assessed. These outcome measures were then used to select the version of the test which was most accurate in differentiating between NH and HFHL listeners, while remaining sufficiently valid and reliable.



## Methods

### Subjects

An a priori power analysis indicated that at least 15 subjects per NH and HI group would be necessary in order to attain a power of 80%, assuming a relevant difference in test outcome (SRT) of 2.6 dB SNR between the groups (Leensen et al., 2011b). A loss of subjects (due to non-attendance, drop-out or exclusion) was anticipated. Therefore up to 40 subjects were invited to participate. Study subjects were sampled by means of a two-gate design. NH participants were mainly students, recruited from the university and a neighbouring high school. HFHL subjects exposed to noise at the workplace were recruited from different industries with high noise exposure, including an orchestra, the construction industry, and a newspaper factory. All subjects were adults (18 years or older), and native speakers of Dutch. NH was defined as pure-tone thresholds of 20 dB HL or better at 0.125 to 6 kHz. HFHL was defined as pure-tone thresholds of 20 dB HL or better at the frequencies 0.125 to 1 kHz, and thresholds ≥ 25 dB HL for at least one frequency between 2 to 6 kHz. Subjects were excluded if they experienced language problems, had an asymmetrical hearing loss (i.e. a difference between the left and the right ear >30 dB at all frequencies), or a type of hearing loss other than HFHL.

In total, 36 subjects participated of which three subjects did not meet the inclusion criteria, and were excluded from further testing. The study population consisted of 18 NH subjects, and 15 subjects with a HFHL. Most of the participants were unfamiliar with online speech-in-noise testing. Details of the participants are listed in Table 2.1. The majority of the participants were male (66.7%). An independent samples t-test showed that the HFHL subjects were significantly older than NH subjects (p<0.001). The exact cause of the HFHL is unknown; however, all of these subjects had a self-reported history of occupational or leisure noise exposure. Seventeen participants (51.5%) were tested on the right ear. The mean volume level chosen by the NH subjects was 75.3 dBA (SD=4.9), and by the HFHL subjects 76.7 dBA (SD=5.0). An independent samples t-test showed that the chosen volume level did not differ significantly between the groups (p=0.363). Mean hearing threshold levels for NH and HFHL subjects are presented in Figure 2.1.

**Table 2.1.** Participant characteristics.

| | | NH (N=18) | HFHL (N=15) |
|---|---|---|---|
| *Male* | | 7 (38.9%) | 15 (100%) |
| *Mean age (years)* | | 27.3 (SD=12.7) | 56.3 (SD=7.0) |
| *Profession* | | | |
| | *Student* | 14 (77.8%) | 0 |
| | *Construction-related* | 1 (5.6%) | 9 (60%) |
| | *Music-related* | 1 (5.6%) | 2 (13.3%) |
| | *Other* | 2 (11.2%) | 4 (26.7%) |
| *Occupational noise-exposure* | | 2 (11.1%) | 14 (93.3%) |
| *Leisure noise exposure* | | 15 (83.3%) | 8 (53.3%) |
| *Use of hearing protection* | | 9 (50%) | 15 (100%) |

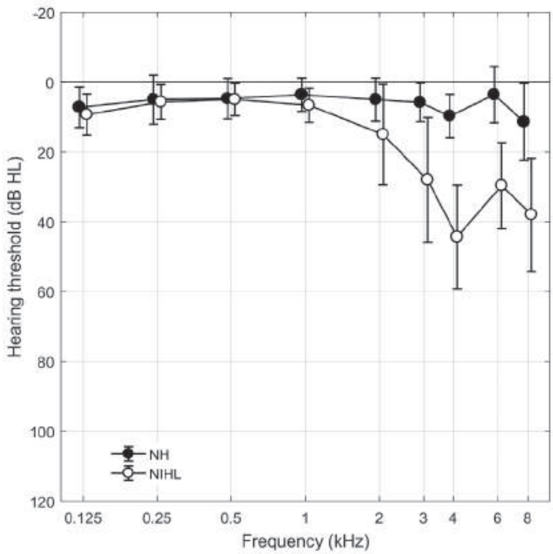


**Figure 2.1.** Audiometric thresholds for NH and HFHL subjects (for test ear). Error bars represent SDs.

### OEC

The speech material of the original OEC consisted of a closed set of nine Dutch consonant-vowel-consonant (CVC) syllables, represented by nine response buttons (identified by pictures and written words) on a computer screen. A tenth button labelled 'not recognized' was included. The words were selected from the Dutch word list used for diagnostic speech audiometry (Bosman, 1989), with a phonemic distribution representative of the Dutch language (Albrecht et al., 2005). The words contained high-frequency consonants, and were paired to contain a matching vowel (bed /bɛt/, knife /mɛs/, bag /tɑs/, pan /pɑn/, cat /pus/, book /buk/, sock /sɔk/, sun /zɔn/, arrow /pɛil/). By matching the vowels, listeners especially need the high-frequency speech information in order to identify the words. The recording was made using a female Dutch speaker. The OEC had no bandwidth limitations, and words were randomly presented in a stationary broadband noise, matched to the long-term average speech spectrum (LTASS). The volume level of the speech could be set by the user prior to testing.

The test was administered by means of the simple adaptive up-down procedure, with a step size of 2 dB. The first stimulus was presented at a SNR of 0 dB. After every correct response the subsequent stimulus speech level was decreased by 2 dB. After every incorrect response the SNR was increased by 2 dB. The SNR's presented ranged from -14 to +4 dB. For every listener, the SNRs until the first incorrect response was given, were not included in the SRT calculation, which resulted in an individual starting level. From this level, a total of 35 stimuli were presented to all listeners. The SRT at 50% was calculated by averaging the SNRs of the last 30 stimuli, for both ears separately. After finishing the test, the results ('good', 'moderate', 'insufficient', 'poor', or 'very poor') for both ears were directly reported to the user together with the appropriate advice. The intra-test standard deviation (SD) was calculated, showing the variation of SRT within the adaptive procedure. The intra-test SD gives an insight into the variation within a single test measurement, and can therefore be used as a measure of the accuracy of a test performed by an individual.

### Optimisation of the original OEC

The original OEC was optimised in three different ways: a) homogenization of the speech material; b) filtering of the masking noise and c) adaptation of the test procedure. In order to optimize the original OEC, past OEC test results

(N=7,933, of the period May 2007 to May 2014) were examined. Test results with intra-test SDs >3 dB were considered unreliable, and were therefore excluded. The mean age of test users was 36.5 years (SD=15.8).

***Homogenization of the speech material: Word selection and level corrections***

In order to develop a precise test, the intelligibility of the individual words included in the test should be as homogenous as possible. Therefore, the slopes of the word-specific psychometric functions were determined, and word intelligibility was equalised with level adjustments based on the average SRTs for the individual words. A logistic regression model was applied to past OEC data for each of the individual words, with the speech intelligibility (percentage correct words) as a function of the SNR of the test presentations. First, the data were corrected for the relative performance of each user. Then the data of all users, and for both ears were pooled for each word. To obtain the mean SRT, and slope at the 50%-point, following function was used (Smits et al., 2004):

$$SI(SNR) = \gamma + (1-\gamma)\frac{1}{1+e^{[-(SNR-SRT)4s]}}$$

Where, SI is speech intelligibility (the proportion correct at a given relative SNR), γ is guess level, and s is slope of the psychometric function at SRT. The model took into account the guess level γ (1/9=0.11), resulting from the closed set of nine words. The psychometric function of the word arrow /pɛil/ had a deviant slope (21.5%/dB SNR), which was much steeper as compared to the slopes of the other words-specific functions, ranging from 9.3 %/dB SNR to 15.6 %/dB SNR. To avoid the relatively easy recognition of this word based on its unique vowel (i.e. diphthong), this word was removed from the test. The remaining eight words were amplified (perceptually difficult words) or attenuated (perceptually simple words) according to their word-specific SRTs.

This procedure is in agreement with procedures used for other speech-in-noise tests with closed response sets (e.g. (Leensen et al., 2011b)), but deviates from recommendations in ISO 8253-3. Where the standard prescribes "to base such curves on a sufficiently large number of otologically normal persons of both sexes, aged between 18 and 25 years inclusive and for whom the test material is appropriate", there  was no information available about the pure-tone audiogram and it could not be verified which subjects were otologically normal, because the results were collected through the internet. However, a procedure was maintained that strongly reduced the differences between NH and HI listeners and it is expected that the relative difficulty within subjects is comparable for subgroups of NI and HI subjects.



### *Filtering of the masking noise: Enhancing the sensitivity for NIHL*

The LP filtered masking noises were created according to the methods described by Leensen et al (2011b). First, a broadband stationary masking noise was created, with the same spectral shape as the LTASS of the optimised word material. Then a set of four different LP filtered masking noises (indicated with LP) was derived by filtering the broadband stationary masking noise. Appropriate cut-off frequencies and noise floors were determined by speech intelligibility index (SII) predictions, according to ANSI S3.5 (1997). SII predictions were performed, in which relevant parameters of filtered noise conditions were varied to predict the effects on SRT for various audiograms. A more detailed description of the SII predictions can be found in Leensen et al (2011b). According to these predictions, LP filtered noises with cut-off frequencies of 1.4 and 1.6 kHz both discriminate well between NH, and HI individuals. To mask potential ambient noise levels, noise floors were presented at two different levels: -12 dB, and -15 dB. Both cut-off frequencies were combined with both noise floors. The five test versions are described in Table 2.2.

### *Adaptations of the test procedure: Test length*

To prevent unnecessarily long testing, and consequently, potential concentration problems in listeners, the influence of the number of stimuli per test on SRT and on intra-test SD was assessed. This was based on all past test results, i.e. including test results with intra-test SDs > 3 dB (n=9,429). Mean SRTs and intra-test SDs were calculated for different test lengths, in steps of five presentations, including total test lengths of 35, 30 and 25 presentations (starting from the individual starting level). The first five presentations were not included in the calculations. The test length did not influence SRT scores, with a mean SRT of -8 dB SNR for all test lengths. Mean intra-test SDs for the different test lengths did not differ either (range: 2.2-2.3 dB). The smallest mean intra-test SD was found for a total test length of 25 stimuli. Therefore the test length was shortened from 35 to 25 stimuli per ear.

**Table 2.2.** Characteristics of the test noises.

| Noise version | Filtering | Cut-off frequency | Noise floor |
|---|---|---|---|
| *LTASS* | - | - | - |
| *LP 1* | Low-pass | 1.4 kHz | -12 dB |
| *LP 2* | Low-pass | 1.4 kHz | -15 dB |
| *LP 3* | Low-pass | 1.6 kHz | -12 dB |
| *LP 4* | Low-pass | 1.6 kHz | -15 dB |

## Measurement procedures

The study protocol was approved by the medical ethics committee of the University of Amsterdam (number NL45730.018.13). All participants were informed and recruited by information letters. Informed consent was given before the start of the measurements.

All audiometric, and speech-in-noise tests were carried out in a soundproof booth at the audiological research department of the AMC. Pure-tone thresholds were assessed first using a Decos clinical audiometer (Decos Systems B.V., Noordwijk, the Netherlands), and TDH-39P headphones (Telephonics, Farmingdale, NY). Audiometric equipment was regularly calibrated using a B&K 2260 sound level meter (Brüel & Kjaer, Naerum, Denmark), and a B&K artificial ear type 4153 (Brüel & Kjaer). The audiogram was recorded at the octave frequencies from 0.125 to 8 kHz, including 3 and 6 kHz. Bone conduction was measured at 0.25, 0.5, 1, 2 and 4 kHz. Pure-tone audiometry was carried out by trained personnel. Subsequently each subject completed a session with the five different test versions of the OEC. The OEC was fully automated and presented using an Adobe Macromedia Flash player web application on a personal computer (Dell Precision T3500, US), which was directly connected to HDA 200 audiometric headphones (Sennheiser, Wedemark, Germany). The speech-in-noise tests were presented monaurally. All tests were presented to one ear of each subject, which was randomly assigned by the web application of the OEC. The order in which the different masking versions were presented was counterbalanced. The tests started after entering the participant's personal log-in code, which was linked to a certain sequence of tests. Instructions were given prior to testing, and the speech stimuli were presented once to familiarise the subject with the stimuli, and the response on the computer screen. The tests were performed at a volume level that was selected by the individual subject as comfortable and loud enough to understand the stimuli easily (ranging from 64 to 84 dBA). SNRs ranged from -30 dB to 0 dB, accounting for speech recognition in LP noises. The actual test started at the SNR after the first incorrect response, resulting in an individual starting level. The SRT was then calculated by averaging the last 20 out of 25 presentations. After completion of this test session a short break was given, followed by the retest (repetition of the OEC tests that were completed in the first session).

After completing the speech-in-noise tests, the participants were asked to fill in a short questionnaire. Details concerning age, gender, profession, occupational and non-occupational noise exposure, and use of hearing protection were requested. A flowchart of the measurement procedure is shown in Figure 2.2. Total test duration (audiometry, speech-in-noise testing, retesting, and questionnaire, including breaks) was 1.5 to 2 hours per subject. Participants were financially compensated.

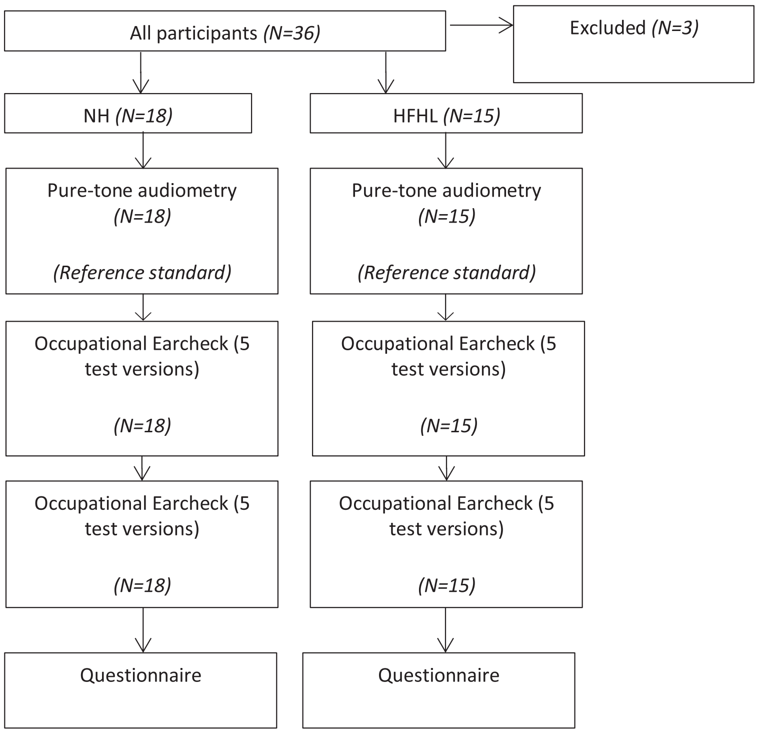


**Figure 2.2** Participant flowchart

# Results

## Test results of NH and HFHL subjects on the OEC

Test results of younger NH and older HFHL subjects were compared, in order to assess how well the different OEC test versions discriminate between the two subject groups. Mean SRT results of the first test for each test version are presented in Figure 2.3. The highest SRTs were obtained with the LTASS test version, while the lowest results are found for LP 2 and LP 4, the LP filtered versions with a noise floor of -15 dB. SRTs of NH and HI subjects for all test versions were compared by means of independent samples t-tests. The results are presented in Table 2.3. The differences in test results between groups were significant for all test versions. The difference in SRT scores was greater for the LP versions compared to the LTASS version.

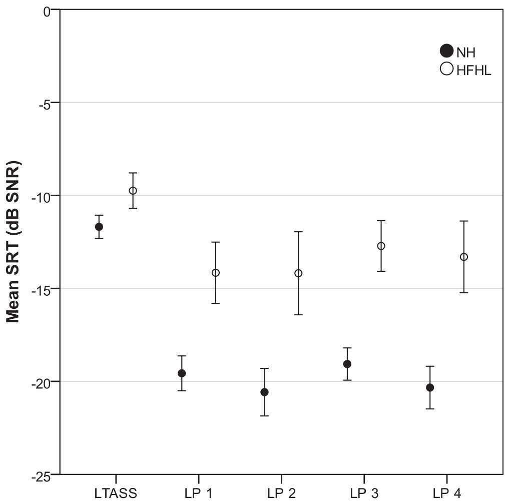


**Figure 2.3.** Mean SRT in dB SNR, for NH and HFHL subjects, for all test versions (OEC test). Error bars represent 95%-confidence intervals.

**Table 2.3.** Mean differences in SRT (dB SNR) (SD) for NH and HFHL subjects.

| Test version | NH Mean SRT | HFHL Mean SRT | Δ NH-HFHL | 95% CI | |
|---|---|---|---|---|---|
| *LTASS* | -11.7 (1.3) | -9.7 (1.7) | -1.9* | -3.0 | -0.9 |
| *LP 1* | -19.6 (1.9) | -14.2 (3.0) | -5.4* | -7.1 | -3.7 |
| *LP 2* | -20.6 (2.6) | -14.2 (4.0) | -6.4* | -8.8 | -4.0 |
| *LP 3* | -19.1 (1.7) | -12.7 (2.5) | -6.3* | -7.8 | -4.9 |
| *LP 4* | -20.3 (2.3) | -13.3 (3.5) | -7.0* | -9.1 | -5.0 |

**Differences are significant at p<0.001. All p-values are corrected using Bonferroni correction for multiple comparisons.*

## Sensitivity and specificity for HFHL

A receiver operating characteristics (ROC) analysis was performed to assess the monaural sensitivity (percentage HI subjects correctly classified as being HI), and specificity (percentage NH subjects correctly classified as being NH) of the different test versions of the OEC. A cut-off value for a dichotomous pass/fail outcome was chosen, based on a proper trade-off between sensitivity, and specificity values. Area under the curve (AUC), sensitivity, specificity and cut-off values for all test versions are shown in Table 2.4. The highest AUC value (0.98), and the highest sensitivity (93%) and specificity (94%) were found for test version LP 3.

## Test validity

In order to assess the validity of the OEC, the SRT results of the OEC tests were compared to the pure-tone audiogram. Pearson correlation coefficients for SRT, and the PTA of the frequencies important for overall speech intelligibility (PTA0.5,1,2,4), and the PTA of the higher noise-sensitive frequencies (PTA3,4,6) for all test versions are shown in Table 2.5. Correlations for all subjects, and for HFHL subjects only are given. For all subjects, the LP versions correlated slightly better with PTA compared to the LTASS version. For all subjects, SRT results of LP 2, LP 3 and LP 4 in particular were highly correlated with PTA3,4,6 (r=0.83 to r=0.85). A scatterplot showing SRT results against PTA3,4,6 for LP3, separated for NH and HFHL subjects, is presented in Figure 2.4 (upper figure). For the total group, all correlations were statistically significant. For the HFHL subjects, the correlations with PTA3,4,6 were significant.

## Test reliability

The test reliability was assessed in several ways. First, the test-retest variability was studied by analysing test and retest results. Then, the mean within-subject SD was calculated to assess the consistency of the test results. To get an insight into the degree of agreement between test and retest results, the ICC was calculated. Finally, to assess the precision of the test, psychometric functions were determined for all test versions. Test reliability measures are shown in Table 2.6.

Paired samples t-tests showed that there were small variations in test and retest results. The differences between test and retest were 1.2 dB SNR or smaller, and not significant for LTASS, LP 1 and LP 2. The mean within-subject SD was calculated by dividing the SD of the differences by the square root of 2. Mean within-subject SDs of 1.0 to 1.7 dB were found, with the smallest value for the LTASS version. A high degree of agreement was found between test and retest results of all subjects, for all LP versions, with ICCs of 0.84 to 0.89. A scatterplot showing test against retest results for LP3, separated for NH and HFHL subjects, is presented in Figure 4 (lower figure). The psychometric functions for all test versions were determined by means of logistic regression, with the speech intelligibility (percentage correct words) as a function of the SNR. For this purpose, the SNRs of all presentations within a test were corrected by the individual SRT of that test. Then the data of all users, and of NH and HFHL subjects separately, were pooled for each test version. The model took into account the guess level γ resulting from the closed set of eight words (1/8=0.125).The functions for NH and HFHL subjects separately are presented in Figure 2.5. The psychometric functions were shifted to the average SRT at 50% for each test version. Differences were found in the steepness of the slopes of the functions for the different test versions for the total group. The LTASS and LP 2 yielded the steepest slopes (14.8%/dB SNR and 13.6%/dB SNR, respectively), followed by LP 3 and LP 4 (13.1%/dB SNR and 12.5%/dB SNR, respectively). LP 1 yielded a slightly shallower slope of 10.6%/dB SNR.

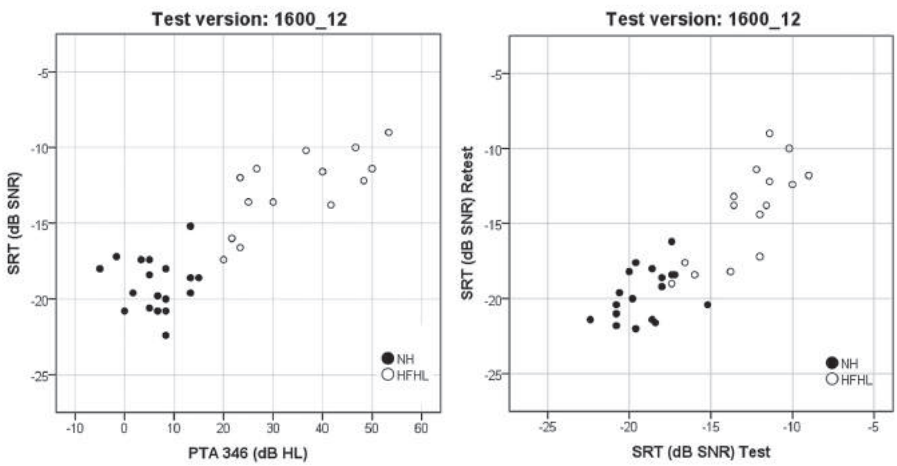


**Figure 2.4.** Scatterplots of SRT values against $PTA_{3,4,6}$ (left Figure), and test against retest results (right Figure), for test version LP 3, for NH and HFHL subjects.

**Table 2.4.** AUC, sensitivity and specificity, and cut-off value for pass/fail, for all test versions.

| Test version | AUC (95% CI) | Sensitivity (%) | Specificity (%) | Cut-off value SRT (dB SNR) |
|---|---|---|---|---|
| *LTASS* | 0.85 (0.70-1.00) | 80 | 78 | -10.7 |
| *LP 1* | 0.94 (0.85-1.00) | 87 | 94 | -17.5 |
| *LP 2* | 0.87 (0.74-1.00) | 80 | 94 | -17.3 |
| *LP 3* | 0.98 (0.95-1.00) | 93 | 94 | -16.9 |
| *LP 4* | 0.94 (0.86-1.00) | 87 | 100 | -16.3 |



**Table 2.5.** Bivariate correlation coefficients (Pearson's r) of the SRT values against the PTA of the frequencies 0.5,1,2,4 kHz ($PTA_{0.5,1,2,4}$) and 3,4,6 kHz ($PTA_{3,4,6}$) for all test versions (OEC test).

| | All subjects (N=33) | | HFHL subjects (N=15) | |
|---|---|---|---|---|
| **Test version** | $PTA_{0.5,1,2,4}$ | $PTA_{3,4,6}$ | $PTA_{0.5,1,2,4}$ | $PTA_{3,4,6}$ |
| *LTASS* | 0.65** | 0.74** | 0.62* | 0.75** |
| *LP 1* | 0.68** | 0.76** | 0.59* | 0.56* |
| *LP 2* | 0.66** | 0.83** | 0.50 | 0.73** |
| *LP 3* | 0.73** | 0.85** | 0.50 | 0.68** |
| *LP 4* | 0.68** | 0.83** | 0.41 | 0.61* |

**Significant at p<0.05. ** Significant at p<0.01.*

**Table 2.6.** Test-retest characteristics of NH and HFHL subjects.

| Test version | Group | Test Mean SRT (dB SNR) (SD) | Retest Mean SRT (dB SNR) (SD) | Mean Δ test-retest (dB) | Mean within-subject SD (dB) | ICC** *All (N=33)* | ICC** *HFHL (N=15)* |
|---|---|---|---|---|---|---|---|
| *LTASS* | NH | -11.7 (1.3) | -11.9 (0.9) | 0.3 | 1.0 | 0.63* | 0.60* |
| | HFHL | -9.7 (1.7) | -10.0 (1.3) | | | | |
| *LP 1* | NH | -19.6 (1.9) | -19.0 (1.6) | -0.5 | 1.4 | 0.84* | 0.84* |
| | HFHL | -14.2 (3.0) | -13.8 (3.0) | | | | |
| *LP 2* | NH | -20.6 (2.6) | -21.2 (2.3) | 0.4 | 1.6 | 0.87* | 0.83* |
| | HFHL | -14.2 (4.0) | -14.3 (4.0) | | | | |
| *LP 3* | NH | -19.1 (1.7) | -19.7 (1.7) | 1.0* | 1.5 | 0.84* | 0.68* |
| | HFHL | -12.7 (2.5) | -14.2 (3.2) | | | | |
| *LP 4* | NH | -20.3 (2.3) | -21.9 (2.3) | 1.2* | 1.7 | 0.87* | 0.74* |
| | HFHL | -13.3 (3.5) | -14.0 (3.3) | | | | |

** Significant at p<0.01. All p-values are corrected using Bonferroni correction for multiple comparisons.*
*** Intra-class correlation (ICC): using a two-way random model, type: absolute agreement, single measures.*

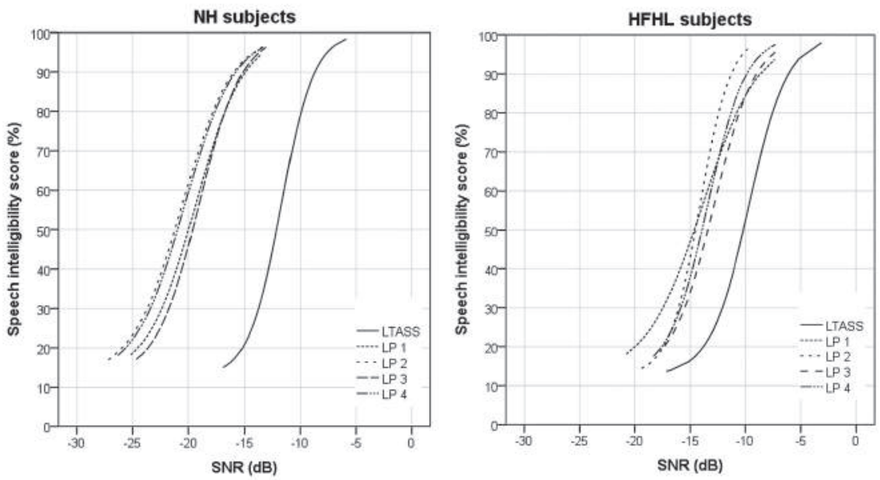


**Figure 2.5.** Psychometric functions for normal-hearing (NH) subjects (N=18) (left), and for high-frequency hearing loss (HFHL) subjects (N=15) (right), per test version.

# Discussion

This study evaluated the optimised internet-based speech-in-noise self-test, the OEC, in young NH subjects and older subjects with HFHL.

## Test results after optimisation

Overall, the improved OEC LP 3 version with a LP filtered stationary noise (with a cut-off frequency of 1.6 kHz and a -12 dB noise floor) appeared to be the most appropriate test, showing a reasonable sensitivity and specificity, and a strong correlation with PTA3,4,6 for the whole target group, while remaining reasonably reliable. Earlier work showed that the original OEC was not yet suitable for NIHL screening purposes (Leensen et al., 2011a). After adapting the speech and noise material of the OEC, substantial improvements in test characteristics were attained. A higher specificity of 94% was found. Also, a better correlation with PTA3,4,6 was achieved. The original test and the improved test were both evaluated in a different study sample, though both study samples showed similarities in demographic distribution. In another study a similar online speech-in-noise test developed for NIHL screening among teenagers, the Earcheck, was also improved by filtering of the masking noise (Leensen et al., 2011b). Earcheck with LP filtered noise discriminated best between NH and NIHL, and improved test sensitivity to 95%. In this study the

LP filtering resulted in test improvements in the same order of magnitude as those that were found for Earcheck.

### Sensitivity and specificity for HFHL

As a proof of concept, mean SRT results of younger NH subjects were compared to mean SRT results of older HFHL subjects. This demonstrated the feasibility of the test, as the test was able to distinguish between subjects with and without HFHL. The test versions with LP noises differentiated better between NH, and HFHL performance as compared to the unfiltered version. LP 2 and 4 showed the highest variation in SRT results. This may due to the lower noise floors, resulting in a higher masking release. Therefore, LP 3 appeared to be the best version, with a large mean SRT difference of about 6 dB SNR, and a low SD. The discriminative power of LP 3 was also reflected in the highest values for sensitivity and specificity, respectively 93 and 94%. The results can be well compared with the results of Vlaming et al (2014). They developed two high frequency (HF) tests, both with a LP filtered speech shaped noise masker, one using digit triplets, and one using CVC words. For the comparable HF-CVC test a sensitivity of 87% and a specificity of 94% was reported, using a similar definition of HFHL (i.e. PTAHF >20 dB).

Of the 18 NH subjects, one subject had a high SRT (of 15.2 dB SNR), and was therefore incorrectly classified. The subject was 52 years old, and had no specialties in the pure-tone audiogram (all hearing levels were 20 dB HL or better). For the retest, this subject obtained a much lower SRT of 20.4 dB SNR. This subject was assigned to a test sequence in which the LP 3 test version was presented first. This may have resulted in the large difference between test and retest. Of the 15 HFHL subjects, one subject obtained a low SRT of 17.4 dB SNR, and was therefore incorrectly classified. The subject was 60 years old, with a hearing level of 40 dB HL at 4 kHz (the hearing levels at all other frequencies were better than 20 dB HL). For the retest, the subject obtained a lower SRT of 19.0 dB SNR. For the retest, five HFHL subjects performed better, with SRTs smaller than the chosen cut-off value of 16.9 dB SNR. These subjects had a lower PTA3,4,6 as compared to the other HFHL subjects (mean PTA3,4,6 of 26 dB HL, and 38 dB HL, respectively). The test may therefore distinguish better between NH and more profound HFHL. Subjects with small degrees of HFHL may be classified incorrectly.

It is important to note, however, that this evaluation took place in a study sample which was not representative for the target group of noise-exposed employees. A two-gate design was used in order to establish a clearly defined group of known cases on the one hand, and healthy controls on the other. This biased selection resulted in significant age differences between the NH and the HI group. Moreover, this may have overrated the sensitivity and specificity of the test, as it became easier for the test to distinguish NH performance from HFHL performance. Subjects with other types of hearing loss, were excluded. This may have introduced an artefactual reduced variation, which may also have resulted in a biased estimation of the discriminative power of the test.

**Test validity**

Relatively high correlations of SRT results with the audiogram were found, especially with the higher noise-sensitive frequencies (PTA3,4,6). The strong correlation with the reference standard that was used for verification, reflected in a high criterion validity of the improved OEC. The correlations for the HFHL subjects group were lower, because of the smaller number of data points and greater variation in SRT.

Although the differences between the different OEC LTASS and LP versions were small, LP 3 showed the strongest correlations with PTA0.5,1,2,4 (r=0.73), and even a higher correlation with the higher frequencies (r=0.85), in all subjects. Vlaming et al (2014) reported a similar correlation with PTAHF of 0.79, and 0.82, for the high frequency triplet and CVC tests, respectively. Jansen et al (2014b) compared the broadband French digit triplet test (DTT) with a CVC test in standard speech-shaped noise, and with the CVC test in a LP filtered masking noise. They found comparable correlations with the higher frequencies (PTA2,3,4,6) for the DTT (r=0.85), and the CVC test in LP filtered noise (r=0.83).

**Test reliability**

Overall, the improved OEC had a better test reliability compared to the original OEC. Though the LP filtering of the masking noise did result in a loss of reliability compared to without the filtering.

The original OEC had a test-retest difference of 0.5 dB (Leensen et al., 2011a), while for the improved OEC in broadband noise this was 0.3 dB, though not significant. The test-retest differences for the improved OEC in LP noises were greater, although not significant for LP 1 and LP 2. For LP 3 a significant mean test-retest difference of 1.0 dB was found. The test-retest differences were greater for HFHL subjects, which is also reflected in the lower ICC of 0.68. Due to the applied measurement procedure (where a test and a retest session were compared, and in which the same word material was used in five different noise conditions within one session), the test-retest differences found in this study do not imply the expected learning effects in a screening context. Though, the applied procedure was necessary in order to select the most appropriate masking noise condition. The learning effect for OEC in a practical setting, in which the same word material is presented multiple times in the same noise conditions, needs to be established in future research. To eliminate a potential learning effect, OEC might have to be performed multiple times in a screening context.

The mean within-subject SD of the improved OEC in a broadband noise was smaller than those of the improved OEC in LP conditions. Jansen et al (2014b) reported similar measurement errors for the CVC test in broadband noise (1.0 and 1.1 dB, for the Flemish and French versions), and for the CVC test in LP filtered noise (1.2 and 1.6 dB for the Flemish and French versions).

The original OEC had a slope of 11.0%/dB SNR (Leensen et al., 2011a) and 11.6%/dB SNR. After homogenisation of the speech material, the slope of the improved OEC in stationary broadband noise was found to be 14.8%/dB SNR. The LP filtering of the noise, however, resulted in shallower slopes. LP 3 had a slope of 13.1%/dB SNR, which still surpassed the original broadband test. Vlaming et al (2014) reported a comparable slope of 12.1%/dB SNR for the HF CVC test. The slope that was found for OEC LP 3 was somewhat shallower as compared to the slopes that were reported for the DTT (Smits et al., 2004, Jansen et al., 2010), and for the HF-triplet test (Vlaming et al., 2014).

### Implications and future research

The current study was performed in a laboratory setting in a soundproof booth. To study whether the OEC is conceptually right, it was important to evaluate the test in clearly defined NH and HFHL groups under controlled conditions. However, the test is developed for occupational screening and monitoring purposes, and expected to be performed in poorly controlled, occupational environments. The OEC should be evaluated in samples of noise-exposed subjects with an unknown hearing status, and in more realistic occupational conditions. This is needed in order to study test properties more accurately, such as learning effects, and sensitivity and specificity, to establish an appropriate cut-off value for the pass/fail categories.

## Conclusions

An internet-based speech-in-noise self-test, the OEC was designed as a screening test for occupational noise-induced HFHL. This test was optimised, and validated among younger NH listeners and older listeners with HFHL, most probably related to noise exposure. The improved OEC, using a more homogenous set of monosyllables with high-frequency consonants and paired vowels, in combination with a LP filtered masking noise (with a cut-off frequency of 1.6 kHz, in combination with a noise floor of -12 dB) is an appropriate and reasonably reliable test for the discrimination between the study groups in a well-controlled setting. A good discriminative power, reflected in reasonable sensitivity and specificity values, was achieved. Awaiting further evaluation in the field, this study shows that the OEC is a potential tool for online self-screening and monitoring in occupational settings.

# Chapter 3.

# Cross-sectional evaluation of an internet-based hearing screening test in an occupational setting.

M. Sheikh Rashid[1]
M.C.J. Leensen[1]
J.A.P.M. de Laat[2]
W.A. Dreschler[1]

1 Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands
2 Department of Audiology, Leiden University Medical Center, Leiden, The Netherlands

Scandinavian Journal of Work, Environment & Health, 2017, 43(3), 279-286.

## Abstract

**Objectives.** The Occupational Earcheck (OEC) is an online internet test to detect high-frequency hearing loss, for the purposes of occupational hearing screening. In this study we evaluated the OEC in an occupational setting, in order to assess test sensitivity, specificity, and validity.

**Methods.** A cross-sectional study was conducted in 2015, in which the optimized OEC was evaluated on 94 employees from the army and three different companies in construction and manufacturing. Subjects underwent OEC in an office-like room. Pure-tone air conduction audiometry was performed as a reference test. The OEC was repeated for a subset of subjects (N=19). Important test characteristics (i.e. sensitivity and specificity, test validity, and test-retest reliability) were assessed.

**Results.** When analysed on the individual level, the sensitivity and specificity of OEC were 90% and 77%, respectively. The speech reception threshold results correlated strongly with the pure-tone average of the frequencies 3,4 and 6 kHz, reflecting good test validity (r=0.79). The difference between test and retest was not significant. The intra-class correlation coefficient was moderate (r=0.57), indicating a reasonable agreement between test and retest.

**Conclusions:** The OEC appears to be a suitable test for the detection of high-frequency hearing loss among noise-exposed employees, with good sensitivity and specificity values, even when performed in a semi-controlled occupational setting, though a possible learning effect should be taken into account.

## Introduction

High-frequency hearing loss (HFHL) caused by excessive exposure to noise in the workplace (also known as noise-induced hearing loss (NIHL)) is one of the most commonly reported occupational illnesses in the Netherlands (van der Molen et al., 2014). Various primary preventive measures for occupational HFHL exist, from interventions to control noise at the source, to the use of personal hearing protection devices. Primary preventive measures are not always effective (Verbeek et al., 2014). For this reason, secondary prevention of HFHL by screening employees exposed to noise becomes important. Early identification of HFHL may prompt actions to prevent progression of the hearing loss (Meyer-Bisch, 1996).



In many European countries, including the Netherlands, professional associations recommend that employees, who are exposed to noise levels greater than a time-weighted average of 80 dBA, be provided with a periodic audiometric evaluation (Sorgdrager et al., 2006). This evaluation should be offered annually in order to monitor the employees' hearing abilities closely. However in practice, audiometric evaluation is incorporated into the preventative occupational health examinations, which are not offered this frequently. Moreover, participation rates among the employees are often low (Jellema, 2014).

The traditional approach for occupational hearing evaluation is pure-tone air conduction audiometry. Though pure-tone air conduction audiometry is the reference standard in clinical assessments, it is a costly and time-consuming method for screening. Hearing threshold assessment for both ears may take 15 minutes, depending on the tester's and participant's experience and motivation, and the number of frequencies measured. Moreover, obtaining reliable pure-tone hearing thresholds in an occupational setting is challenging. Pure-tone thresholds are subject to variability due to tester, participant, and environmental factors, but test procedure and equipment also play a role (Schlauch and Carney, 2012, Carter et al., 2014).

Online speech-in-noise testing (for the measurement of auditory speech recognition abilities in noise), promises to be a valuable alternative tool for hearing screening. It is easily accessible, low cost, and broadly applicable (Smoorenburg, 1992, Smits et al., 2004, Culling et al., 2005, Smits et al., 2006, Jansen et al., 2010, Smits et al., 2013). It allows hearing assessment of at-risk employees in a remote setting, as it does not require specialized and costly technical equipment and therefore facilitates more frequent hearing assessments (Stenfelt et al., 2011). The test measures the speech reception threshold (SRT), a measure of the ability to understand speech in noise. The SRT is defined as the critical signal-to-noise ratio (SNR) necessary for a person to recognize 50% of speech material correctly.

Several online tests have been developed for the Dutch language. The first test was a digit triplet test: the National Hearing Test (Smits et al., 2004). Commissioned by the Netherlands Hearing Health Foundation, the department of Audiology of the Leiden University Medical Center developed the Occupational Earcheck (OEC), which is based on similar principles. It was specifically developed to monitor the hearing ability of employees in noisy occupations and raise awareness of the damaging effects of noise on hearing. The test is designed to be very precise, as it tests both ears monaurally. The OEC was optimized and validated in a well-controlled laboratory setting at our department and showed a sensitivity of 93%, and a specificity of 94% for the detection of HFHL (Sheikh Rashid et al., 2017c). However, the test should also be evaluated in a noise-exposed population in an occupational environment in order to assess whether it is appropriate for screening purposes. In this study, we evaluated the OEC further in an unselected sample of noise-exposed subjects and in more realistic occupational settings than the laboratory environment. Our main objective was to evaluate whether the improved OEC is a valid and reliable screening test to detect HFHL in a high-risk population.

## Methods

### Study population

The study participants were recruited from the army and three different companies in construction and manufacturing. With consent of the company management, information letters were sent to employees of several noisy departments in the companies and the army. In total, 102 employees volunteered to participate. Participants were adults (≥18 years) and Dutch speakers. The medical ethics committee of the University of Amsterdam approved the study protocol (number 2013_231). Informed consent was obtained for all subjects.

### Measurement procedure

A cross-sectional study was carried out in 2015. The index test (OEC) and the reference test (pure-tone air conduction audiometry) were performed in a single test session during which the subject's demographic details (including gender, age, and occupational noise exposure) were also collected by means of a short questionnaire. The question concerning occupational noise exposure was: "*How many days a week do you work in noise* [*noise is defined as sound levels >80* dBA*, or when talking with a raised voice at a distance of 1 m is required*]?" The measurements were performed at five representative occupational test locations, in quiet office-like rooms. One of the companies had multiple sites, therefore the measurements were performed at two different locations. Ambient noise level measurements were performed at the test sites prior to testing. The audiometric test conditions of all test locations met the international standards for hearing screening (i.e. unmasked air conduction starting at 500 Hz; ISO 8253, part I) when sound attenuating cups are used in combination with the headphones.



Each subject completed the OEC on their own with minimal supervision by the testers. A subgroup (every 5th subject) repeated the OEC a second time. Hereafter pure-tone air conduction audiometry was performed as a reference. Both ears were measured at the octave frequencies 500–8000 Hz, including 3000 and 6000 Hz. Pure-tone air conduction audiometry was performed by two trained test operators using an Interacoustics AC40 or AD 229b clinical audiometer in combination with TDH 39 headphones with sound attenuating cups (Amplivox audiocups). For the OEC measurements, a research laptop and Sennheiser HDA 200 headphones were used. The testers who evaluated OEC were not aware of the results of the pure-tone air conduction audiometry, and vice versa. A complete measurement including instructions and informed consent (5 minutes), questionnaire (5 minutes), OEC (5 minutes) and pure-tone air conduction audiometry (15 minutes), took about 30–35 minutes per subject (5 minutes extra for a retest).

### Occupational Earcheck

The speech material of OEC consists of a closed set of eight Dutch consonant-vowel consonant (CVC) words: *(bed /bɛt/, knife /mɛs/, bag /tɑs/, pan /pɑn/, cat / pus/, book /buk/, sock /sɔk/, sun /zɔn/)*. They are represented by eight response buttons on a visual screen, identified by a picture and a written word. A ninth button labelled "not recognized" is included. The words were selected from the Dutch wordlist used for diagnostic speech audiometry (Bosman, 1989) and contain matching vowels and high-frequency consonants, making the test more sensitive for the detection of HFHL. In order to acquire a precise test, the intelligibility of the individual words in noise was equalized with level adjustments. These level adjustments were derived from the slopes of word-specific psychometric functions, based on previously performed tests (Leensen et al., 2011b). The test is presented in a stationary masking noise, matched to the long-term average speech spectrum of the words, except for the higher frequencies: the matched masking noise is low-pass filtered (cut-off frequency 1.4 kHz), and has a noise floor of -12 dB SNR. The test consists of 25 stimuli per ear, making it a relatively short test which can be performed within five minutes.

Test presentation is monotic: both left and right ear are tested separately. The sequence of the ears is randomly assigned by the OEC. The volume level of the stimuli can be set by the user to a comfortable loudness by means of a volume scale, resulting in individual test intensities. The test is administered by means of the simple adaptive up-down procedure with a step size of 2 dB. The first stimulus is presented at a SNR of 0 dB. With every correct response, the subsequent stimulus level is decreased by 2 dB, and with every incorrect answer the stimulus is increased by 2 dB. The noise level remains fixed throughout the test. The SNRs presented range from -30–0 dB. The actual calculation starts at the SNR of the first incorrect response, resulting in an individual starting level. The SRT is calculated by averaging the SNRs of stimuli 6–25 per ear. The intra-test standard deviation (SD) is calculated using the same stimuli and gives an insight into the variation within a single test measurement. It can therefore be used as a measure of the accuracy of a test performed by an individual.

### Statistical analyses

A sample size calculation was performed, indicating that ≥79 subjects were needed in order to detect a meaningful correlation of r=0.58 between SRT results and a pure-tone average (PTA) of the higher frequencies (Leensen and Dreschler, 2013a). This sample size would provide 80% power to discover a correlation which is statistically different from a moderate correlation of r=0.30 at the 0.05 significance level. Descriptive statistics were performed on demographic information, and pure-tone thresholds. True HFHL on the basis of pure-tone air conduction audiometry was defined as a PTA of the frequencies 3,4 and 6 kHz (PTA346) of 25 dB HL or worse. SRT results of the OEC in dB SNR for the first ear tested were compared for HFHL and non HFHL ears by means of an independent samples t-test. To assess test validity, the OEC SRT results of the first ear measured were compared to PTA346 in dB HL of the corresponding ear by means of a Pearson product correlation coefficient. To further assess the discriminative power of the test, a receiver operating characteristic (ROC) analysis was performed on SRT results of the first ear measured. By means of this analysis, an appropriate cut-off value for pass/fail of the OEC was estimated, and corresponding test sensitivity and specificity values for detecting HFHL were assessed monaurally. To assess the sensitivity and specificity on the individual level, true HFHL was defined as a PTA346 of 25 dB HL or worse for at least one ear (HFHL 1+). Both ears of one subject had to have a lower score than the cut-off value of OEC in order to pass the screening test. An individual with a test result equal to or higher than the cut-off value for at least one ear would get a positive test result. To assess test reliability, test and retest results of the first ear measurement of a subgroup were compared with a paired sample t-test. Two parameters were calculated, the intra-class correlation coefficient (ICC, two-way random, absolute agreement, single measures), and the measurement error. The ICC was calculated to get an insight into the degree of agreement between test and retest results. In order to assess the consistency of the test results, the measurement error was calculated by taking the quadratic mean of the within-subject standard deviations of the repeated measurements. Data were analysed using SPSS statistics version 22 (IBM Corp, Armonk, NY, USA).

## Results

In total, 102 subjects volunteered to participate; 6 did not attend on the day of the test and 2 were excluded from analysis due to invalid OEC measurements (OEC test was presented on both ears at the same time, instead of one ear). The remaining 94 subjects all performed the index test (OEC) and the reference test (pure-tone air conduction audiometry). The flow of the participants through the study is depicted in Figure 3.1: 30 subjects had a HFHL (1+), of which 17 had a HFHL at both ears, 4 at the right ear only, and 9 at the left ear only; 64 subjects did not have a HFHL. Of the 30 subjects with HFHL (1+), all were male, with a mean age of 52.3 years (SD 7.3). The majority reported working in noise for at least half a working day per week, with an average of 3.8 days a week (SD 1.5) (Information concerning this question was missing for 1 subject in this group). A majority of the 64 non HFHL subjects were male (92.2%), with a mean age of 36.4 years (SD 10.6). The majority reported working in noise for at least half a working day per week, with an average of 3.1 days a week (SD 1.9) (Information concerning this question was missing for one subject in this group). Across the five test locations, only small variations in gender, age, and SRT scores were observed. The distribution of audiometric hearing threshold levels for HFHL and non HFHL ears is shown in Figure 3.2.

In order to assess how well the OEC discriminates HFHL from non HFHL, SRT test results of HFHL ears were compared to those of non HFHL ears (for the first ear tested). The mean SRT was -11.4 dB SNR (SD=4.2) for HFHL ears, and -16.7 dB (SD=2.2) for non HFHL ears. The difference of 5.3 dB SNR was statistically significant ($P<0.001$). To assess the validity of the OEC, the SRT results of the first ear tested were compared to the pure-tone audiogram of the corresponding ear. As shown in Figure 3.3, SRT results correlated strongly with PTA346 ($r=0.79$, $P<0.001$).

A ROC analysis was used to assess the most appropriate cut-off value for a dichotomous pass/fail outcome with the best trade-off between sensitivity and specificity values using monaural data of the first measurement. The highest agreement between hearing thresholds and OEC test results was found when the cut-off value was set at -14.9 dB SNR (Figure 3.4). This setting resulted in a sensitivity of 83% and a specificity of 75% in order to identify HFHL with PTA346 of 25 dB HL or worse. The area under the curve (AUC) was 0.89 [95% confidence interval (95% CI) 0.81–0.97]. Table 3.1 presents the OEC results (positive for at least one ear and negative for both ears) compared to pure-tone air conduction audiometry results (HFHL for at least one ear and non HFHL) on the individual level. When taking both ears into account, the sensitivity was 90% and the specificity was 77%.

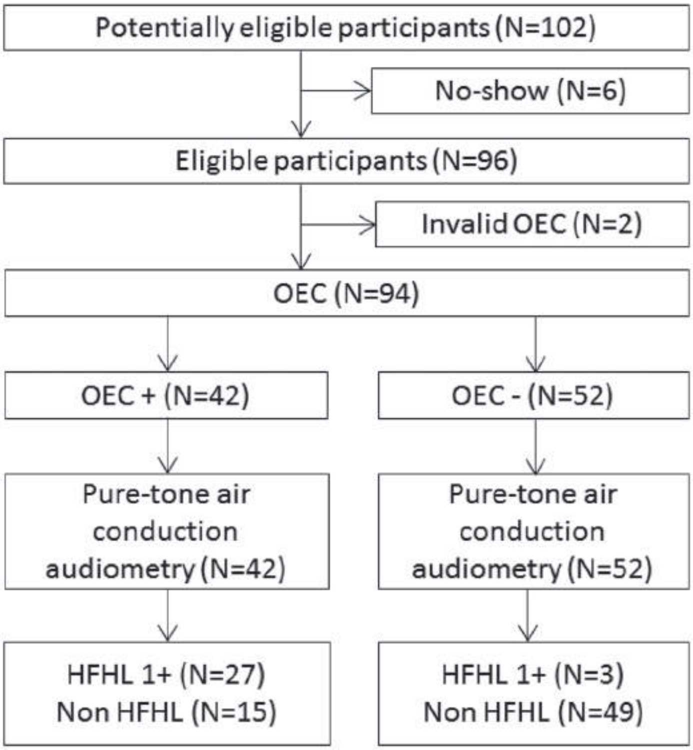


**Figure 3.1.** Flowchart of study participants.
*[N=number of participants. OEC=Occupational Earcheck. HFHL= High-frequency hearing loss].*

A subgroup of 19 subjects performed the OEC twice. The mean SRT scores for test and retest (for the first ear) were compared. Performance on retest, with a mean SRT of -16.9 dB SNR (SD=2.4) was better than on the initial test, with a mean SRT of -16.0 dB SNR (SD=3.0). This indicated a learning effect of 0.9 dB SNR, but this was not statistically significant (95% CI -0.3–2.1, P=0.12). The test and retest results were moderately correlated, with an ICC of 0.57 (P=0.003). The measurement error was 1.8 dB SNR.

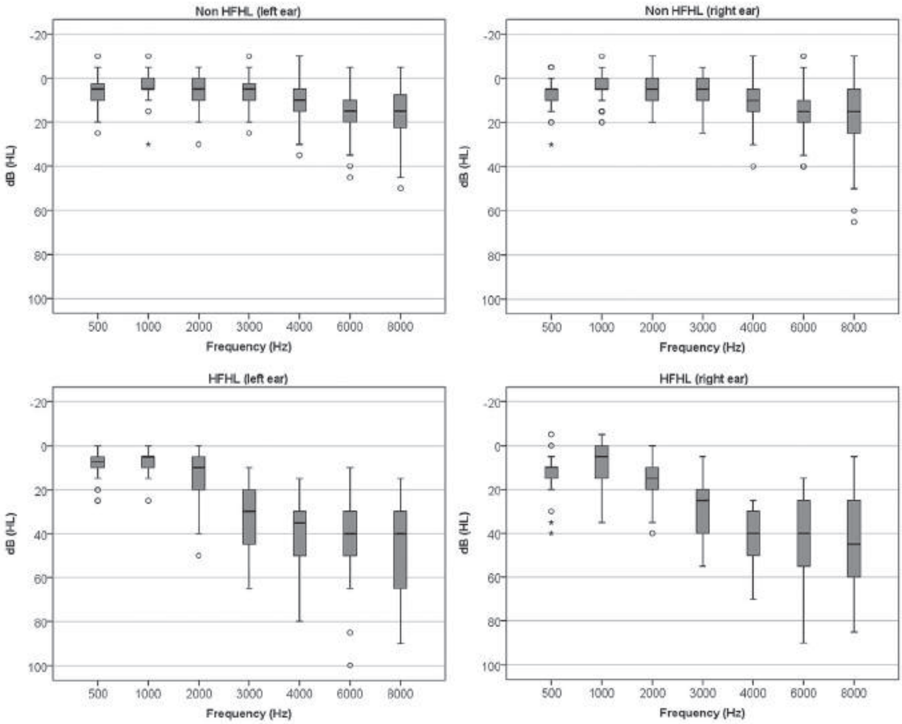


**Figure 3.2.** Boxplots presenting pure-tone air conduction audiometry threshold distribution for non high-frequency hearing loss (Non HFHL) ears and high-frequency hearing loss (HFHL) ears, left and right ears separately.

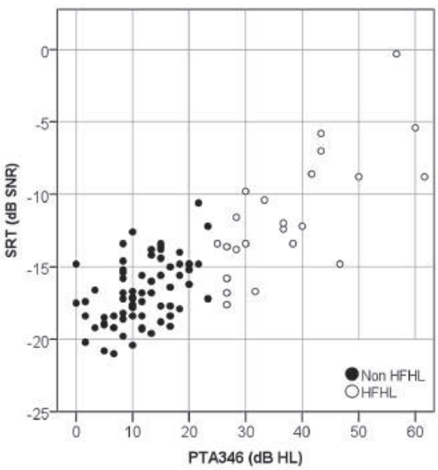


**Figure 3.3.** Scatterplot of speech reception threshold (SRT) values in dB SNR against the pure-tone average of the frequencies 3,4, and 6 kHz (PTA346) for the first ears measured. Black symbols represent non high-frequency hearing loss (Non HFHL) ears, and white symbols high-frequency hearing loss (HFHL) ears.

**Table 3.1.** Two-by-two Table: Test scores on the individual level.

| | | Pure-tone air conduction audiometry[a] | | |
|---|---|---|---|---|
| | | *HFHL 1+* | *Non HFHL* | *Total* |
| **OEC result[b]** | *Positive 1+* | 27 | 15 | 42 |
| | *Negative* | 3 | 49 | 52 |
| | *Total* | 30 | 64 | 94 |

*[a] True high-frequency loss for at least one ear (HFHL 1+) is defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) according to the pure-tone air conduction audiometry test. b Occupational Earcheck (OEC) result based on a cut-off value of -14.9 dB SNR to discriminate between a positive result for at least one ear (1+), and a negative result for both ears.*



## Discussion

The OEC distinguished well between HFHL and non HFHL ears, with a significant difference between the mean SRT results of 5.3 dB SNR for the first ear measurement. The test showed a high correlation of 0.79 between SRT results and PTA346. In this study, a sensitivity of 83% and a specificity of 75% was found. The high AUC (0.89) value indicated good test accuracy. These analyses were based on test results of single ear measurements. Results of each of a subject's ear were studied separately in order to properly assess the OEC's test properties. In order to reduce the possible influence of a learning effect, we used the results of the first ear tested for this measurement. However, for practical screening purposes, the main focus is on the outcome at the level of the individual tested, and both ears per subject should be taken into account. The assessment of test results on the binaural level is important in order to make the correct decisions for referral, further comprehensive audiological assessment, and recommendations for the appropriate intervention. Therefore, sensitivity and specificity values were established on the individual level as well. Based on the classification of HFHL for at least one ear versus no HFHL for both ears, the sensitivity (or proportion of true positives) on the individual level increased to 90% and the specificity (or the proportion of true negatives) to 77%.

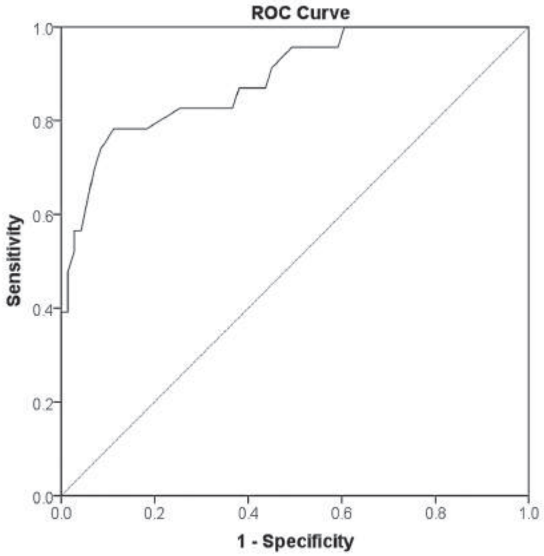


**Figure 3.4.** ROC curve, presenting sensitivity and specificity for the Occupational Earcheck on monaural basis, for different cut-off values for pass/fail outcome.

In a well-controlled laboratory-based study of the OEC with normal-hearing subjects and HFHL subjects a sensitivity of 93% and a specificity of 94% were found, as well as a high correlation between SRT results and high frequency PTA (r=0.83) (Sheikh Rashid et al., 2017c). We found poorer test characteristics (sensitivity, specificity and correlation with high frequency PTA) in this study in an occupational setting. A possible explanation for the differences found is that the laboratory study had a study sample, which consisted of young normal-hearing students on the one hand, and known HFHL cases on the other. The current study consisted of an unselected group of noise-exposed employees, classified as either having a HFHL or not. Noise-induced HFHL might have been the most probable hearing loss in this high-risk population, however, age-related hearing losses (i.e. presbyacusis) could not be ruled out, as the HFHL group was significantly older compared to the non HFHL group. This can be attributed both to a longer period of noise exposure and to the (early) effects of presbyacusis. Furthermore, the reference standard was carried out differently in both studies. In the lab study, clinical pure-tone air conduction audiometry was performed in a soundproof booth, while in the current study, it was performed in poorer testing conditions, which may have led to less reliable measurements.

Jansen et al (2013), performed a similar study, in which noise-exposed workers completed the broadband digit triplet SRT self-test in an office-like room at five different industrial settings. Their findings were slightly more favorable relative to the findings presented in this paper. They found a higher sensitivity and specificity for detecting mild HFHL (92% and 89%, respectively), and a lower measurement error (0.8 dB). The differences in findings may be explained by their use of digit triplets in a broadband noise. The simplified speech material is less influenced by non-auditory cognitive abilities, and – in combination with the broadband noise – leads to more reliable estimations of the SRT. The use of meaningful words in a speech-in noise test such as in OEC, however, may be valuable for screening purposes as it is representative of daily communication situations experienced by the population being screened. Also, the use of a low-pass filtered noise instead of an unfiltered broadband noise has shown to improve the discrimination between HFHL and normal hearing/other losses ((Leensen et al., 2011b, Jansen et al., 2014b). Differences in study methods (such as the chosen definition of HFHL, measurements for one or both ears, and the calculation of the measurement error) and study population may also have explained the differences found between the studies.



The OEC can serve as a valuable screening method for HFHL in occupational settings. We aimed to develop a test that can improve a reliable differentiation of HFHL from normal hearing, and isolated low-frequency hearing losses. A comprehensive diagnostic audiological evaluation, is only indicated when the OEC result is positive. HFHL identified by OEC is probably related to noise exposure, but may also reflect another form of HFHL. The actual type and degree of the hearing loss should then be specified in further full diagnostic audiological evaluation after which appropriate measures can be advised.

This study showed some difficulties concerning the practical implementation of the OEC. An important issue was the reasonable test–retest reliability. The relatively large measurement error found may be due to a learning effect between both ear measurements within one test. Only a small subgroup performed the test twice, so even though we did not find a statistically significant difference between test and retest, a possible learning effect cannot be ruled out, and its influence on test results remains unknown. A learning effect may have led to higher estimated SRT values (especially for the first ear measured) and the relatively high number of false positive HFHL classifications.

The 77% specificity found at the individual level would in practice result in a large proportion of employees incorrectly identified as having a HFHL, and consequently unnecessarily referred for comprehensive testing. The high false-positive rate may decrease by introducing a retest for subjects with a positive test score. It is important to further investigate the effects of a direct automatic retest on test sensitivity and specificity of the OEC applied at an individual level.

Another important limitation is that the study population consisted of volunteers, creating a risk of sample selection bias. This type of bias should not influence the comparison of pure-tone air conduction audiometry results with OEC results, as both tests were performed by all participants. However, this bias may have affected certain study population characteristics such as the prevalence and the severity of HFHL, as more health conscious employees or employees with significant hearing problems may have volunteered to participate. As the severity of hearing loss is associated with sensitivity and specificity, the values that were established in this population may not be entirely applicable to other populations of noise-exposed employees.

The study demonstrated a good agreement between test result and hearing status according to the conventional audiogram. However, this optimal cut-off value of the pass/fail outcomes was determined post hoc, and may have led to an overestimation of the accuracy of the OEC. For these reasons it is important to validate the new threshold criteria in other noise-exposed samples.

Future studies concerning the development of the OEC should focus on its applicability to specific populations, its feasibility in different testing environments, and its special requirements. For instance, the OEC may be used as a monitoring tool and be applied on an annual basis to identify small changes in hearing. Therefore, the test–retest reliability of OEC should be assessed in more detail, taking into account the learning effect between tests.

### Concluding remarks

In this study, we assessed the accuracy of OEC for screening purposes in realistic occupational settings. This paper demonstrated that the OEC is able to detect HFHL, even in less optimal occupational settings. A good discriminative power was achieved, as reflected by the sensitivity and specificity values of 90% and 77%, respectively.

# Chapter 4.

# Accuracy of an internet-based speech-in-noise hearing screening test for high-frequency hearing loss: Incorporating automatic conditional rescreening.

M. Sheikh Rashid[1]
W.A. Dreschler[1]

1 Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands

Revised manuscript submitted to International Archives of Occupational and Environmental Health.

## Abstract

**Purpose.** To validate the accuracy of an internet-based speech-in-noise hearing screening test for high-frequency hearing loss (HFHL) 'Occupational Earcheck (OEC)' incorporating an automatic conditional rescreening, in an occupationally noise-exposed population. Secondary objectives were to assess the effects of age on test accuracy measures, and to assess the test accuracy for different degrees of HFHL.

**Methods.** A study was conducted on cross-sectional data of occupational audiometric examinations, including the index test OEC and reference standard pure-tone air conduction audiometry, of 80 noise-exposed workers. Sensitivity, specificity, and likelihood ratios were calculated for the OEC, after automatic conditional rescreening, for a younger and an older age group, and for two degrees of HFHL (HFHL25: PTA3,4,6 ≥ 25 dB HL, and HFHL35: PTA3,4,6 ≥ 35 dB HL, both for at least one ear).

**Results.** Test specificity for HFHL25 after a single test was 63%, and improved to 93% after the automatic conditional rescreen. Test sensitivity for HFHL25 decreased from 65% to 59%. Test sensitivity and specificity including automatic conditional rescreening for HFHL35 was 94% and 90%, respectively. The positive likelihood ratio for HFHL25 was 8.4, and for HFHL35 9.4. The negative likelihood ratio for HFHL35 was below 0.1.

**Conclusions.** The OEC is an appropriate screening test, especially for HFHL35. Normal-hearing workers who obtained a positive test result for the first test for one or two ears, benefit from having an automatic rescreen, resulting in an improvement of the test specificity, and hence prevent unnecessary referral.

## Introduction

High-frequency hearing loss (HFHL) caused by excessive exposure to noise in the workplace, also known as noise-induced hearing loss (NIHL), is an important public health problem worldwide (May, 2000, Sliwinska-Kowalska and Davis, 2012). In the Dutch construction industry it is one of the most commonly reported occupational diseases (van der Molen et al., 2016b). Therefore, secondary prevention (i.e. early identification) of HFHL by screening is of great importance, and stimulates to take actions to prevent progression of the hearing loss (Meyer-Bisch, 1996).

Over the past few years several internet-based speech-in-noise self-tests have been developed and investigated (Smits et al., 2006, Jansen et al., 2010, Leensen et al., 2011b, Watson et al., 2012, Molander et al., 2013, Paglialonga et al., 2014, Vlaming et al., 2014, Williams-Sanchez et al., 2014). Studies have shown that these tests can be used as a proper screening tool (Smoorenburg, 1992, Smits et al., 2004, Culling et al., 2005, Smits et al., 2006, Jansen et al., 2010, Leensen et al., 2011b, Smits et al., 2013). These tests facilitate audiometric hearing evaluation of noise-exposed workers in the workplace: a trained audiometrist, a soundproof room, and specialized and costly technical equipment are no longer required, as is the case for the more conventional pure-tone air conduction screening audiometry (Stenfelt et al., 2011, Leensen and Dreschler, 2013a).



This study focuses on the Occupational Earcheck (OEC), a Dutch internet-based speech-in-noise hearing screening test for occupational HFHL, developed at the Department of Audiology of the Leiden University Medical Center, commissioned by the Netherlands Hearing Health Foundation (Ellis et al., 2006). A phased approach was maintained to evaluate this test for screening purposes in noise-exposed workers. In the first phase the concept was improved for HFHL and tested in a well-controlled laboratory setting in a population that was recruited by means of a two-gate design, with normal-hearing cases on the one hand, and known HFHL cases on the other (Sheikh Rashid et al., 2017c). In the second phase, the improved test was evaluated in an unselected group of noise-exposed employees in a quiet office-room at the work place (Sheikh Rashid et al., 2017b). The discriminative ability of OEC was calculated on the individual level, which means that the results of both ears were taken into account. Based on the classification of HFHL for at least one ear versus no HFHL for both ears, the sensitivity on the individual level was 90% and the specificity was 77%. A relatively large measurement error was found, possibly due to a learning effect between the single ear measurements within one test. The learning effect may have led to higher estimated SRT values, especially for the first ear measured, and the relatively high number of false positive HFHL classifications. Though learning was accounted for by training, and a long individual run-up to the actual measurement was incorporated in the test, a learning effect still appeared.

In a screening setting, even a small learning effect may result in an incorrect classification due to the dichotomous test outcome. Normal-hearing listeners who have trouble with understanding the test procedure or who are not yet familiar with the speech material, may incorrectly receive a positive test score. A potential solution to this problem is to provide a second test opportunity for the initial referrals. Listeners may benefit from an automatically offered rescreen, provided for the ear(s) with a poor result, as the final classification (pass or referral) will be based on the last test result.

The objective of this study was to validate the test accuracy of OEC incorporating a new procedure with an automatic conditional rescreening, in a representative study population of noise-exposed workers. Test accuracy measures, including sensitivity, specificity, predictive values, and likelihood ratios were calculated. Secondary objective was to assess the effect of automatic conditional (i.e. sequential) rescreening of the positives on test accuracy measures. Another secondary objective was to establish the test accuracy for different degrees of HFHL, and for different age groups.

## Methods

### Study population

The study population consisted of occupationally noise-exposed employees from two manufacturing companies in the Netherlands who voluntarily performed an occupational audiometric examination provided by their employers, which is according to the Dutch Working Conditions Act. Subjects were 18 years or older and were speakers of the Dutch language. There were no exclusion criteria. The employees were informed by their employer by means of an information letter, and gave approval for sharing their results with researchers of the Amsterdam Medical Center for research purposes. According to the Medical Ethics Committee of the University of Amsterdam official approval of this study was not necessary, as the Medical Research Involving Human Subjects Act does not apply to this study (reference number W17_254 # 17.297).

### Measurement procedure

This prospective cross-sectional study was based on data from occupational audiometric examinations of noise-exposed workers that were performed in 2016. For every employee results of the index test OEC were collected. As a reference, pure-tone air conduction thresholds were collected by means of pure-tone air conduction audiometry. Demographical data on gender and age were collected.



### Occupational Earcheck

The speech material of OEC consisted of a closed set of eight equally intelligible Dutch consonant-vowel consonant (CVC) words with matched vowels, represented by eight response buttons on a visual screen, identified by a picture and a written word. A ninth button labelled 'not recognized' was included. The speech material was presented in a stationary low-pass filtered masking noise. Test presentation was monotic; both left and right ear were tested separately. The sequence of the ears was randomly assigned by OEC. The first stimulus was presented at a signal-to-noise ratio (SNR) of 0 dB, and with every correct response the subsequent stimulus level was decreased by 2 dB, while with every incorrect answer it was increased by 2 dB. The noise level remained fixed throughout the test. The noise level could be set by the user to a comfortable loudness by means of a volume scale, resulting in individual test intensities. The actual measurement started at the SNR of the first incorrect response, resulting in an individual starting level. Total test length per ear measurement was shortened to twenty presentations. The speech-reception threshold (SRT) was calculated by averaging the SNRs of the last ten stimuli. The intra-test standard deviation (SD) of the last ten stimuli gave an insight into the variation within a single test measurement. The previously established cut-off value of -14.9 dB SNR was used for pass/fail (Sheikh Rashid et al., 2017b). In order to achieve a good (i.e. negative) result for OEC, a subject would need a SRT score of <-14.9 dB SNR for both ears. A subject would get a poor (i.e. positive) result for OEC if the test result of at least one ear was ≥-14.9 dB SNR. More details on the development of OEC are described elsewhere by Sheikh Rashid et al. (2017c, 2017b).

The test was performed on an Apple Ipad with on-ear HQ-HP113LW headphones in a quiet office room at the work setting. OEC self-tests were minimally supervised by testers of the Netherlands Hearing Health Foundation. The testers were not aware of the results of the pure-tone air conduction audiometry. A sequential test design was applied. Listeners with a positive test result on the first test, automatically received a rescreen. The rescreening was conditional: A retest was only provided for the ear(s) with a positive test result, or with an intra-individual SD of >3 dB. Based on previous research, test results with an intra-individual SD of >3 dB were considered unreliable (Sheikh Rashid et al., 2017c).

### Pure-tone air conduction audiometry

Pure-tone air conduction audiometry was performed by professional audiometrists in sound-insulated office cabins, with ambient sound levels of 31 and 34 dBA, at both work settings, with the use of the clinical audiometers Madsen Micromate 304 (Otometrics) and Voyager 522, connected to TDH39 headphones. The headphones were provided with sound-attenuating Amplivox audiocups, because it could not be guaranteed that the audiometric test conditions of the office cabins met the international standards for hearing screening (ie, unmasked air conduction starting at 500 Hz; ISO 8253, part I, 2010). The audiometers were calibrated and were in compliance with the norm EN 60645-1 (ANSI S3.6, Type 2). Pure-tone air-conducted hearing thresholds were collected for both ears for the octave frequencies between 0.25 and 8 kHz (and additionally for 3 and 6 kHz). The audiometrists were not aware of the OEC results of the workers.

### Statistical analyses

Descriptive statistics were performed on demographic information, and pure-tone thresholds. True HFHL on the basis of pure-tone air conduction audiometry was defined as a pure-tone average (PTA) of the frequencies 3, 4, and 6 kHz (PTA346) of 25 dB HL or worse for at least one ear (HFHL25). A second, higher, degree of HFHL was defined as a PTA346 of 35 dB HL or worse for at least one ear (HFHL35). When thresholds for certain frequencies were missing, the adjacent thresholds were interpolated. Two-by-two contingency tables were used to compare the performance of OEC with pure-tone air conduction audiometry. Test properties were calculated, including sensitivity and specificity[1], positive and negative predictive values[2], and positive and negative likelihood ratios[3] (sensitivity/1-specificity, and 1-sensitivity/specificity), for the single screen versus the conditional rescreen, for two degrees of HFHL, and for separate age groups. To assess the effect of age, the workers were divided into a younger age group (≤45 years), and an older age group (>45 years). Likelihood ratios were calculated to overcome the disadvantage of a single cut-off value, and to apply the results of OEC to the individual (Parikh et al., 2009), making them useful for screening practice. Data were analyzed using IBM SPSS Statistics 24.



## Results

In total, data of 80 noise-exposed workers were available. All workers performed the index test (OEC) and the reference test (pure-tone air conduction audiometry). A STARD diagram is given in Figure 4.1, to report the flow of participants in the study. We could not analyze the effects of gender, because the vast majority of the subjects were male (N=78 (97.5%)). The mean age was 44.0 years (SD=11.5). About half of the participants underwent a rescreen for at least one ear (N=42 (52.5%)). In total, 55 ears were rescreened, of which 52 ears with a positive test result (8 of these ears also had an intra-individual SD>3 dB). Three ears with a negative test result were rescreened due to an intra-individual SD>3 dB. Figure 4.2 presents a scatterplot of first test and rescreen results for

1 The sensitivity of the test reflects the proportion correctly identified individuals with HFHL among all individuals with HFHL. The specificity reflects the proportion correctly identified non HFHL individuals among all non HFHL individuals.

2 The positive predictive value is the probability that the individual has hearing loss when OEC shows a positive result. The negative predictive value is the probability that an individual is non HFHL when OEC shows a negative result.

3 The positive likelihood ratio is the ratio of the probability of a positive OEC test in workers with HFHL to the probability in non HFHL workers. The negative likelihood ratio is the ratio of the probability of a negative OEC test in workers with HFHL to the probability in non HFHL workers.

all ears that were retested. The prevalence of HFHL25 (for at least one ear) was 42.5% (34 out of 80 workers). Four workers (5%) had a HFHL25 at the right ear only, and nine (11.3%) workers had a HFHL25 for the left ear only. Twenty one workers (26.3%) had a HFHL25 for both ears. The remaining 46 subjects (57.5%) showed normal results on both ears at the OEC test. Figure 4.3 presents mean hearing thresholds of both ears, for non HFHL25 individuals, and individuals with HFHL25 for at least one ear. The prevalence of HFHL35 was 22.5% (18 out of 80 workers). The group of ≤45 years (N=41) had a mean PTA346 of 12.8 dB HL (SD=13.5) for the right ear and 15.0 dB HL (SD=15.0) for the left ear. The older age group (N=39) had a mean PTA346 of 26.3 dB HL (SD=16.6) for the right ear and 28.4 dB HL (SD=15.3) for the left ear. The differences between the younger and the older group in mean PTA346 for both the left ear and the right ear were statistically significant ($p<0.001$).

The mean SRT score based on the single screen was -15.5 dB SNR (SD=3.1) for the right ear, and -15.5 dB SNR (SD=3.3) for the left ear. The mean intra-individual standard deviation was 2.0 dB for both the left ear and the right ear. The mean SRT score including the conditional rescreen was -16.2 dB SNR (SD=3.1) for the right ear, and -16.0 dB SNR (SD=3.2) for the left ear. The mean intra-individual standard deviation for the right ear was 1.9 dB (SD=0.6), and for the left ear 2.0 dB (SD=0.6). The correlation coefficient for PTA346 and OEC results including conditional rescreen, was 0.57 for the right ears ($p<0.01$), and 0.61 for the left ears ($p<0.01$).

Table 4.1 presents the OEC results (positive for at least one ear versus negative for both ears) compared to pure-tone air conduction audiometry results (HFHL and non HFHL) for HFHL25. Thirty-four workers had a HFHL for at least one ear, as determined by the reference test. In the first test, 24 of these workers with a HFHL were correctly identified by OEC (i.e. the true positives). In seventeen workers, the OEC wrongly identified a hearing loss (i.e. the false positives). Twelve workers with HFHL were wrongly labeled as non HFHL (i.e. the false negatives), while 29 non HFHL correctly received a negative result (i.e. the true negatives). The sensitivity was 65%, and the specificity was 63%. When taking the results into account of the automatic conditional rescreen, sensitivity decreased to 59%, while specificity increased to 93%. Table 4.2 presents the OEC results compared to pure-tone air conduction audiometry results for

HFHL35. Eighteen workers had a HFHL for at least one ear, as determined by the reference test. The sensitivity was 100% and the specificity was 66%. When taking the results into account of the automatic conditional rescreening, sensitivity decreased to 94%, and specificity increased to 90%. Table 4.3 presents the association of the single screen versus the conditional rescreen, with the presence and absence of HFHL25 and HFHL35 for the total group, and the two age groups. For HFHL25 high positive likelihood ratios were found for the conditional rescreen in all workers (8.4), and for the age group >45 years (8.1). For HFHL35 high positive likelihood ratios were found for the conditional rescreen in the total group (9.4), and for the younger age group (20). Also, for HFHL35 low negative likelihood ratios were found in case of the conditional rescreen (0.07 for the total group, and 0.08 for the older group). High negative predictive values were particularly found for HFHL35, with and without the conditional rescreen.

4

**Table 4.1.** Two-by-two contingency tables: HFHL25, for the single screening (upper table), and for the conditional rescreening (lower table).

| Single screen | | **Pure-tone air conduction audiometry**[a] | | |
|---|---|---|---|---|
| | | *HFHL25* | *Non HFHL* | *Total* |
| | *Positive* | 22 | 17 | 39 |
| **OEC result**[b] | *Negative* | 12 | 29 | 41 |
| | *Total* | 34 | 46 | 80 |

| Conditional rescreen | | **Pure-tone air conduction audiometry**[a] | | |
|---|---|---|---|---|
| | | *HFHL25* | *Non HFHL* | *Total* |
| **OEC result**[b] | *Positive* | 20 | 3 | 23 |
| | *Negative* | 14 | 43 | 57 |
| | *Total* | 34 | 46 | 80 |

*[a] True high-frequency hearing loss for at least one ear (HFHL25) is defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) of 25 dB HL or worse, according to the pure-tone air conduction audiometry test. b Occupational Earcheck (OEC) result based on a cut-off value of -14.9 dB SNR to discriminate between a positive result for at least one ear, and a negative result for both ears.*

**Table 4.2.** Two-by-two contingency tables: HFHL35, for the single screening (upper table), and for the conditional rescreening (lower table).

| Single screen | | Pure-tone air conduction audiometry[a] | | |
|---|---|---|---|---|
| | | *HFHL35* | *Non HFHL* | *Total* |
| **OEC result[b]** | *Positive* | 18 | 21 | 39 |
| | *Negative* | 0 | 41 | 41 |
| | *Total* | 18 | 62 | 80 |

| Conditional rescreen | | Pure-tone air conduction audiometry[a] | | |
|---|---|---|---|---|
| | | *HFHL35* | *Non HFHL* | *Total* |
| **OEC result[b]** | *Positive* | 17 | 6 | 23 |
| | *Negative* | 1 | 56 | 57 |
| | *Total* | 18 | 62 | 80 |

[a] *True high-frequency hearing loss for at least one ear (HFHL35) is defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) of 35 dB HL or worse, according to the pure-tone air conduction audiometry test. b Occupational Earcheck (OEC) result based on a cut-off value of -14.9 dB SNR to discriminate between a positive result for at least one ear, and a negative result for both ears.*

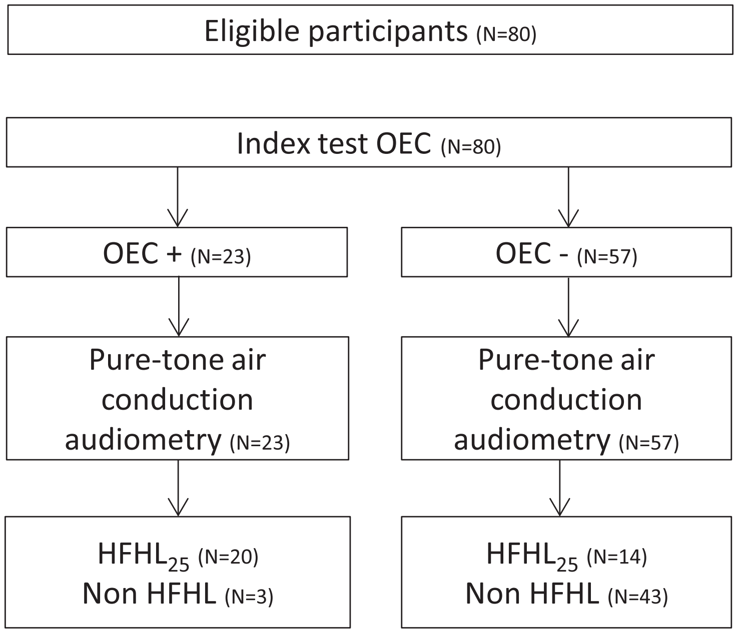




**Figure 4.1.** STARD diagram, with classification based on results of automatic rescreen. Index test is OEC. Reference test is pure-tone air conduction audiometry. Target condition is HFHL25 (for at least 1 ear). N=number of participants. OEC=Occupational Earcheck. HFHL= High-frequency hearing loss.

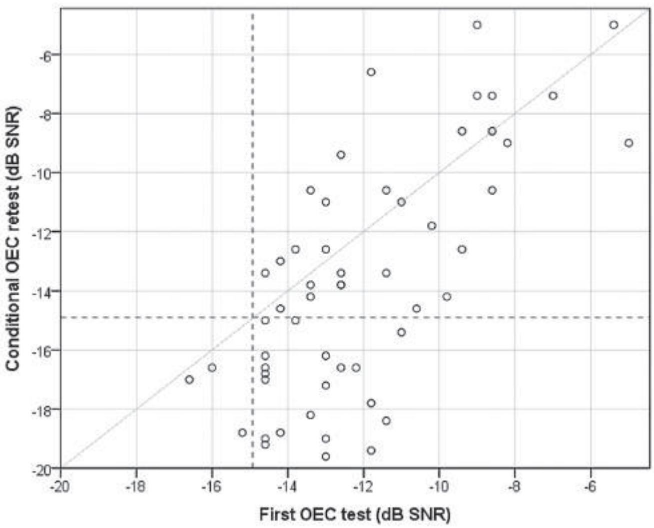


**Figure 4.2.** Scatterplot of (first) OEC test and retest (rescreen) results for all retested ears (N=55). The horizontal and vertical interrupted lines depict the cut-off value for pass/fail, set at -14.9 dB SNR.

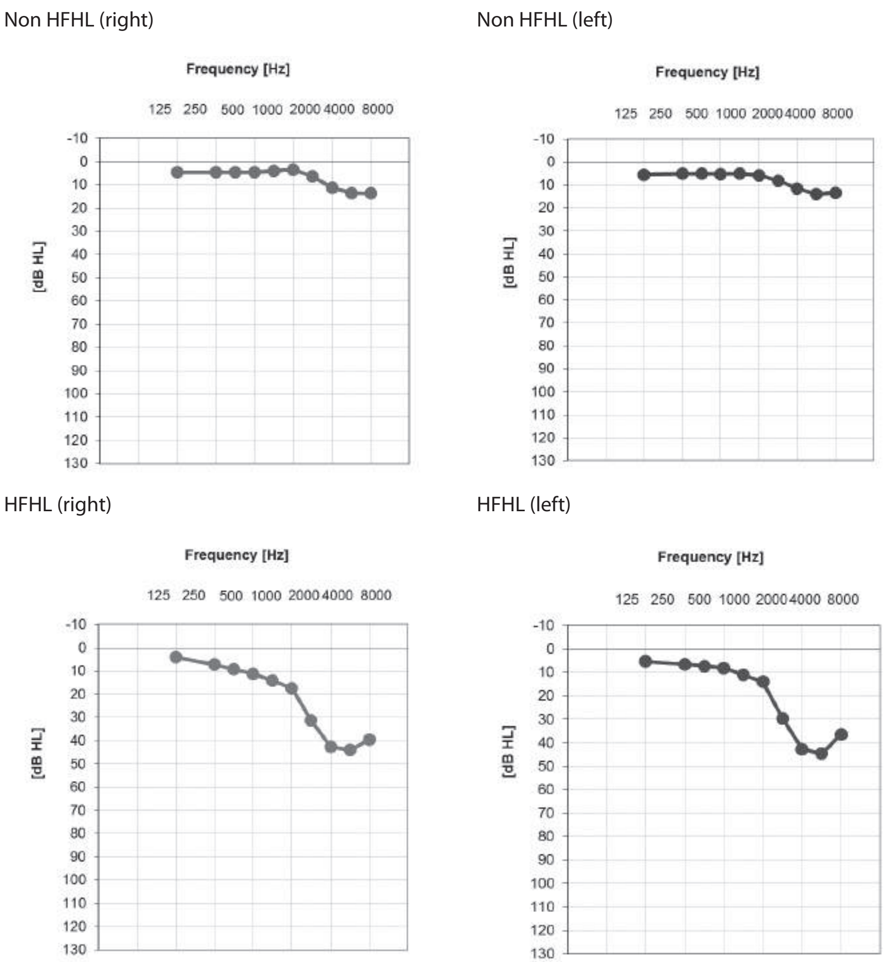


**Figure 4.3.** Mean pure-tone air conduction audiometry thresholds for non high-frequency hearing loss (Non HFHL) ears (upper panels) and for ears with high-frequency hearing loss defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) of 25 dB HL or worse (HFHL25) (lower panels), The thresholds for left and right ears are presented separately.

**Table 4.3.** Association of single screen and conditional rescreen, and population (all, young, and old) with the presence and absence of HFHL25 and HFHL35*, expressed as sensitivity, specificity, predictive values, and likelihood ratios**.

| Screen | Degree HFHL | Population | Sensitivity | Specificity | PV+ | PV- | LR+ | LR- |
|---|---|---|---|---|---|---|---|---|
| Single screen | $HFHL_{25}$ | All | 65% (22/34) | 63% (29/46) | 56% (22/39) | 71% (12/41) | 1.8 | 0.6 |
| | | ≤45 years | 50% (4/8) | 70% (23/33) | 29% (4/14) | 85% (23/27) | 1.7 | 0.7 |
| | | >45 years | 69% (18/26) | 46% (6/13) | 72% (18/25) | 43% (6/14) | 1.3 | 0.7 |
| | $HFHL_{35}$ | All | 100% (18/18) | 66% (41/62) | 46% (18/39) | 100% (41/41) | 2.9 | - |
| | | ≤45 years | 100% (3/3) | 71% (27/38) | 21% (3/14) | 100% (27/27) | 3.4 | - |
| | | >45 years | 100% (15/15) | 58% (14/24) | 60% (15/25) | 100% (14/14) | 2.4 | - |
| Conditional rescreen | $HFHL_{25}$ | All | 59% (20/34) | 93% (43/36) | 87% (20/23) | 75% (43/57) | 8.4 | 0.4 |
| | | ≤45 years | 38% (3/8) | 94% (31/33) | 60% (3/5) | 86% (31/36) | 6.3 | 0.7 |
| | | >45 years | 65% (17/26) | 92% (12/13) | 94% (17/18) | 57% (12/21) | 8.1 | 0.4 |
| | $HFHL_{35}$ | All | 94% (17/18) | 90% (56/62) | 74% (17/23) | 98% (56/57) | 9.4 | 0.07 |
| | | ≤45 years | 100% (3/3) | 95% (36/38) | 60% (3/5) | 100% (36/36) | 20 | - |
| | | >45 years | 93% (14/15) | 83% (20/24) | 78% (14/18) | 95% (20/21) | 5.5 | 0.08 |

**HFHL25 = high-frequency hearing loss for at least one ear, defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) of 25 dB HL or worse. HFHL35 = high-frequency hearing loss for at least one ear, defined as a pure-tone average of the frequencies 3,4,6 kHz (PTA346) of 35 dB HL or worse.*
*** PV+ = positive predictive value, PV- = negative predictive value, LR+ = positive likelihood ratio, LR- = negative likelihood ratio.*

## Discussion

In this study conventional pure-tone air conduction audiometry results were compared to results of the online speech-in-noise hearing screening test OEC for HFHL in a population of noise-exposed workers. For HFHL25 a moderate sensitivity of 65% and specificity of 63% was found. Automatic conditional rescreening significantly improved the specificity of the test to 93%. Especially the older population seemed to benefit from a second chance, with an increase in specificity of 46% to 92%. Sequential testing seems to be beneficial as it further reduced the number of false positives. Although, testing duration increased, the total number of false-positives incorrectly referred for further audiological assessment significantly decreased. The positive likelihood ratio of 8.4 indicates that OEC is particularly able to rule in HFHL25with a reasonably high degree of confidence. In other words, if workers achieve a positive (i.e. poor) test score on OEC, it can be quite certain that they actually have HFHL25, as the majority of non HFHL individuals would not have such high SRT results. On the other hand, the sequential rescreening lead to a deterioration in test sensitivity of 65% to 59%, especially in the younger population (50% to 38%), which indicates that part of the younger workers with a HFHL were still able to achieve a negative result on the rescreen. The lower negative predictive values indicate the uncertainty of the actual hearing status of the workers with a negative (i.e. good) score.

For the more moderate HFHL35, however, OEC is both highly sensitive and specific. The positive likelihood ratio of nearly 10 indicates that the OEC is able to rule in HFHL with a high confidence, while the negative likelihood ratio below 0.1 provides strong evidence that OEC is also able to rule out HFHL. Furthermore, with a positive likelihood ratio of 20, the OEC is strongly predictive of the detection of HFHL35 in younger workers.

Test accuracy was investigated for two age categories. The test sensitivity was lower in the younger population (except for HFHL35, after the conditional rescreen), while the specificity was lower in the older population. This implies that the younger workers were more often able to achieve a negative test result despite of a HFHL, as compared to the older workers. This may be due to the severity of the HFHL, as the severity of the target condition determines the probability of finding positive test results (Moons et al., 1997). Age is associated with the severity of the HFHL; the older workers showed larger hearing losses as compared to the younger workers.

In an earlier evaluation of OEC in a noise-exposed population higher sensitivity and specificity values were found, even without rescreening, namely 90% and 77%, respectively (Sheikh Rashid et al., 2017b). This may be due to the fact that the cut-off point for pass/fail was derived post hoc from the same population, which may have overestimated the accuracy of the test. Furthermore, sensitivity and specificity values may vary across populations due to selection bias, as well as due to variations in population characteristics (Moons et al., 1997), including age, and the severity of the hearing loss. Leensen & Dreschler investigated the internet-based speech-in-noise test Earcheck, which is based on the same principles as OEC, and found a comparable moderate sensitivity of 68% and specificity of 71% in 249 male construction employees (mean age= 49.7 years) for one screening round (2013a). Jansen et al. compared the Digit Triplet test, a consonant-vowel-consonant test with words with the same vowel (CVC), and a CVC test with a low-pass filtered (CVC_LP) with high-frequency PTA in 118 noise-exposed workers (age range= 22-59 years) (2013, 2014b). A higher sensitivity of 92%, and a specificity of 89% to detect mild HFHL (defined as a PTA2346 above 10 dB HL) was found for the Digit Triplet test (Jansen et al., 2013).

For the CVC tests an increased measurement error and a weaker correlation with PTA2346 was found as compared to the more reliable Digit Triplet test (CVC: R=0.86, CVC_LP: R=0.79, Digit Triplet: R=0.86) (Jansen et al., 2014b). These studies, however, did not account for different ages when investigating sensitivity and specificity. Also, they did consider a single screening round only.



Sekhar et al. considered the effect of a two-step screening on test sensitivity and specificity in HFHL screening in adolescents (2016). State school-based hearing screens, threshold tests at 250 to 8000 Hz using pulsed pure tones conducted in the school library, were compared to the gold standard sound-treated booth testing. Initial referrals returned for repeated screening. Following the two test rounds, specificity improved (from 49.5% to 84.6%), while sensitivity maintained (76.7%). In the current study specificity improved as well, however, sensitivity decreased slightly. In the study by Sekhar et al. (2016), the two test rounds of threshold testing only reduced the number of false-positives, while for OEC, the number of false-negatives increased as well. This may be well explained by the learning effect that OEC encounters.

An important limitation of this study was that the study participants were not randomly selected. The employees voluntarily participated in an occupational audiometric examination, because they were more health-conscious, or more worried about their hearing ability. This may have resulted into selection bias, affecting the prevalence and severity of HFHL. Therefore, the values of the test properties of OEC may differ in other noise-exposed populations. Another important limitation of this study is that one of the two audiometrists did not include the octave frequencies 3 and 6 kHz, which are important for the diagnosis of HFHL according to the audiogram. Therefore, the adjacent frequencies were interpolated for 60 of the 80 workers. As a consequence, the measurement accuracy of the high frequency test point 4 kHz weighted more heavily as compared to that of the other frequencies. Furthermore, it is not clear whether the HFHL in the workers was related to noise. It is important to note that the HFHL could have been a combination of noise-induced hearing loss and presbycusis. For the purpose of this study the most important result is that OEC is able to discriminate between HFHL and non HFHL, despite the actual cause of the hearing loss.

For further practice it is important to consider the actual goal of screening with OEC in certain situations. According to this study OEC appears to be quite suitable if the goal is to rule in/out moderate HFHL or worse, especially in younger populations. This means that OEC provides an important tool for the identification of individuals who are likely to benefit from preventive measures in order to prevent worsening of the hearing loss, or in more severe cases, from hearing aids. If the goal is, however to screen for early/mild HFHL (HFHL25), OEC would probably miss out on a significant percentage of cases, but would be quite specific (i.e. low number of false-positives). In that case, the chance that non HFHL workers will have a positive result and unnecessarily be referred to further audiological assessment would be small. This may be cost efficient, as unnecessary expensive and invasive audiological diagnostic assessment can be avoided. The false-negatives could possibly be detected in another screening round, for instance by means of annual screening. Future studies on OEC may therefore focus on (the potential learning effects on) periodic screening. Furthermore, future research may also focus more on variations in test accuracy parameters due to variations in (sub)populations, including differences in prevalence and severity of HFHL.

## Conclusions

In this study the test accuracy of OEC for screening of HFHL in a noise-exposed population was validated. Automatic conditional rescreening seems to be beneficial, considerably improving test specificity. With a moderate test sensitivity of 59%, but a high test specificity of 93%, the test is particularly able to rule in mild HFHL25 with a reasonably high degree of confidence. OEC appears to be a more accurate screening test for higher degrees of HFHL (HFHL35), with a high test sensitivity of 94%, and a high test specificity of 90%. The accuracy of OEC may vary across different occupational noise-exposed populations. This should be explored further.

# Part II

## The application of internet-based screening tests in children and teenagers.

# Chapter 5

# Application of the online hearing screening test "Earcheck": Speech Intelligibility in Noise in Teenagers and Young Adults.

M. Sheikh Rashid[1]
M.C.J. Leensen[1]
W.A. Dreschler[1]

[1] Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands

Noise Health,2016, 18(85), 312-318.

## Abstract

**Objective:** The objective was to describe the speech intelligibility in noise test results among Dutch teenagers and young adults aged 12–24 years, using a national online speech reception threshold (SRT) test, the Earcheck. A secondary objective was to assess the effect of age and gender on speech intelligibility in noise.

**Design:** Cross-sectional SRT data were collected over a 5-year period (2010–2014), from participants of Earcheck. Regression analyses were performed, with SRT as the dependent variable, and age and gender as explaining variables. To cross-validate the model, data from 12- to 24-year olds from the same test distributed by a hearing aid dispenser (Hoorscan) were used.

**Results:** In total, 96,803 valid test results were analyzed. The mean SRT score was –18.3 dB signal-to-noise ratio (SNR) (standard deviation (SD)= 3.7). Twenty-five percent of the scores was rated as insufficient or poor. SRT performance significantly improved with increasing age for teenagers aged 12–18 years by 0.49 dB SNR per age-year. A smaller age-effect (0.09 dB SNR per age-year) was found for young adults aged 19–24 years. Small differences between male and female users were found.

**Conclusion:** Earcheck generated large quantities of national SRT data. The data implied that a substantial number of users of Earcheck may have some difficulty in understanding speech in noise. Furthermore, the results of this study showed an effect of gender and age on SRT performance, suggesting an ongoing maturation of speech-in-noise performance into late adolescence. This suggests the use of age-dependent reference values, but for this purpose, more research is required.

## Background

Prolonged exposure to high levels of noise is known to result in noise-induced hearing loss (NIHL). This results not only in decreased detection of sounds, as reflected by poorer pure-tone thresholds, but also in deterioration in supra-threshold processing affecting speech processing abilities and speech discrimination. Tinnitus is also often reported as an additional consequence of noise-induced hearing damage. Dose–response relationships are predominantly based on the effects of occupational noise exposure. However, in recent years, recreational noise exposure has received increasing attention. Technological advances, such as the proliferation of personal music players (PMPs), are believed to have dramatically increased recreational noise exposure (Zhao et al., 2010, Levey et al., 2011, Portnuff et al., 2011). This had led to an increase in the risk of NIHL, especially among young people, who are not only exposed to noise during the use of PMPs, but also when visiting clubs, music festivals, or concerts (Serra et al., 2005).



Literature regarding the relationship between recreational noise exposure and hearing loss has revealed inconsistent results. Some studies reported an increase in the prevalence of NIHL as a result of exposure to recreational noise (Meyer-Bisch, 1996, Niskar et al., 2001a, Shargorodsky et al., 2010), whereas others failed to prove a dose–response relationship (Mostafapour et al., 1998, Zhao et al., 2010, Henderson et al., 2011, Jin et al., 2013). A recent systematic review of the literature on this subject by Carter et al. (2014), based on 265 articles, concluded that there is insufficient evidence to determine the extent of the risk of recreational noise. Nevertheless, the review also concluded that a significant proportion of young people are exposed to noise levels that are high enough to cause hearing damage. In most of the studies summarized by Carter et al. (2014), NIHL was measured by pure-tone audiometry screening, and differences in testing conditions and definitions of NIHL used resulted in inconsistent results. The authors suggest using other methods of detecting or quantifying NIHL, for example, supra-threshold tests like speech-in-noise assessments.

Difficulty in understanding speech in noisy situations is often one of the first signs of NIHL. This specific hearing disability can be tested more accurately by means of a speech-in-noise test than by traditional pure-tone audiometry. As such, speech-in-noise tests have more relevance than pure tone audiometry, and a poor result on such a test is more convincing to the listener. Speech-in-noise tests have been shown to be suitable for use as self-administered internet-based hearing screening tests (Smits and Houtgast, 2005, Jansen, 2013, Leensen and Dreschler, 2013a, Smits et al., 2013). They are less burdening compared to pure-tone audiometry tests. Moreover, the online application of these tests provides the opportunity to reach a large population of individuals at risk for NIHL and to collect large quantities of national data. Results of online speech-in-noise tests performed at home or at a remote setting have been previously studied, however not specifically in teenagers and young adults. To enhance targeted hearing education and prevention, it is essential to have a good understanding of potential threat of NIHL in this age group.

Earcheck (Dutch: Oorcheck, www.oorcheck.nl), a Dutch online speech-in-noise test, was developed to detect NIHL. The test was developed, validated, and improved by audiological scientists of the Leiden University Medical Center (LUMC) and the Academic Medical Center Amsterdam (AMC), and a new version was launched nationwide in 2010 (Leensen et al., 2011b, de Laat et al., 2016). The test targets teenagers and young adults aged 12–24 years, to raise their awareness of NIHL and to screen for NIHL in an easily accessible and relevant way. The test discriminates quickly between normal hearing individuals and individuals with hearing difficulties by means of a pass/fail outcome.

This study aims to describe the intelligibility of speech in noise among teenagers and young Dutch adults aged 12–24 years using Earcheck responses, collected nationally. As about 20,000 youngsters perform the online Earcheck each year, test results will give an insight into the intelligibility of speech in noise among teenagers and young adults. It will also reveal what proportion of users have poor scores, whom potentially have an incipient hearing loss. A secondary objective is to study the relationship between performance in speech-in-noise testing and the gender and age of the respondents.

## Materials and Methods

### Test characteristics

Earcheck is a Dutch speech-in-noise test which consists of nine different monosyllabic words, randomly presented in a low-pass filtered masking noise. Subjects respond via a screen showing nine response buttons and a tenth button saying “not recognized.” This last button was added to prevent respondents from guessing. Words are presented to the subject who is asked to identify the word by clicking on the corresponding button on the computer screen. The level of the noise is fixed and the level of presented words varied using an up–down procedure with a 2 dB step size. This test procedure is based on the method developed by Plomp and Mimpen (1979a), with the exception of the fact that the first stimulus of Earcheck is presented only once at a fixed signal-to-noise ratio (SNR) of –10 dB. The SNRs in the test range from –6 dB SNR to –30 dB SNR.

A list of 27 stimulus words is used to estimate SNR at which 50% of the speech material was identified correctly. This is defined as the speech reception threshold (SRT), and is calculated by taking the arithmetic average of the SNRs of the last 20 presentations. The result is shown immediately after completing the test, and is classified into categories “good,” “insufficient,” or “poor.” When a user fails the online test (an insufficient or poor test result), diagnostic audiological evaluation by a general practitioner, hearing aid dispenser, or at an audiological center is recommended. This recommendation may also encourage users to protect their hearing by making behavioral changes or by actively seeking medical help. The Earcheck is also applied in the adult population under the name Hoorscan (www.hoorscan.nl). The Hoorscan is aimed at adults considering using hearing aids and is provided online by the hearing aid dispenser.



### Test validation

Previous research in our department indicated that a test with a stationary low-pass filtered masking noise, instead of a broadband noise, discriminated better between normal hearing and hearing-impaired subjects with different degrees of NIHL (n = 98). This resulted in a high sensitivity of 95% and a high specificity of 98%, with SRT thresholds of -18.4 dB SNR (cut-off value for the categories good and insufficient) and -12.7 dB SNR (cut-off value for the categories insufficient and poor), and without a reduction in test reliability (Leensen et al., 2011b). The SRT performance was compared to clinical pure-tone audiometry, which was considered the gold standard. This validation took place in an adult population (age range: 18–72 years), and in a laboratory setting, but currently gives the best approximation of the validity of the test. Exact values for test sensitivity and specificity for the teenagers and young adult population in a home-based test situation are yet to be determined.

Home-based application of Earcheck and Hoorscan may result in different test results due to poorer testing conditions resulting from uncontrollable parameters such as ambient background noise and the quality of the sound cards or transducers used. A previous study investigating the influence of test environment on the applicability of Earcheck showed that SRTs measured at home were poorer than those obtained in the laboratory (Leensen and Dreschler, 2013b). As a consequence, cut-off values for NIHL should be 1.2 dB SNR higher in a home-based setting than the cut-off values that were determined in a well-controlled lab setting (Leensen and Dreschler, 2013a). To account for the observed differences in SRT when completing the test at home, we applied a correction factor of 1.2 dB SNR.

### Test procedure

Earcheck is performed at an individually set presentation level. Prior to starting the test, a word is presented repeatedly without noise. Respondents use their personal computer (PC) volume control or a slider on screen to adjust the volume to a level at which the presented word is clearly intelligible. This user-selected presentation level is used for the presentation of all subsequent test stimuli. All testing is done binaurally (in diotic presentation, i.e. the same signals are presented to both ears) and either headphones or loudspeakers can be used for testing. Headphones are recommended to obtain a more reliable test result and in this study only the headphone data are included. The test can be performed in less than 5 min, including introduction and instruction, presentation of test results, and recommendations.

## Data measures

Cross-sectional data were derived from all participants of Earcheck over a 5-year period (January 2010 until December 2014). The participants represented an Internet convenience sample, that is, the study sample was not actively recruited for the purpose of this study. The test is embedded in the website of the National Hearing Foundation (www.hoorstichting.nl), alongside with other educational materials, which are all available free online. Users reached the test website in several ways, and all voluntarily performed the online hearing screening test, for example, at home or at school as part of an education program.

Earcheck collected self-reported information on age (in years), gender (male/female), self-rated hearing status (good, less, or poor), and type of transducer used (headphones or speakers). It also collected test results, including SNRs per stimuli, mean SRT scores (in dB SNR), test result category (good, insufficient, or poor), and intra-individual standard deviations (SDs) (in dB). Only test results of teenagers and young adults with reliable intra-individual SDs and tests performed by headphones were analyzed. Subjects younger than 12 years old or older than 24 years were excluded. In addition, participants with invalid intra-individual SDs of 0 dB or ≥3 dB were excluded. The intra-individual SD describes the variation within a single test measurement, and therefore is a measure for the accuracy of a test performed by an individual.



## Data analysis

Cross-sectional statistical analyses were performed using International Business Machines Corp. (IBM) Statistical Package for the Social Sciences (SPSS) version 22 (SPSS Inc., Chicago, IL, USA). Descriptive statistics were performed for the variables age, gender, self-rated hearing, and test results. In addition, relationships between the factors were assessed. At first, bivariate relationships between SRT score and gender, and SRT score and age were explored by means of simple linear regression analyses. Then, multivariable regression analyses were performed, with SRT score (in dB SNR) as primary outcome variable, significant explanatory factors, and relevant interactions. The results are presented as beta values, 95% confidence intervals (95% CIs), *P*-values, and explained variance ($R$-squared). The multivariable regression model was cross-validated in a data sample of users of Hoorscan between the age of 12 and 24 years, collected over the same period. The same exclusion criteria as for Earcheck data were applied. The same regression model was applied to the Hoorscan data sample, and the beta values were compared.

## Results

In total, 242,383 completed Earcheck tests were registered for the period January 2010 until December 2014. After excluding data according to the before-mentioned exclusion criteria, 96,803 valid test results remained for analysis. 69,647 results were excluded, as subjects were younger than 12 years or older than 24 years. 26,208 test results were excluded due to invalid intra-individual SDs of 0 dB or ≥3 dB. Finally, 49,725 results were excluded, as these tests were not performed using headphones.

Because there was no great variation in SRT scores between the years, data of all years were pooled for further analyses. Table 5.1 displays the SRT scores per year (in percentiles). To assess whether there were differences in mean SRT score between the years, a one-way Analysis of Variance (ANOVA) was performed, with SRT score as the dependent variable and year as factor. There was a significant effect for year ($F = 6.715$, $P < 0.001$). When performing a post-hoc test, only the year 2010 differed in mean SRT score from other years. However, this difference was rather small, and not relevant (mean difference of 0.15 dB).

**Table 5.1.** SRT scores (in dB SNR) in percentiles, per year.

| Year | SRT score (dB SNR) | | |
|---|---|---|---|
| | 25th percentile | 50th percentile | 75th percentile |
| 2010 | -20.8 | -19.4 | -17.4 |
| 2011 | -20.8 | -19.4 | -17.0 |
| 2012 | -20.7 | -19.4 | -17.2 |
| 2013 | -20.7 | -19.3 | -17.2 |
| 2014 | -20.7 | -19.3 | -17.2 |

Overall, the proportion of male users was slightly smaller than the proportion female users (48 and 52%, respectively). The mean age of the users was 15.7 years (SD = 2.8). The majority of all users (76%) rated their own hearing as good, 23% as insufficient, and only 1% as poor. The mean SRT score of all users was –18.3 dB SNR (SD= 3.7). The largest proportion of users (74.5%) had a good result, while 18.5% had an insufficient result and 7.0% a poor result. Results from the simple linear regression analyses for the bivariate relationships are shown in Table 5.2.

**Table 5.2.** Bivariate relationships with age and gender.

| | Outcome: SRT score (dB SNR) | | |
|---|---|---|---|
| | **ß-Value** | **95% CI** | ***P*-value** |
| Age (years)* | –0.28 | –0.29 to –0.27 | <0.001 |
| Gender** | 0.32 | 0.27 to 0.37 | <0.001 |

**Reference category: 12-year-old, **Reference category: male.*

The factors age and gender significantly explained variation in SRT outcome. Mean SRT scores decreased with increasing age. Overall, male users had slightly better scores compared to female users. Then, a multivariable regression analysis was performed, including the outcome factor SRT and explanatory factors age and gender. The model included an interaction term for age and gender, resulting in the following formula:

5

$$SRT = intercept + age*b1 + gender*b2 + age*gender*b3$$

Results are presented in Table 5.3. The main factors age and gender were both significantly related to SRT score. SRT score decreased (improved) with 0.31 dB SNR per age-year. To illustrate, there was a 3.6 dB SNR difference in SRT performance between a 12-year-old and a young adult male user aged 24 years. Female users had a slightly better score as compared to male users. For the reference category of 12-year olds, this difference was 0.13 dB SNR. However, the interaction term between age and gender was significant, indicating a different relation between SRT score and age for male and female users. According to the model, SRT score improved by age for both male and female users. The mean SRT score for male users was more favorable than for female users from the age of 15 years and above.

**Table 5.3:** Multivariable regression results.

| | Outcome: SRT score (dB SNR) | | |
|---|---|---|---|
| | ***ß*-Value** | **95% CI** | ***P*-value** |
| Intercept* | –17.21 | –17.26 to –17.15 | <0.001 |
| Age (years) | –0.31 | –0.32 to –0.30 | <0.001 |
| Gender | –0.13 | –0.21 to –0.05 | 0.001 |
| Age*gender | 0.07 | 0.05 to 0.09 | <0.001 |

**Reference category: 12-year-old male user, R2 = 0.042.*

A graphical presentation of the relationship between age and SRT score, for both male and female users, is given in Figure 5.1. SRT scores tend to decrease much more sharply for teenagers (12–18 years) compared to young adults (≥18 years), for both male and female users. This consistent decrease in SRT scores with age is displayed in more detail in Figure 5.2, by means of percentiles. A Mann–Whitney *U* test showed that the SRTs of teenagers were significantly worse than those of young adults ($P < 0.001$), with a difference in mean SRT score of 1.4 dB SNR. For this reason, multiple regression analyses were repeated for teenagers and young adults separately.

Both models included the outcome factor SRT score, explanatory factors age and gender, and an interaction term for age and gender. Results are presented in Table 5.4. For the teenagers, the main factors age and gender were both significantly related to SRT score. SRT score decreased (improved) with almost half a dB SNR per age-year. Female users had a slightly lower score compared to male users (–0.32 dB SNR). However, the interaction term between age and gender was significant. SRT score improved with age for both male and female users, but the difference between SRT scores of male and female users became greater with age with a more favorable mean SRT score for male users. For the young adults, the age-effect was significant but quite small, with a change of –0.09 dB SNR per age-year. The mean SRT score of female users was about half a dB poorer than those of male users. There was no significant interaction between age and gender for the young adults.

**Table 5.4.** Multiple regression results for teenagers (*n* = 76,070) and young adults (*n* = 20,733).

| | Outcome: SRT score (dB SNR) | | |
|---|---|---|---|
| | ***ß*-Value** | **95% CI** | ***P*-value** |
| Teenagers | | | |
| Intercept* | –16.76 | –16.84 to –16.68 | <0.001 |
| Age (years) | –0.49 | –0.52 to –0.47 | <0.001 |
| Gender | –0.32 | –0.43 to –0.22 | <0.001 |
| Age*gender | 0.14 | 0.10 to 0.17 | <0.001 |
| Young adults | | | |
| Intercept** | –19.45 | –19.53 to –19.38 | <0.001 |
| Age (years) | –0.09 | –0.12 to –0.07 | <0.001 |
| Gender | 0.58 | 0.47 to 0.70 | <0.001 |
| Age*gender | 0.00 | –0.04 to 0.04 | 0.883 |

**Reference category: 12-year-old male user, **Reference category: 18-year-old male user, $R^2$ model teenagers = 0.028, R2 model young adults = 0.015.*

**Table 5.5.** Multiple regression results for teenagers (*n* = 10,555) and young adults (*n* = 6557): Hoorscan data.

| | Outcome: SRT score (dB SNR) | | |
|---|---|---|---|
| | ***ß*-Value** | **95% CI** | ***P*-value** |
| Teenagers | | | |
| Intercept* | –17.92 | –18.11 to –17.74 | <0.001 |
| Age (years) | –0.30 | –0.36 to –0.24 | <0.001 |
| Gender | –0.01 | –0.27 to 0.25 | 0.963 |
| Age*gender | 0.16 | 0.08 to 0.24 | <0.001 |
| Young adults | | | |
| Intercept** | –19.85 | –19.99 to –19.71 | <0.001 |
| Age (years) | –0.04 | –0.08 to 0.00 | 0.060 |
| Gender | 0.59 | 0.36 to 0.81 | <0.001 |
| Age*gender | 0.05 | –0.02 to 0.12 | 0.165 |

**Reference category: 12-year-old male user, **Reference category: 18-year-old male user, $R^2$ model teenagers = 0.015, R2 model young adults = 0.016.*

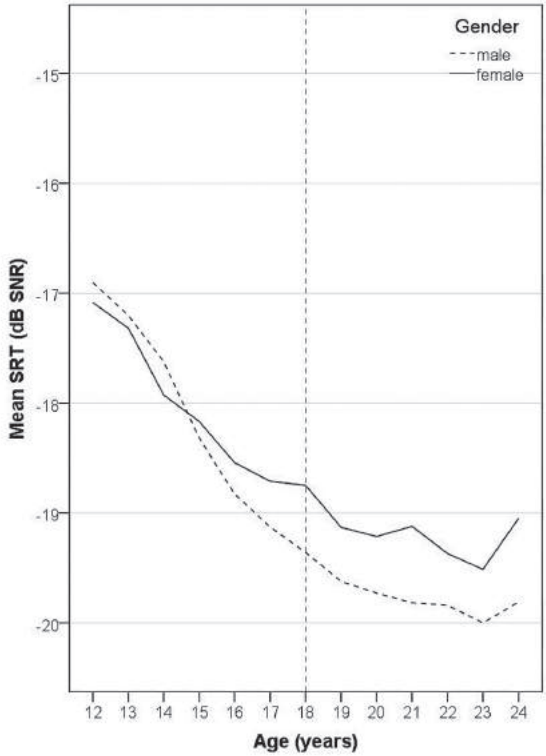


**Figure 5.1.** Average SRT score as a function of age, for male (black line) and female (interrupted lines)

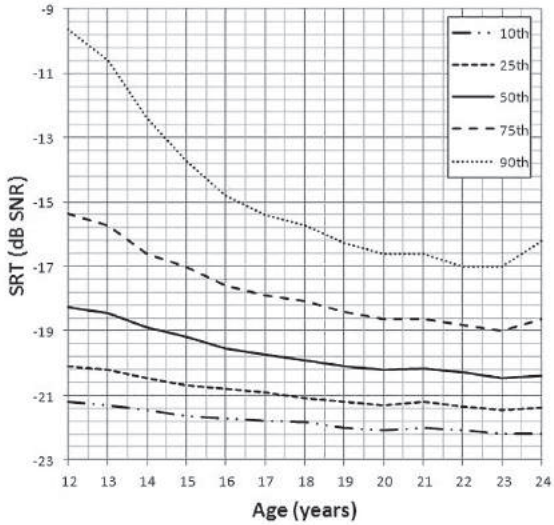


**Figure 5.2:** SRT score by age, in percentiles.

The models for teenagers and young adults were cross-validated in a Hoorscan sample using the same inclusion criteria [Table 5.5]. These analyses showed models similar to those obtained with Earcheck data, except that the main effect of gender was not significant for the teenagers. For the teenagers in the Hoorscan sample, the main effect of age was significant, and the beta-value was somewhat smaller than for the teenagers in the Earcheck sample. There was a similar interaction effect between age and gender. For the young adults in the Hoorscan sample, there was no significant relationship between age and SRT score. The effect of gender was very similar to the effect of gender in the Earcheck sample. The interaction model was not significant in this sample either.

## Discussion

The improved and validated online speech-in-noise screening test Earcheck generated large quantities of national data. In total, 242,383 completed tests were registered for the period January 2010 until December 2014. We analyzed 96,803 valid test results for teenagers and young adults (between 12 and 24 years) with reliable intra-individual SDs (≤3 dB) and usage of headphones. The proportion of users with a good result was 74.5%, while 18.5% had an insufficient result and 7.0% a poor result. This implies that a substantial number of users of Earcheck may have some difficulty in understanding speech in noise. The cause of these difficulties is not known, but NIHL is one of the potential causes. According to the final multiple regression models that were fitted for teenagers and young adults separately, SRT score tends to improve with age, especially among teenagers between the age of 12 and 18 years. Furthermore, the effect of age appears to be somewhat different for male and female users. Similar effects of age and gender for teenagers and young adults were found in an independent sample of online users of the same test (Hoorscan), which means that the fitted Earcheck model can be generalized to a different internet convenience sample. An important note, however, is that both samples do not accurately represent the real population, so it is not possible to draw conclusions for all Dutch teenagers and young adults based on this research.



Results of online-speech-in-noise tests performed at home or at a remote setting have not been studied earlier in this specific age group. Studies mainly concern online SRT performance of (older) adults, and focus on whether ageing reduces speech intelligibility (Smits and Houtgast, 2005, Stam et al., 2015). Other studies mainly focus on age-specific normative data of school-aged children (Hnath-Chisolm et al., 1998, Eisenberg et al., 2000, Fallon et al., 2000, Johnson, 2000, Stuart, 2005, Talarico et al., 2007, Vaillancourt et al., 2008, Jansen, 2013, Koopmans and Smits, 2015). According to these studies, there are age-related improvements in speech in noise recognition among young children (5–12 years); however, no unambiguous age-effect was found. Differences in findings may depend on the type of task, the speech and background noise material used in the task, the study sample (the size of the sample and the number and range of age groups), and the age of the adult reference group (Elliot, 1979). The majority of the studies support the statement that auditory maturation is more or less completed by the time children reach adolescence, with a speech-in-noise performance equal to adults' performance. Our study shows that the SRT performance improves even after the age of 12 years. It suggests that the maturation process of speech-in-noise performance in this type of speech-in-noise test is not complete until the age of 18 years.

In this sample, 25.5% of respondents failed the test, of which 18.5% of test results were categorized as insufficient and 7.0% as poor. This percentage is higher than previously reported results from population-based surveys. A national cross-sectional survey found prevalence rates of hearing loss in the young U.S. population ranging from 12.5 to 19.5% (Niskar et al., 2001a, Shargorodsky et al., 2010). According to another survey, 16% of the young U.S. adults entering an industrial workforce showed high frequency hearing loss (Rabinowitz et al., 2006b). It is important to note that these studies used screening audiometry to assess hearing loss, and maintained a very strict criterion of NIHL, defined as hearing thresholds greater than 15 dB hearing level (HL) in either ear at 3, 4, or 6 kHz. Moreover, there were important issues in these audiometric surveys that question the accuracy of the prevalence estimates, for example, the imprecision of screening audiometry and the unknown influence of other otological problems, such as conductive hearing losses (Schlauch and Carney, 2012). Although pure-tone audiometry is the reference standard for assessing hearing threshold levels, it is subject to variability due to calibration issues, test–retest reliability, and test environment. These factors, which are particularly present in screening settings, are critical when determining minimal deteriorations in the lowest signal level a person can hear.

An important strength of this study is that Earcheck, a functional supra-threshold test that measures speech intelligibility in noise, was used to assess hearing loss. Because of its test characteristics, the main limitations that are linked to pure-tone audiometry could be avoided. By measuring a SNR, the influence of testing conditions is minimized (Culling et al., 2005). In addition, speech-in-noise tests are insensitive to conductive hearing losses. Therefore, Earcheck results might yield more accurate prevalence estimates than previously reported. Another strength is that the test is convenient and easily accessible. The self-administered online speech-in-noise test made it possible for teenagers to measure SRT performance in the comfort of their own home or school. The test could be performed free of charge, and the online applicability resulted in a large quantity of data, collected nationally and in an interesting age group of 12- to 24-year olds. Finally, the estimated models were cross-validated in a separate sample of similar data. This strengthened the reliability of the relationships among age, gender, and SRT observed in the Earcheck data.



Despite the above-mentioned strengths of Earcheck, the percentage of subjects failing the test is high. The most important explanation for this high percentage of respondents with non-normal performance is the use of a convenience sample of Dutch adolescents and young adults who performed the test voluntarily and on their own initiative. This sort of sampling is usually biased by selection. It most likely does not fit the definition of a random sample, where everyone in the population has an equal chance of being selected for participation. Since it does not truly represent the population, the study is limited when it comes to generalization. Although a large proportion of the study sample performed Earcheck in their school class as part of an educational program (about 40%), a higher response rate for subjects that have doubts about their hearing is expected in this study. This selection bias probably resulted in an overestimation of hearing losses in this population. However, we do not expect that the likely selection bias affected our findings concerning the influence of age and gender on SRTs results.

Another explanation for the high percentage of poor results is that the tests were performed in uncontrolled home settings by anonymous users; hence, the results were based on self-testing. Although we tried to limit the influence of inaccurate self-testing by using a criterion for reliable intra-individual SD, and including only tests performed by headphones, unreliable self-tests cannot be completely ruled out. In addition, test characteristics of home-based application using a sample of young adults are unknown, but it is likely that the values for sensitivity and specificity are poorer than in laboratory testing (Leensen and Dreschler, 2013b). Finally, fixed, age-independent, cut-off values for the result categories were used in this study. Although the influence of home-based testing was accounted for, the effects of age were not. As an improvement in speech intelligibility with age was proven, the use of fixed cut-off values may have overestimated the prevalence of poorer test results.

To gain a better insight into speech discrimination abilities of teenagers using the online speech-in-noise test Earcheck, it is important to study normal age-specific SRT performance of this target group in a controlled study. Earcheck was comprehensively validated among adults in earlier studies; however, it is important to study the test in normal-hearing teenagers as well, to set appropriate reference values and to correctly interpret internet screening outcomes. In addition, further work investigating the maturation effect of speech understanding among normal-hearing teenage students aged 12–18 years is needed.

## Conclusion

The goal of this study was to investigate intelligibility of speech in noise among teenagers and young adults in the Netherlands, using Earcheck responses. Earcheck is a Dutch online speech-in-noise test, specifically designed to detect high-frequency hearing loss. The majority of the respondents scored “good”; however, an “insufficient” or “poor” test result was obtained by part of the respondents, indicating that hearing loss may be present in this population of teenagers and young adults. The percentage of respondents with a poor result was higher than previously reported results from population-based survey studies. It is important to note that these findings are only applicable to the convenience sample used in this study and cannot be generalized to the general population due to the significant likelihood of selection bias. This research also gave insight into the relationship among SRT score, gender, and age. The results of this study show a significant effect of gender and age on

SRT performance. SRT scores tended to improve with age, especially among teenagers between the age of 12 and 18 years, and this effect was greater in male than in female respondents. The results of this study suggest that the maturation process of Earcheck performance is not complete until adulthood and suggest the use of age-dependent reference values for Earcheck; however, more research is required.

# Chapter 6

# Age dependence of thresholds for speech in noise in normal-hearing adolescents.

I. Jacobi[1]
M. Sheikh Rashid[1]
J.A.P.M. de Laat[2]
W.A. Dreschler[1]

1 Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands
[2] Department of Audiology, Leiden University Medical Center, Leiden, The Netherlands

Trends in Hearing,2017, 21, 1-9.

## Abstract

Previously found effects of age on thresholds for speech reception thresholds in noise in adolescents as measured by an online screening survey require further study in a well-controlled teenage sample. Speech reception thresholds (SRT) of 72 normal-hearing adolescent students were analysed by means of the online speech-in-noise screening tool Earcheck (In Dutch: Oorcheck). Screening was performed at school and included pure-tone audiometry to ensure normal-hearing thresholds. The students' ages ranged from 12 to 17 years. A group of young adults was included as a control group. Data were controlled for effects of gender and level of education. SRT scores within the controlled teenage sample revealed an effect of age on the order of an improvement of -0.2 dB per year. Effects of level of education and gender were not significant. Hearing screening tools that are based on SRT for speech in noise should control for an effect of age when assessing adolescents. Based on the present data, a correction factor of -0.2 dB per year between the ages of 12 and 17 is proposed. The proposed age-corrected SRT cut-off scores need to be evaluated in a larger sample including hearing-impaired adolescents.

## Introduction

Reduction in speech intelligibility performance in background noise is an early indicator of hearing impairment. Because expert assessment of hearing impairment and classical pure-tone audiometry have their drawbacks in large-scale studies, time-efficient and easily accessible self-assessment screening tests, either by telephone or online, have been developed. They focus primarily on mid- to high-frequency hearing loss and the intelligibility of words in stationary masking noise (Smits and Houtgast, 2005, Leensen et al., 2011b, Jansen et al., 2013, Dillon et al., 2016). These internet-based speech-in-noise tests provide the opportunity to reach a large population and have proven to be reliable as self-administered hearing screening tools.

The Dutch Earcheck (Oorcheck, www.oorcheck.nl) is such an online hearing screening test and was developed by the Leiden University Medical Center (LUMC) and the Academic Medical Center Amsterdam (AMC) (Leensen et al., 2011b). It is specifically aimed at young people to raise awareness of the consequences of uncontrolled noise or music exposure, and reaches about 30,000 to 40,000 participants a year.

The speech-in-noise test uses nine monosyllabic words that are randomly presented in a fixed masking noise, while the signal level is varied in 2 dB steps to assess the speech reception threshold (SRT). In 2015, Oorcheck data comprising the five preceding years were analysed. The test results of 96,803 Oorcheck users aged 12 to 24 years revealed a trend in SRT scores, improving by about 0.3 dB signal-to-noise ratio (SNR), especially between 12 and 18 years of age (Rashid et al., 2016).



While younger children and adults are age-groups that are regularly investigated in hearing research, there are fewer studies on (changes during) adolescence. Research on adolescents is made difficult by compulsory school attendance and class schedules, the consent process, and inter-subject variation during this period of biological and psychosocial change.

There is well-documented evidence that normal-hearing (NH) children aged 5 to 12 years differ from adults in speech recognition performance (Neuman et al., 2010, Hall et al., 2016). The few studies on speech recognition during adolescence differ in stimuli and masker conditions. Nonetheless, they all indicate a steady improvement in the SRT from childhood to adulthood. In an early study, developmental changes were found between 3 and 17 years of age (Elliot, 1979)(Goldman & Fristoe & Woodcock as cited in Elliott et al., (1979)). In 2005, the data of a Dutch telephone survey revealed worse SRTs in 15- to 19-year-olds compared with 20- to 24-year-olds (Smits and Houtgast, 2005). In a more recent study by Corbin et al. (2016), the recognition of monosyllabic words in a speech-shaped noise masker was worse in 8- to 12-year-olds compared with 13- to 16-year-olds and adults. Wightman and co-workers (Wightman and Kistler, 2005, Wightman et al., 2006, Wightman et al., 2010), assessed subjects aged 5 to 18 years of age and found that the rate of change with age to be slower in ipsilateral masking with a single talker than in contralateral masking with a single talker, suggesting that informational masking in the two conditions is mediated by different processes.

The changes across adolescence may be explained by the well-characterized changes in brain structure and functioning during that period (Gogtay et al., 2004, Litovsky, 2015, Vinette and Bray, 2015). During adolescence, many brain regions are still in development (Gogtay et al., 2004, Vinette and Bray, 2015), and auditory processing in the brainstem and cortex matures (Ponton et al., 2000, Wunderlich and Cone-Wesson, 2006, Mahajan and McArthur, 2012, Skoe et al., 2013, Skoe et al., 2015, Krizman et al., 2016). Studies that correlate auditory brainstem responses with speech perception reveal a highly complex dynamic auditory system with sound representations that undergo changes during adolescence, with large effects of enriched or limited experience on auditory functioning and the subcortical system and continuous fine-tuning (de Boer and Thornton, 2008, Strait et al., 2014, Krizman et al., 2015, Tierney et al., 2015).

Cognitive control of speech perception improves from childhood to adulthood. There are changes in the effects of attention on auditory stream segregation, and there is an increase in the precision of acoustic-phonetic properties and boundaries (Sussman, 1993, Wunderlich and Cone-Wesson, 2006, McNealy et al., 2010, Medina et al., 2010, Westerhausen et al., 2015). Sensory processing is refined significantly by cognitive skills (Kraus et al., 2012, Strait et al., 2014) and a comparison of auditory-evoked potentials in neurobiology studies shows that changes in speech perception processing and structural changes develop concurrently (Eggermont and Ponton, 2003). Less well investigated is the effect of experience on the maturation of the adolescent neurodevelopment (Tierney et al., 2015). Although the age effect seen in the 5 years of Oorcheck data shows some correspondence with effects of maturation in other fields of research, the internet survey carries some bias. This includes inclusion bias, a poorly controlled test condition, unknown hearing thresholds, and uncertainties with reference to the participants' age specifications. Additional research is required to confirm the age-related findings from the online Oorcheck survey.

The primary aim of the present study was therefore to analyse to what extent an age-related trend can be found in the speech-in-noise test data of adolescents in a well-controlled sample. Oorcheck SRT data for NH adolescents were collected at two high schools, after which the effects of age, level of education, gender, and test repetition were analysed. Adolescents were compared with a control group of young adults. A secondary aim was to estimate correction factors to compensate for the potential unwanted effect of age in the online screening tool.

## Materials and Methods



### Subjects

A total of 104 subjects were assessed in February 2016. Recruitment of the adolescents took place at two schools in the Netherlands: Zandvliet College, a higher secondary school in The Hague; and Haarlem College, a lower secondary school in Haarlem. With the consent of the school management and parents, students were sent information about the purpose and procedure of the study. Inclusion criteria required participants to be native Dutch speakers aged 12 to 17 years with NH. NH was specified as hearing thresholds of 25 dB Hearing Level (HL) or better at 250 and 500 Hz, and hearing thresholds of 20 dB HL or better at 1, 2, 3, 4, and 6 kHz in each ear. The threshold of 25 dB HL at 250 and 500 Hz was chosen to account for potential environmental noise, since no sound proof booths were available. The participants' (intra-individual) standard deviation for the Oorcheck had to be lower than 3 dB (compare (Rashid et al., 2016)). A control group of young adult college students aged 18-20 years were recruited from the Avans Hogeschool in Breda.

### Procedure

This cross-sectional study protocol was approved by the medical ethics committee of the University of Amsterdam (identification code 2015_297). Data on the subjects' SRT in noise were obtained by their responses during the online hearing test Oorcheck, completed in a quiet room at school. Prior to testing, information on the subject's age, grade, and gender was collected. To confirm NH, the Oorcheck was preceded by pure-tone audiometry in the same quiet room.

### Pure-tone audiometry

Pure-tone audiometry was performed by two trained test operators and included air conduction thresholds at frequencies 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz, using calibrated clinical audiometers (AC40 and Decos audioNigma), connected to TDH 39 headphones with sound attenuating cups (Amplivox audiocups).

### Ambient noise level measurements

Ambient noise-level measurements were performed in all test rooms. Using a DVM805 digital sound level meter (applicable standard: IEC651 type 2), the sound level in the room where the adult control subjects were measured was 37 dBA, and the sound level in both rooms at Zandvliet college was 35 dBA. The sound level measurements in the two rooms used at Haarlem College were done using a sound level meter B&K 2260. The Z-weighted maximum sound levels for the mid-frequency third-octave bands (250-8000 Hz) ranged from 23.3 to 38.7 and 18.6 to 37.4 dB SPL, respectively. The audiometric test conditions at all test locations met the requirements of the international standards for hearing screening with sound attenuating cups in combination with headphones (i.e. unmasked air conduction starting at 500 Hz; ISO 8253, Part I).

### Earcheck

After pure-tone audiometry, the subject's SRT was assessed with the Oorcheck tool. The speech material used in this speech-in-noise test is based on a closed set of nine monosyllabic words (thumb [dœym], goat [xεIt], chicken [kIp], lion [lew], cat [pus], rat [Rat], fire [vyr], wheel [wil], and saw [zax]) taken from the Dutch word lists for speech audiometry (Bosman, 1989), spoken by a Dutch female speech therapist. These stimuli were presented in random order with a stationary masking noise that was low-pass filtered with a cut-off frequency of 1600 Hz and a slope of 100 dB per octave. The original broadband masking noise and the speech stimuli had a matching long-term average spectrum. A more detailed description of the test material can be found in Leensen (2013). Testing is binaural and diotic. Prior to testing, a stimulus without noise is presented and the subject is instructed to adjust the volume to a comfortable level at which the stimuli can be clearly understood. Starting at a signal-to-noise ratio of -10 dB, the level of noise is fixed while the signal level is varied adaptively in 2 dB steps according to the up-down procedure described by Plomp and Mimpen (1979a). After each word presentation, the subject has to choose one of nine corresponding pictograms on the screen or the button "not understood." SRT is defined as the SNR at which 50% of the word material is identified correctly. In the Oorcheck, SRT is calculated as the average SNR for stimuli 8 to 27 and is stored in an online database. The result of the Oorcheck is either "pass" or "fail" (using a cut-off value of -18.4 dB SNR established during controlled experimental settings; (Leensen et al., 2011b)) and is directly shown to the participant. The participants performed the online test individually twice (test and immediate retest) with minimal instructions from the researchers.

For the Oorcheck, a research laptop (HP) and a tablet (Surface) were used, in combination with Sennheiser HDA 200 and Sennheiser HD330 headphones. The control group performed the test using a Sennheiser HD330 headphone on their own mobile phone. Previous research on the Oorcheck presentation levels on the SRT's of NH subjects revealed no significant effects at presentation levels well above the absolute threshold, ranging from 65 to 77 dBA (Leensen and Dreschler, 2013b).



## Statistical analyses

Descriptive statistics were applied on hearing thresholds, age, gender, level of education, and on the SRT results (test and retest) derived with the Oorcheck. To explore SRT scores as a function of age (in years), gender (male or female), and education level (low or high), multiple regression analyses were performed. To explore the correction factor for age, regression analysis was applied to the SRT scores. All data were analysed using IBM SPSS Statistics 23.

# Results

## Subjects

One hundred and four subjects participated, but 23 of them (22%) did not fulfil the audiometric inclusion criteria: Seven subjects (7%) had a hearing loss in both ears, seven had a hearing loss in the right ear, and nine (9%) in the left ear. One of the remaining 81 (78%) NH subjects did not perform a retest of the Oorcheck and was therefore also excluded from the analysis. Of the remaining 80 NH subjects with two Oorcheck runs, eight (8%) had an intra-test standard deviation greater than 3 dB for the Oorcheck and were therefore excluded as well (cf. (Rashid et al., 2016), leaving 72 subjects (69%) for data analysis. Table 6.1 shows the mean hearing thresholds of pure-tone audiometry (PTA) 0.5-1-2 kHz and PTA 1-2-4 kHz of the 72 subjects for each age- group and by ear, as well as the number of subjects per group. The group of young adults (control group) consisted of 10 participants with a mean age of 19 years (SD= 0.94). The adolescent group consisted of 41 female and 21 male students ranging from 12 to 17 years of age. Thirty-two of them were students at the higher level secondary school, while 30 attended a lower level secondary school.

## Test and retest SRT

The mean SRT scores for the first test and the retest are presented in Table 6.2 and Figure 6.1, grouped by age and including the young adults. The SRT scores of the first test (M=-19.2 dB SNR, SD=1.4) and the retest (M=-19.7 dB SNR, SD=1.5) differ significantly (t=2.421, $F(1,71)=5.86$, and $p=0.018$, paired t-test, two-tailed) and indicate a learning effect.

Repeated measures analysis on the test and retest SRT with age-group as a covariate showed a significant main effect of age-group, $F(1,71)=8.508$, $p=0.005$. The effect of the test versus the retest situation showed a trend of improvement but did not reach significance, $F(1,70)=3.331$, $p=0.072$. There was no significant interaction effect of age-group with the test–retest SRTs, $F(1,70)=2.334$, $p=0.131$.

## Mean SRT and effects of age, gender, and level of education

A regression analysis was conducted on the mean SRT results of test and retest. The mean SRT improved significantly by 0.2 dB with each year of age (95% CI:

-0.364; -0.072, p=0.004), while effects of gender (95% CI: -0.613; 1.323, p=0.467) and level of education (95% CI: -0.660; 1.391, p=0.479) were not significant. Post hoc tests between the age-groups (corrected a<.008) revealed a significant difference between the 12-year-olds, (mean SNR= -18.7 dB, SD=0.9) and the adults, (mean SNR=-20.3 dB, SD= 0.9, t(18)= 3.75, p=0.001); all other comparisons were not significant.

## Age-based cut-off score

Initially, the cut-off score for a good Oorcheck SRT in a controlled laboratory setting was -18.4 dB SNR. According to the results of the regression analysis, the effect of age can be predicted by a factor of -0.2 dB per year of age for the present study group. A correction of -0.2 dB per year was applied to the original overall cut-off score to control for age, which resulted in a cut-off value of -17.2 dB SNR for the 12-year-olds, that decreased with age up to the initial cut-off value of -18.4 dB SNR for young adults of 18 years and older. In Table 6.3, the proposed new cut-off scores based on an SRT improvement of -0.2 dB per year of age are presented by age-group, next to the SRT distributions in percentiles. As can be seen, the 75th or 90th percentiles of the subjects tested are at lower scores than the cut-off values (except for the 15-year-olds), and the vast majority of the SRTs in all age-groups are within the proposed cut-off value in all age-groups.



## Categorized test results and corrections for age and rapid learning

Table 4 shows the number (%) of subjects who scored “good” versus “poor” in their first and second Oorcheck, respectively, based on the cut-off score of -18.4 dB (Leensen et al., 2011b). Seventy-two percent of the NH subjects scored a “good” Oorcheck in the first test and 93% scored a “good” in the second Oorcheck. In the lower part of Table 6.4, the test and retest Oorcheck results are given according to the new categorization with a correction factor of -0.2 dB per year of age. The present study showed a trend of improvement in the repeated measures analysis. For subjects whose first age-corrected score was “poor,” the retest score was used. Figure 6.2 shows the statistical distribution (in percentiles) of the deviations from the age-corrected cut-off scores after the application of both improvements, that is, after the application of age-corrected cut-off scores and with a replacement of the first test score by the retest score in those cases where the first score was categorized as “poor.”

Ninety percent scored a “good” first test score, and in 10% of the cases, the retest score replaced a “poor” first test score. One of the 72 subjects failed to reach the age-corrected Oorcheck criterion both in test and in retest. As can be seen in Figure 2 and Table 4, with this procedure, 99% of the NH subjects obtained a good score.

**Table 6.1.** Age group, number of participants, and mean (SD) of PTA0.5/1/2 and PTA1/2/4 in dB HL for right ear, left ear, and better ear.

| | | PTA right ear | | PTA left ear | | PTA better ear | |
|---|---|---|---|---|---|---|---|
| **Age group** | **N** | **$PTA_{0.5/1/2}$** | **$PTA_{1/2/4}$** | **$PTA_{0.5/1/2}$** | **$PTA_{1/2/4}$** | **$PTA_{0.5/1/2}$** | **$PTA_{1/2/4}$** |
| 12 | 10 | 4.50 (4.38) | 4.33 (4.81) | 3.50 (4.26) | 3.25 (3.86) | 2.33 (4.32) | 2.25 (4.23) |
| 13 | 8 | 7.50 (5.84) | 6.35 (4.45) | 5.83 (4.63) | 5.00 (3.65) | 5.21 (4.67) | 4.17 (4.15) |
| 14 | 11 | 3.93 (4.43) | 1.67 (5.69) | 3.33 (6.67) | 1.74 (6.41) | 2.12 (5.63) | 0.53 (5.76) |
| 15 | 17 | 3.23 (4.77) | 1.47 (4.81) | 4.12 (4.53) | 2.89 (4.53) | 2.16 (3.67) | 0.64 (4.64) |
| 16 | 8 | 4.79 (3.14) | 2.70 (3.75) | 3.33 (5.56) | 2.92 (3.42) | 2.08 (4.78) | 1.15 (2.67) |
| 17 | 8 | 5.83 (3.14) | 1.46 (2.12) | 6.04 (4.79) | 2.71 (2.98) | 4.38 (3.56) | 1.25 (2.36) |
| Adults | 10 | 5.83 (5.62) | 4.66 (4.11) | 3.67 (5.82) | 2.83 (4.81) | 3.50 (5.90) | 1.63 (4.33) |

**Table 6.2.** Mean SRT (SD) of test and retest and the average of test and retest in dB SNR by age group.

| | Mean SRT (SD) in dB SNR | | |
|---|---|---|---|
| **Age group** | **First test** | **Retest** | **Mean of test-retest** |
| 12 | -17.92 (1.53) | -19.61 (1.11) | -18.76 (0.94) |
| 13 | -19.10 (0.89) | -19.58 (0.95) | -19.34 (0.75) |
| 14 | -19.46 (1.09) | -19.59 (1.13) | -19.52 (0.90) |
| 15 | -19.21 (1.69) | -19.55 (1.80) | -19.37 (1.42) |
| 16 | -19.17 (0.94) | -19.44 (2.00) | -19.31 (1.36) |
| 17 | -19.69 (0.73) | -19.63 (2.01) | -19.66 (1.19) |
| Adults | -20.05 (1.34) | -20.57 (0.96) | -20.31 (0.90) |
| Total | -19.22 (1.38) | -19.70 (1.48) | -19.46 (1.16) |

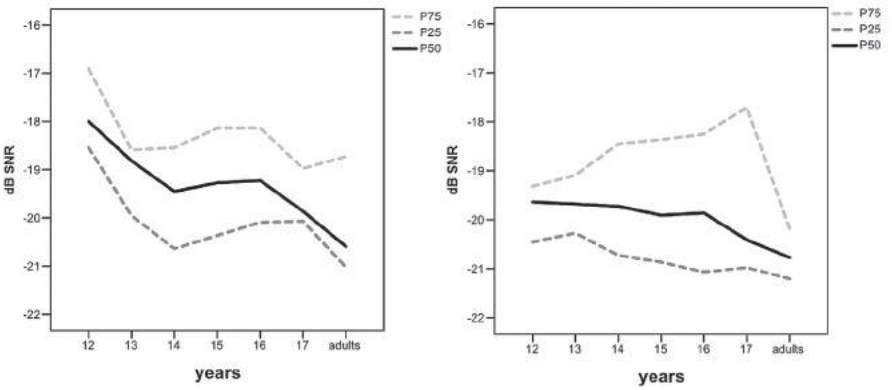


**Figure 6.1.** SRT in dB SNR (y-axis) by age-group (x-axis): Distribution in percentiles (25th, 50th, and 75th) of first test (left) and retest (right).

**Table 6.3.** SRT distribution of mean of first and second Oorcheck in percentiles by age group. To the right, the proposed Oorcheck cut-off SRTs are given based on a correction factor of 0.2 dB per year of age. N=72

| | Percentiles | | | | | Age-corrected |
|---|---|---|---|---|---|---|
| **Age group** | **10th** | **25th** | **50th** | **75th** | **90th** | **cut-off score** |
| **12** | -20.59 | -19.33 | -18.68 | -18.03 | -17.51 | -17.2 |
| **13** | -20.45 | -20.14 | -19.07 | -18.88 | | -17.4 |
| **14** | -20.63 | -20.14 | -19.64 | -19.05 | -17.88 | -17.6 |
| **15** | -21.41 | -20.48 | -19.41 | -18.84 | -16.75 | -17.8 |
| **16** | -20.86 | -20.23 | -19.84 | -18.23 | | -18.0 |
| **17** | -20.86 | -20.56 | -20.07 | -18.72 | | -18.2 |
| **Adults** | -21.57 | -21.02 | -20.32 | -19.47 | -18.98 | -18.4 |

**Table 6.4.** Number (%) of subjects who scored good (poor) on their first (second) Oorcheck, according to a cut-off value -18.4 dB SNR, and according to the age-dependent criteria (N=72).

| | | | **First test good** | **First test poor** | **Total** |
|---|---|---|---|---|---|
| **Cut-off** | | | | | |
| **-18.4 dB SNR** | **Retest** | **good** | 45 (62%) | 15 (21%) | 60 (83%) |
| | | **poor** | 7 (10%) | 5 (7%) | 12 (17%) |
| | **Total** | | 52 (72%) | 20 (28%) | 72 (100%) |
| **Age-dependent** | **Retest** | **good** | 58 (81%) | 6 (8%) | 64 (89%) |
| | | **poor** | 7 (10%) | 1 (1%) | 8 (11%) |
| | **Total** | | 65 (90%) | 7 (10%) | 72 (100%) |

6

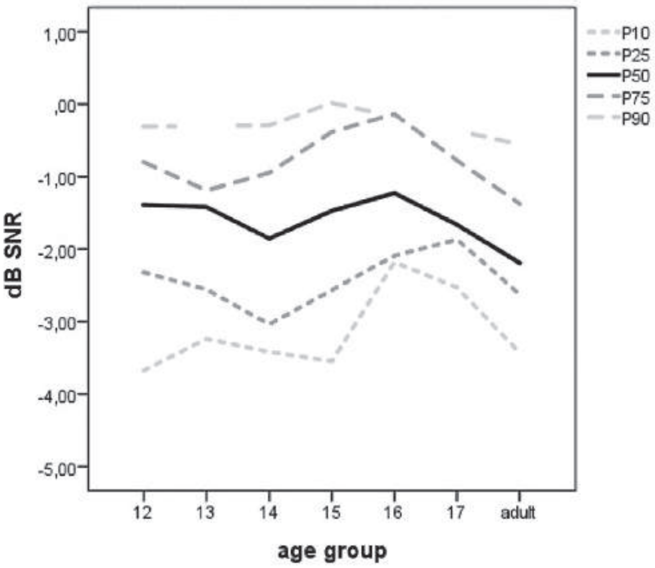


**Figure 6.2.** Deviations from the age-corrected cut-off scores in dB SNR (y-axis) by age-group (x-axis): Distribution in percentiles (10th, 25th, 50th, 75th, and 90th) after the application of the Oorcheck procedure, that is, after "poor" first test scores were replaced by retest scores.

## Discussion

The SRT results of the present sample of students with NH obtained by the Oorcheck revealed an effect of age. This supports the recent findings of a large survey that covered five years of (uncontrolled) online Oorcheck data (Rashid et al., 2016), which showed an effect of 0.31 dB SNR per age-year in the SRT score of 12- to 24-year-old males, and slightly better scores for females.

Furthermore, repeated testing revealed a small but consistent learning effect ($p=0.072$) in our study. The SRT outcome of the 17-year-olds was comparable to the adults' outcome. The effect of age on the present Oorcheck SRTs was independent of level of education and gender. However, given the small subgroups, the latter should be interpreted with caution.

The improvement in the SRTs from 12 years to adulthood was approximated by a regression analysis, and the effect of age could be corrected for by an age-dependent cut-off score for the pass or fail criteria of 17.2 dB SNR for 12-year-olds, with a decrease of 0.2 dB per year of age to a cut-off score of -18.4 dB SNR for 18 years and older. The resulting SRT of -18.4 dB SNR for adults coincides with the outcome of previous Oorcheck studies (Leensen et al., 2011b, Rashid et al., 2016). The correction of -0.2 dB per year of age seems to be a valid approximation and compensates for the unwanted effect of maturation on the Oorcheck outcome in NH adolescents. It is comparable to the results of school-age children who performed a similar Dutch online speech-in-noise test (Sheikh Rashid et al., 2017a).

The present results also correspond with the results of Wightman and coworkers (Wightman and Kistler, 2005, Wightman et al., 2006, Wightman et al., 2010), that show monotonic improvement with age for subjects between 10 and 20 years of age in the ipsilateral masking condition. Corbin et al. (2016) found gradual improvement in the recognition of mono-syllabic words in a speech-shaped noise masker when comparing a group of 8- to 12-year-olds with a group of 13- to 16-year olds, and hardly any improvement when comparing a group of 13- to 16-year-olds with a group of adults. While the difference between the group of 8- to 12-year-olds and the 13- to 16-year-olds corresponds with our study, the difference between the adolescents and the group of adults is difficult to compare, as the adults evaluated by Corbin et al. (2016) included listeners as old as 44 years of age.



The study by Elliott (1979) also revealed worse outcomes for children aged 13 years or younger compared with 15- and 17-year-olds, but used highly predictable sentences at three signal-to-babble ratios. When sentences are used, SRT screening is prone to effects of vocabulary or syntactic knowledge and differences in the ability to access the lexicon (Kaandorp et al., 2016). The Oorcheck tool used in the present study may be assumed to reduce effects of lexical context by using short and context-free words that are familiar to children, and by using labels with pictograms instead of written labels only.

Given the difference between child and adult perception, progressive improvement in SRT as seen in our data can be expected from childhood to adulthood. Younger listeners require a wider bandwidth to perform comparably with adults in speech identification tasks (Eisenberg et al., 2000, Hall et al., 2016). With the age effect depending on the spectral match between noise masker and speech stimuli, the low-pass filtered masker of the Oorcheck probably accentuated the performance differences between teenagers and adults in comparison to a speech-shaped masker.

Since our study group was homogeneous with respect to hearing thresholds (NH), the age effect in the adolescents' SRT probably reflects the fact that the structures that facilitate behavioral learning and better speech-in-noise perception were not available to teenagers to the same degree as to adults. The SRT "deficits" in the adolescents and which decreased with age might be attributed to their developing cognitive abilities and cognitive control, including memory capacities, experience, and selective attention. They presumably affected auditory sensitivity, such as discrimination and processing of acoustic-phonetic cues (Sussman, 1993, Gogtay et al., 2004, Wunderlich and Cone-Wesson, 2006, Sussman and Steinschneider, 2009, Anderson and Kraus, 2010, Medina et al., 2010, Parbery-Clark et al., 2011, Kraus et al., 2012, Hornickel et al., 2013, Moon et al., 2014, Strait et al., 2014, Westerhausen et al., 2015).

In addition to the age effect in our data, there was also a trend for SRT to improve between the first test and the immediate retest. The SRT improvements from test to retest in our data are probably an example of fast adaptation of auditory processes to incoming speech-in-noise signals (Skoe et al., 2013), including fast adaptation to the task, stimulus, or phonetic inventory of the speaker. In case of a failed first test, a retest is recommended. The relatively better first SRT score found in adults which hardly improved with retesting might show that the adults' auditory system was already well-tuned or adapted instantly, leaving little room for improvement.

In summary, when screening by SRT, the effects of the still maturing cortical and subcortical system on auditory speech and noise processing have to be considered. The studies referred above indicate that experience-related factors or auditory pathologies other than increased hearing thresholds are involved in the tuning of speech (-in-noise) processing in adolescents. While online speech-in-noise tests can offer an efficient way to screen for hearing impairment on a regular basis, the auditory system is in flux during childhood and adolescence, and age-related cut-off scores in SRT should be considered for this age period.

To check for an effect of potential differences in cognitive skills within an age-group, we included the level of education. Larger samples are needed to confirm the insignificance of the level of education and gender in our data. Future research might also assess the adolescents' musical training and bilingualism, since both have a significant effect on auditory processing of speech-in-noise (Tierney et al., 2015, Krizman et al., 2016). For musicians, auditory brain responses revealed superiority in the representation of timbre, pitch, and timing (Slater et al., 2015).

Since our study focused on the verification of an age effect in SRT scores of NH adolescents, it is not yet clear to what extent the proposed age-dependent "easing" of the cut-off scores might affect the test's ability to detect hearing loss in the respective age-groups. More research is needed to rule out possible negative effects of an age correction on the sensitivity and specificity of Oorcheck as a screening test. Our results should be considered in light of these caveats.

From the point of view of awareness and prevention of hearing loss, and considering the effect that hearing loss can have on a student's development, education, employment, rehabilitation costs, and retention rate (Bess et al., 1998) it should be noted that 22% (23/104) of the participants in the present study had to be excluded due to their elevated pure-tone hearing thresholds.



## Conclusion

Hearing screening tools which are based on thresholds for speech in noise should control for an effect of age when assessing adolescents. Based on the present data, we propose a correction factor of -0.2 dB per year of age for Oorcheck SRT cut-off scores for adolescents between the ages of 12 and 17 years. More data are needed to verify the present findings and proposed corrections.

# Chapter 7

# Evaluation of an internet-based speech-in-noise screening test for school-age children.

M. Sheikh Rashid[1]
W.A. Dreschler[1]
J.A.P.M. de Laat[2]

1 Clinical and Experimental Audiology, ENT Department, Amsterdam Public Health Research Institute, Academic Medical Center (AMC) Amsterdam, The Netherlands
2 Department of Audiology, Leiden University Medical Center, Leiden, The Netherlands

International Journal of Audiology,2017, 56 (12), 967-975.

## Abstract

**Objective.** To evaluate a Dutch online speech-in-noise screening test (in Dutch: "Kinderhoortest") in normal-hearing school-age children. Sub-aims were to study test–retest reliability, and the effects of presentation type and age on test results.

**Design.** An observational cross-sectional study at school. Speech reception thresholds (SRTs) were obtained through the online test in a training condition, and two test conditions: on a desktop computer and smartphone. The order of the test conditions was counterbalanced.

**Study sample.** Ninety-four children participated (5–12 years), of which 75 children were normal-hearing (≤25 dB HL at 0.5 kHz, ≤20 dB HL at 1-4 kHz).

**Results.** There was a significant effect for test order for the two test conditions (first or second test), but not for presentation type (desktop computer or smartphone) (repeated measures analyses, $F(1,75)=12.48$, $p<0.001$; $F(1,75)=0.01$, $p=0.982$). SRT significantly improved by age year (first test: 0.25 dB SNR, 95%-CI: -0.43- -0.08, p=0.004. Second test: 0.29 dB SNR, 95%-CI: -0.46- -0.11; p=0.002).

**Conclusions.** The online test shows potential for routine hearing screening of school-age children, and can be presented on either a desktop computer or smartphone. The test should be evaluated further in order to establish sensitivity and specificity for hearing loss in children.

## Introduction

Untreated mild to severe childhood hearing loss may have serious negative consequences for speech and language, educational and socio-emotional development (Davis et al., 1986, Brookhouser et al., 1991, Bess et al., 1998, Yoshinaga-Itano et al., 1998). Therefore, early identification is of great importance. Well-established neonatal hearing screening programmes in European countries, including otoacoustic emissions (OAE) or automated auditory brainstem response (AABR) screening, identify permanent congenital hearing losses, but they do not detect delayed-onset or acquired sensorineural losses (Skarzynski and Piotrowska, 2012, Winston-Gerson and Sabo, 2016). In the USA, the prevalence of mild permanent sensorineural hearing loss at 6 kHz in children aged six to nineteen years is 12.5%, and in children aged seven years 6% (Niskar et al., 2001a). Up to 90% more children are diagnosed with hearing loss before the age of nine years than are diagnosed as newborns Fortnum et al. (2001), (Prieve et al., 2015). Early diagnosis of hearing loss can be achieved with hearing screening during pre-school, and primary school years, reducing the impact on speech and language development (Lu et al., 2014). There are several screening methods to identify delayed-onset or acquired sensorineural hearing loss in pre-school- and school-age children. OAE and pure-tone screening are the most reliable and commonly used tools, though pure-tone screening is considered to be the preferred reference standard (Prieve et al., 2015). In the Netherlands, childhood hearing assessment is performed in all children between the age of four and six years. The assessment is performed at school by a youth health care nurse through pure-tone threshold screening at regular contact sessions.



When hearing screening is performed in the school setting, a large number of children can be reached (Winston-Gerson and Sabo, 2016). However, pure-tone screening in remote settings, such as schools, is often performed in less than optimal test conditions. High ambient noise levels, but also calibration issues and examiner's and examinee's training, experience and motivation, negatively influence the accuracy of screening results, making pure-tone screening less reliable for detecting hearing losses (Bamford et al., 2007, Schlauch and Carney, 2012, Kam et al., 2013, Prieve et al., 2015). Therefore, there is a need for appropriate, effective and efficient periodic hearing screening that can be performed accurately and reliably in school or other remote settings, to identify suspected mild to severe sensorineural hearing losses in pre-school and school-age children.

One of the early signs of hearing impairment is the difficulty experienced in understanding speech in background noise in daily situations (Smoorenburg, 1992, Kramer et al., 1998). Therefore, one potential approach to identify hearing loss is speech-in-noise testing. Advanced time-efficient online self-administered and automated speech-in-noise tests have been developed, that focus on the detection of sensorineural hearing losses (Leensen et al., 2011b, Jansen, 2013, Smits et al., 2013). The main advantages of such tests are that the tests are easily accessible and less susceptible to environmental noise (Smits et al., 2004, Culling et al., 2005). A Dutch online speech-in-noise hearing screening test for children has been developed by the Leiden University Medical Center and the Academic Medical Center in the Netherlands, and was implemented online in January 2007 (in Dutch: "Kinderhoortest"). The test was developed with the aim of allowing the evaluation of children's speech perception in noise in an easy and accessible way at a remote setting, such as the school environment. The goal of such testing would be the early detection of perceptive sensorineural hearing loss in school-age children. The relatively simple test with suitable speech material may be useful for children aged five years and older. An important limitation is that this test may not be assumed to be sensitive to conductive hearing losses caused by external or middle ear pathologies, such as otitis media. According to the underlying model by Plomp and Mimpen (1979a), speech-in-noise results do not lead to higher critical SNR's for pure conductive hearing losses. Conductive hearing losses are one of the potential forms of hearing losses in school-age children, though they are more common in pre-school-age children (Samelli et al., 2012). Most of the children experience temporary conductive hearing losses, which can be treated medically.

The main objective of this study was to evaluate the suitability of the Dutch online speech-in-noise screening test for use in primary school children. The sub-aims were to evaluate the test–retest reliability of the test, the effect of the presentation type: on a desktop computer or smartphone, and to assess age effects on test results.

# Methods

## Subjects

This study was performed in 94 primary school children. Recruitment took place at the Koningin Wilhelmina school in Rijnsburg, the Netherlands. Information letters, informed consent forms and short questionnaires were sent to the parents. All children were native speakers of the Dutch language. Speech-in-noise data were collected from 94 children. The results of 19 children were excluded from the analyses, because these children were younger than five years old (N=1), had poor hearing thresholds at one or more octave frequencies (≥ 25 dB HL at 0.5 kHz and/or ≥20 dB HL at 1-4 kHz) for at least one ear (N=10), had missing data on the speech-in-noise tests for at least one condition (N=2), had instable SRT measurements for at least one test (N=5), and had a floor score for at least one test (N=1) (the definitions of an instable measurement and of a floor score are explained in the section "Statistical analyses"). The data of the remaining 75 normal-hearing children were analysed further.

## Measurement procedures

This cross-sectional study was approved by the medical ethics committee of Leiden University Medical Center (project number P11-108). Informed consent was given by parents and the school's board of directors. For every child, information of concerning age, gender and grade was collected. All tests took place during school hours in quiet rooms at the school.

### *Pure-tone audiometry*

Pure-tone audiometry was performed as a reference standard. The five-year olds performed play audiometry. Hearing thresholds were measured for both ears at the frequencies 0.5, 1, 2 and 4 kHz. Bone-conduction thresholds were assessed as well when a hearing threshold was 20 dB HL or worse at one frequency. Pure-tone audiometry was performed in the teachers' office room, with an average background noise level of 43 dB(A). Because there was no soundproof cabin, and audiometric tests were potentially subject to environmental noise, a hearing threshold of 25 dB HL or better at 0.5 kHz was defined as normal. Between 1 and 4 kHz a hearing threshold of 20 dB HL or better was defined as normal. Children with poorer thresholds were referred for further investigation, and parents or caregivers were informed. For the pure-tone audiometry measurements the Interacoustics AD229b audiometer was used with Telephonics TDH-39P headphones with Amplivox Audiocups to attenuate ambient sound, and a Radioear B71 bone conductor.

***Speech-in-noise testing***

Children's perception of speech in noise was assessed by means of the online speech-in-noise test for children. First, a training condition was performed in a group session in class, i.e., all children belonging to one grade performed the test in the same computer classroom at the same time, but each performed the test individually on a personal desktop computer. Spoken instructions on the test procedure were given by the research assistant before the training test started. The children were instructed to identify the presented words by clicking on the corresponding pictures on the screen. They were also instructed to click on the picture depicting a question mark if a presented word could not be identified. Then, in two test conditions, all children were tested with a desktop computer, and with a smartphone. For these two test conditions, children performed the test one by one, separate from the other children, in the teachers' office room. The order of the type of presentation was counterbalanced. The computer classroom in which the children underwent the training condition had an average background noise level of 48 dB(A). The two test conditions took place in the same teachers' office room in which the pure-tone audiometry took place, with an average background noise level of 43 dB(A). The online speech in noise tests were presented on a standard desktop computer and DKT Eduline of Philips SHP2000 headphones, and on the smartphones Nokia Lumia 625 or the Huawei G6, with Ewent headphones.

The speech material consisted of a closed set of eight Dutch monosyllable consonant-vowel-consonant (CVC) words. The words were all nouns, and highly familiar for young children, as they were selected from the Dutch word lists used for diagnostic speech audiometry in children (Bosman, 1989). The response buttons on the screen were pictures accompanied by written words. The written words were: "lion", "goat", "book", "rose", "moon", "thumb", "fire" and "chicken" (in Dutch), and were all represented by easily recognisable pictures. To prevent guessing, a ninth response button with a question mark and the text "not understood" was added. The response screen is shown in Figure 7.1.


**Figure 7.1.** Response screen of the Dutch online speech-in-noise test for school-age children, in gray scale (In Dutch: "Kinderhoortest").

In order to enhance test reliability, the words were perceptually homogenised (Sheikh Rashid and Dreschler, 2014). To achieve equal intelligibility of the words, the presentation levels of the specific words were adjusted. These level corrections, based on the average SRTs for the individual words, were derived from the slopes of word-specific psychometric functions according to the method described in Leensen et al. (Leensen et al., 2011b). The word-specific psychometric functions were based on online results of tests that were performed by children, from January 2007 to August 2014 (N=46,742). Perceptually difficult words were amplified, and perceptually simple words were attenuated (the level corrections ranged between 1.51 and -2.55 dB). The words were presented in a masking noise, which was a broadband continuous noise, with a spectrum that corresponded with the long-term average speech spectrum of the homogenised word material. The test was diotic (binaural); i.e., both ears were measured at the same time. The volume level of the stimuli could be set by means of the volume scale to a comfortable level. The minimum and maximum volume levels were controlled for in the clinical setting, and set at 15 dBA (i.e., whisper level) to 85 dBA (without distortion of the sounds). To familiarise listeners with the test and the response buttons, the words were presented in a masking noise prior to the test. The test consisted of 20 stimuli. All words were randomised, and each word was presented two or three times. The test started with a signal-to-noise ratio (SNR) of -1 dB and the intensity of the word was being varied by means of an up-down procedure in steps of 2 dB SNR. The intensity of the masking noise was fixed. The speech reception threshold (SRT in dB SNR) at which 50% of the material is correctly understood, was based on the mean SNR values of the last 10 presentations. The test was presented in an HyperText Markup Language (HTML) format, and could, therefore, be performed on any electronic device that supported the format, such as a desktop computer, tablet, or smartphone. Test duration was approximately 3 min in all age groups.

### Statistical analyses

Data were analysed using SPSS (IBM SPSS Statistics 20 and 22; IBM Corp, Armonk, NY, USA). Results of normal-hearing children between the age of 5 and 12 years old were analysed. The data of children with incomplete or invalid test results, due to instable measurements or a floor effect, were excluded from the analyses. An instable measurement refers to an intra-individual standard deviation of 3 dB SNR or larger. A floor effect refers to a minimal SRT score. A floor effect can be the result of consecutive incorrect responses due to a hearing loss or not understanding the test procedure.

Descriptive analyses were performed on hearing thresholds and SRTs of the subjects. The normality assumption was assessed by means of Q-Q plots and goodness of fit tests. SRT data showed normal distributions. Therefore, General Linear Model Repeated measures analyses were performed on the two test conditions to analyse the effect of the type of presentation: desktop computer or smartphone (within-subject factor), the order: the first or second test (within-subject factor) and age in categories (between-subject factor) on SRT (in dB SNR). Post hoc analyses with Bonferroni corrections were performed when significant effects were found. To analyse SRT (in dB SNR) as a function of age (in years) and test (training condition, first and second tests), multiple regression analyses were performed. In order to assess the consistency of the first test and second results, a measurement error was calculated by taking the quadratic mean of the within-subject standard deviations of the repeated measurements. Finally, in order to assess age-related differences, a regression analysis was performed on SRT scores (in dB SNR) of the first and second test, as a function of age (in years).

## Results

Table 7.1 shows the pure-tone average (PTA) thresholds for the octave frequencies 0.5, 1, 2 and 4 kHz (PTA5124). The mean results are given per age group and per ear. At least 10 children participated per age group, except for the youngest age group (N=7). The 11– 12-year olds (N=8 and N=2, respectively) are clustered in the oldest age group (≥11 years).

**Table 7.1.** Age group (in years), number of participants per age group, and mean pure-tone average (PTA) for the frequencies 0.5, 1, 2, and 4 kHz (PTA5124) (in dB HL) (SD) for right and left ear.

| | | $PTA_{5124}$ | |
|---|---|---|---|
| **Age group** | **N (75)** | **Right ear** | **Left ear** |
| *5* | 7 | 12.9 (1.2) | 13.6 (2.3) |
| *6* | 13 | 8.8 (5.3) | 9.2 (4.6) |
| *7* | 10 | 7.5 (5.0) | 6.3 (3.8) |
| *8* | 11 | 5.7 (5.4) | 5.6 (4.7) |
| *9* | 13 | 5.7 (5.3) | 7.8 (5.5) |
| *10* | 11 | 5.7 (5.1) | 5.6 (6.0) |
| *≥11* | 10 | 8.8 (3.8) | 8.0 (2.8) |

**SRT scores** for all test conditions.

7

The mean SRT scores (in dB SNR) for the training condition and the two counterbalanced test conditions (first and second tests, and on desktop computer and smartphone) were calculated for each age group (Table 7.2). Children performed better on both test conditions (desktop computer and smartphone) as compared to the training condition, with a significant difference in mean SRT of -1.5 dB SNR ($F(2,75)=17.64$, $p<0.001$). For the two test conditions, there was no significant main effect for type of presentation (desktop computer or smartphone) (F(1,75)=0.01, p=0.982), but there was a significant main effect for test order (first or second test) (F(1,75)=12.48, p<0.001). The mean SRT scores of the second test were significantly better than the SRT scores on the first test with a difference of 0.7 dB SNR (95%-CI 0.3- 1.1; p=0.001). The main effect of age was significant as well (F(6,75)=3.09, p=0.01). Post-hoc analyses showed that the 10–11-year olds performed better as compared to the 5–6-year olds (with a difference of 1.9 dB SNR, 95%-CI -3.7- -0.1; p=0.023), and to the 6-7-year olds (with a difference of 1.6 dB SNR, 95%-CI -3.1- -0.1; p=0.028). There were no significant interaction effects between age and type of presentation (F(6,75)=0.67, p=0.674), or age and test order (F(6,75)=1.06, p=0.393).

**Table 7.2.** Mean SRT (in dB SNR) (SD) per age group (in years), for training and two test conditions.

| | | Mean SRT (dB SNR) (SD) | | | |
|---|---|---|---|---|---|
| | | Test order | | Presentation type | |
| **Age group** | **Training** | **First** | **Second** | **Desktop computer** | **Smartphone** |
| *5* | -11.8 (2.4) | -12.6 (1.2) | -13.8 (1.8) | -13.3 (0.6) | -13.1 (2.2) |
| *6* | -11.2 (2.6) | -13.2 (1.2) | -13.8 (1.1) | -13.9 (1.0) | -13.1 (1.3) |
| *7* | -12.0 (2.4) | -13.8 (1.3) | -14.2 (1.1) | -14.0 (1.1) | -14.0 (1.3) |
| *8* | -12.6 (1.6) | -13.9 (1.3) | -14.1 (1.6) | -13.6 (1.8) | -14.3 (1.0) |
| *9* | -12.9 (1.8) | -13.9 (1.6) | -14.0 (1.5) | -14.0 (1.5) | -13.9 (1.5) |
| *10* | -14.0 (1.3) | -14.7 (1.1) | -15.5 (1.8) | -15.1 (1.5) | -15.2 (1.6) |
| *≥11* | -13.6 (1.5) | -13.8 (2.2) | -15.5 (1.4) | -14.6 (2.5) | -14.8 (1.4) |
| *Total* | -12.6 (2.1) | -13.8 (1.5) | -14.4 (1.6) | -14.1 (1.6) | -14.1 (1.6) |

The results of the multiple regression analysis with SRT (in dB SNR) as a function of age (in years) and test (training condition, first test and second test) are shown in Table 7.3. According to the model, the mean SRT score for a five-year-old child in the first test session was -12.6 dB SNR. There was a significant improvement (decrease) in mean SRT score of 0.3 dB SNR per age year. Performance on the first test was 1.2 dB SNR better than on the training condition. Performance on the second test was 0.7 dB SNR better as compared to the first test. These differences were statistically significant. The measurement error between the first and second tests was 1.3 dB.

## Age-related differences

In Figure 7.2, the SRT score by age in percentiles is given for the first test. To analyse age-related differences in SRT, a regression analysis was performed, with SRT (in dB SNR) of the first test as outcome measure and age (5–11 years) as an explaining factor. According to this model, there is a significant improvement of 0.3 dB SNR in mean SRT score per age year (β=0.25, 95%-CI -0.43- -0.08; p=0.004, R2=0.11). A comparable age effect was observed in the second test (β=0.29, 95%-CI-0.46- -0.11; p=0.002, R2=0.13).

**Table 7.3.** Multiple regression analysis with SRT (in dB SNR) as a function of age (in years) and condition (training, first and second test) (reference= first test).

| | | β | *p* | **95% CI** | |
|---|---|---|---|---|---|
| *Constant* | | -12.58 | <0.001 | -13.13 | -12.03 |
| *Age (in years)* | | -0.33 | <0.001 | -0.44 | -0.21 |
| *Condition* | | | | | |
| | *Training* | 1.16 | <0.001 | 0.63 | 1.68 |
| | *Second test* | -0.68 | 0.013 | -1.20 | -0.15 |

*Explained variance* $R^2 = 0.27$

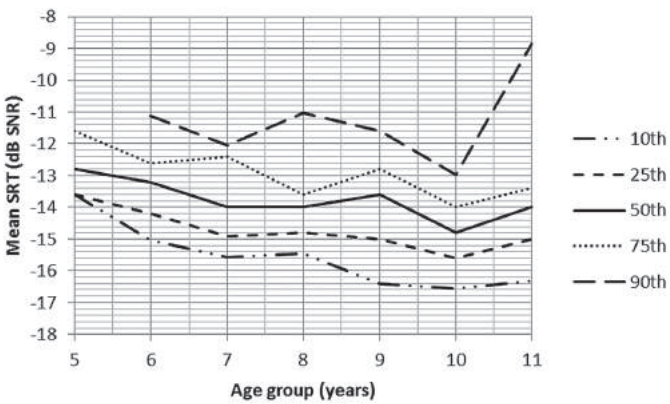


**Figure 7.2.** Reference values. Mean SRT in dB SNR by age group for the first test.
*Distribution in percentiles (10th,25th, 50th, 75th, 90th).*

To establish age-corrected cut-off values for pass-refer criteria for the screening test, the 90th percentile of the SRT results of the test for six-year olds was used as a starting point. The beta-value of -0.25 dB SNR per age year for the first test was then used to correct for age. The age-corrected cut-off values are presented in Table 7.4. These cut-off values were then applied to the results of the first test of the 75 children. Based on this categorisation, 92% (N=69) of the normal-hearing children passed the test.

**Table 7.4.** Age-corrected cut-off values.

| Age category (years) | Cut-off value (score positive result) (dB SNR) |
|---|---|
| *5–6* | >-11.00 |
| *6–7* | >-11.25 |
| *7–8* | >-11.50 |
| *8–9* | >-11.75 |
| *9–10* | >-12.00 |
| *10–11* | >-12.25 |
| *11–12* | >-12.50 |

## Discussion

This study focussed on the practical evaluation of the Dutch online speech-in-noise screening test in normal-hearing school-age children of 5–12 years old. To assess the reliability of the test, the test was performed at a primary school: first, a group training session was performed on a desktop computer, than two test conditions were performed, on a personal desktop computer and on a smartphone. The order of the test conditions was counterbalanced. The two tests were performed better as compared to the training, with a difference between the average of both tests, and the training test of 1.5 dB SNR. There were no significant differences in SRT score by type of presentation. There was an effect of test order between the two test conditions, indicating an additional learning effect after training of 0.7 dB SNR. A measurement error of 1.3 dB was found, indicating reasonable test-retest reliability. The standard deviations on the test and retest conditions were smaller than the step size that was used in the adaptive procedure (2 dB), indicating the homogeneity of the participant's results. Furthermore, according to the regression analysis, the oldest children had better SRT scores as compared to the youngest children, with a difference in the order of 1.5 dB SNR for the first test.

## Presentation mode

According to this study, the type of presentation (i.e. electronic device) did not influence SRT score. The test can be performed either on a desktop computer or on a smartphone combined with commercial headphones. The expectation that the test can be delivered on all types of electronic devices that support HTML applications is supported by the results of this study. Also, the children did not experience any difficulties in using the different electronic devices. According to a study by Culling, Zhao, and Stephens (2005), variations in equipment and listening environment do not present any significant obstacles to the development of a self-administered screening test based on speech in noise. There are several hearing screening tests delivered on different types of electronic devices (Leensen et al., 2011b, Jansen, 2013, Smits et al., 2013, Potgieter et al., 2015). Studies on these computer- and smartphone-based speech-in-noise screening tests have shown that there are indeed no significant effects of transducer type on test outcome. Jansen (2013) showed that there is no significant effect of transducer type (headphones, built-in laptop speakers, in-ear phones and external speakers) on SRT in uncontrolled circumstances, for the Flemish computer-based digit triplet test. Recently, a smartphone-based digits-in-noise hearing test in South African English has been developed and validated (Potgieter et al., 2015). It was investigated whether different types and quality of headphones, including standard smartphone headphones and clinical headphones, would influence SRT. Statistically significant effects were not found. The South African smartphone-based screening test is based on the digit triplet test, developed by Smits, Goverts, and Festen (2013). Although the current test uses CVC words in noise, it is based on the same principles of speech-in-noise hearing testing.



## Test-retest reliability

To assess the test-retest reliability of the speech-in-noise test, the first test was compared to the second test. Children performed significantly better on the second test as compared to the first test, indicating a learning effect of 0.7 dB SNR. It is important to note that the children were already familiar with the test procedure and the word material, because of the training condition that was performed prior to the two test conditions. For this reason, it cannot be ruled out that the actual learning effect is even greater than the learning effect found between the first and the second tests. It is unclear to what extent the training condition may have influenced test results. The difference of 1.2 dB SNR found between the training condition and the first test indicates an initial learning effect, but part of this difference could be ascribed to other factors related to testing together in one classroom, such as distraction.

Test-retest reliability of speech-in-noise tests in children has been studied earlier. Schafer et al. (2012) assessed the test-retest reliability of the Phrases in Noise Test (PINT) in normal-hearing children, a speech-recognition test for use in a clinical or educational setting. The PINT seemed fairly reliable as differences between two lists were within 3 dB SNR for 90% of the children. A smaller learning effect was found in the current study. This may be due to the use of a closed set of highly familiar words instead of sentences that have higher linguistic demands (Smits et al., 2013). Jansen (2013) investigated the feasibility of the digit triplet test as an automated self-test in school-age children, and found a measurement error of 0.5–0.7 dB for different age groups. The smaller measurement error may be a result of the use of digit triplets as speech material instead of single CVC words, leading to more reliable estimates of the SRT (Jansen et al., 2014b).

### Age-related effects

In this study, age-related effects were present; the older children outperformed the younger children in all test conditions. There were no significant interactions with presentation type or order; age effects were consistently the same in all conditions, and were also present in the first and second tests. Several studies have been focussing on (school-age) children's ability to recognise speech in noise, and age-effects in auditory processing abilities (Elliot, 1979, Elliott et al., 1979, Fallon et al., 2000, Johnson, 2000, Talarico et al., 2007, Vaillancourt et al., 2008, Schafer et al., 2012, Jansen, 2013, Koopmans and Smits, 2015). In these studies, several auditory tasks and speech-in-noise tests were performed in different noise conditions. The majority of these studies has demonstrated maturation of the auditory system of normal-hearing children. Speech-in-noise recognition tends to improve with age and adult-like performance is reached in adolescence, depending on the speech-in-noise listening condition (Fallon et al., 2000, Johnson, 2000, Talarico et al., 2007, Vaillancourt et al., 2008). Fallon, Trehub, and Schneider (2000) found that five-year old children required SNRs that were 5 dB more favourable than those of adults to obtain comparable performance on low-context sentences presented in background babble. Talarico et al. (2007) investigated the effect of age and cognition in 6- to 16-year olds with a task that included (non-)words in noise, varying in confusability and difficulty. Mean SNR scores decreased across all age groups, indicating better speech-in-noise recognition in the older children (up to 3 dB SNR). No correlations were found between speech in noise conditions and IQ scores. According to Elliot (1979), there are developmental changes in SRTs of young children up to 16 years of age. Age effects on SRT performance are mainly explained by developing auditory processing abilities, associated with developing linguistic skills, and cognition-related abilities such as memory capacities, experience and attention (Elliot, 1979, Elliott et al., 1979, Boothroyd, 1997, Hnath-Chisolm et al., 1998, Eisenberg et al., 2000, Fallon et al., 2000, Vaillancourt et al., 2008).

For the test session in this study, five-year olds required SNRs that were 1.5 dB more favourable than those of 11–12-year olds to achieve comparable 50% of correct performance. Jansen (2013) assessed the reference SRT for normal-hearing listeners for the digit triplet test and found that the SRTs of the 5th graders are 0.6 dB SNR worser as compared to those of the 7th graders. In the present study, a comparably small age-effect was found. This may be due to the relatively simple test procedure and the use of a closed-set of highly familiar, short and context-free monosyllabic words, supported by visual response buttons with pictograms and written words. The influence of linguistic abilities is expected to be small in this type of task (Fallon et al., 2000, Jansen, 2013). The age-effect found in our study may be mainly a result of immature auditory perceptual abilities, combined with the influence of attentional limitations (Schafer et al., 2012, Jansen, 2013), and the difficulty experienced in understanding test instructions in younger children.



### Study limitations

This research has some limitations. First, the inclusion of normal-hearing children was based on hearing thresholds measured by means of pure-tone audiometry, which was not performed in an sound-isolated booth, but in the teachers' room. Although the room was considered quiet, as confirmed by the ambient noise-level measurements that were performed, environmental noise could not be completely avoided, and this may have influenced the pure-tone measurements. To attenuate the ambient noise, audio cups were used in combination with the headphones. Also, to assure normal-hearing, the criteria for normal hearing (at the lower octave frequencies) were adjusted. Looser threshold criteria are reasonable in school settings (Kam et al., 2013). However, children may have performed worse than they would have under optimal test conditions, which implies that possibly some normal-hearing children may have been excluded.

According to the pure-tone screening, 10 children had poor hearing thresholds at one or more octave frequencies (≥ 25 dB HL at 0.5 kHz and/or ≥20 dB HL at 1-4 kHz) for at least one ear, and their results were, therefore, excluded from the analyses. Based on the established age-corrected pass-refer criteria, only one of them failed the online speech-in-noise test (age=5 years, SRT score for the first test= -10.8 dB SNR). This child had a PTA512 of 20 dB HL for the right ear, and 23 dB HL for the left ear. The large number of false-negatives, however, could possibly be explained by the less reliable test environment of the pure-tone screening (i.e., the false-negatives could actually be true-negatives). Another explanation could be that, since the test was binaural, children with an unilateral hearing loss were still able to pass the test (four out of the remaining nine children had an unilateral hearing loss). This may be an important limitation of the test. Also, the majority of the children with a bilateral hearing loss had a relatively small, and educationally insignificant, hearing loss, and probably were, therefore, still able to perform well on the online speech-in-noise test. In order to assess an optimal cut-off point for a dichotomous pass/fail outcome with a proper trade-off between sensitivity and specificity for clinically relevant hearing losses, it is necessary to include a large representative sample of children showing a wide range of hearing thresholds. This cut-off point would probably correspond to a higher degree of hearing loss than the relatively strict criteria that were proposed in this study.

Another limitation is that the youngest age group (the five-year olds) was underrepresented in this study as compared to the other age groups. Results of this age group may be less reliable. Also, the five-year olds had some trouble understanding the test instructions. For these young children, it was difficult to understand the goal of the test and the procedures. As the suitability of the test is still unclear for young children, reference values are only established for children older than five years. It is important to evaluate the established reference SRT values in larger populations. Also, it is important to have simple, clear and understandable instructions for the youngest children.

Finally, due to the setup of this research, it is difficult to distinguish a learning effect from the effect of test condition for the training condition versus the first test. Testing in a classroom setting may be less reliable than separate testing in a teacher's office room, due to distractions in the classroom that may hinder children's listening and focussing abilities (Knecht et al., 2002). According to Culling, Zhao, and Stephens (2005), group presentation of speech-in-noise tests in classrooms should be discouraged, mainly because of the potential negative effect of high levels of room reverberation. The training effect as well as simultaneous group testing versus separate testing in different settings need to be explored further. To assess test-retest reliability in more detail, tests and (multiple) retests should be performed under the same conditions, and with different time-intervals.

## Implications for practice and future research

The speech-in-noise test has important implications for hearing loss screening purposes in school-age children. In the Netherlands, online-speech-in-noise tests are already being used frequently to raise awareness in teenagers and adolescents (de Laat et al., 2016), but not yet for screening purposes. The current test appears to be suitable to be used in a national hearing screening programme, as it is a simple test, appropriate for small children, which can be performed in 3 min when performed binaurally. Due to the type of speech material, the influence of cognition, attention and linguistic demands is minimal. The independency of the test for soundproof test rooms and type of presentation creates opportunities for time-efficient simultaneous group testing and screening in remote settings. However, before the test can be implemented as a screening test, it is important to assess its sensitivity and specificity for detecting clinically relevant or educationally significant degrees and types of hearing losses in children. The test, therefore, needs to be evaluated in a larger representative sample of school-age children, including hearing-impaired children with a large range of hearing losses, in a realistic testing environment such as a school setting. The hypothesis is that children with hearing loss will have higher SNRs as compared to normal-hearing children; however, differences in performance should be investigated. In addition, the current test was conducted binaurally. However, in order to detect unilateral hearing losses, the test should be evaluated when conducted monaurally, as well. Furthermore, to establish sensitivity and specificity, it is of great importance to compare the speech-in-noise test with a reliable reference standard. Therefore, pure-tone audiometry should be performed in better test conditions as compared to the test conditions in the current study.

This research shows an age-dependency for SRT in normal-hearing children (an amelioration of -0.25 dB per age year). The test result can be misleading if this is not corrected for. Therefore, the suggestion is to use age-corrected cut-off values in order to prevent false interpretations of positive test results of young children. The proposed age-corrected SRT cut-off values need to be validated in hearing-impaired children as well. Furthermore, this research indicated a learning effect. Learning effects may be accounted for by training or repeated conditional testing, i.e., introducing an automatic retest for children who failed the test. The possible influence of a learning effect on screening test outcomes should be studied further.

## Conclusions

The online-speech-in-noise test with simple word material was shown to be appropriate for use in school-age children, and shows potential for a routine-hearing screening test. The test can be conducted simultaneously in a classroom setting, and can be delivered on either a desktop computer or on a smartphone in combination with commonly available headphones. When testing, age and learning effects should be considered. Age-corrected SRT cut-off values for pass/refer categories are proposed for screening purposes. A learning effect exists which could be reduced by training and/or conditional repeated testing. The test should be evaluated further in a larger representative population of school-age children, including hearing-impaired children, in order to evaluate its sensitivity and specificity for identifying childhood-hearing loss in realistic screening settings.

# Chapter 8

## General discussion

# General discussion

The main theme of this thesis is the evaluation of internet-based speech-in-noise hearing screening tests for noise-induced hearing loss (NIHL) in two distinct populations: the working-age population exposed to occupational noise on the one hand, and the general teenage population exposed to recreational noise on the other. The main objectives of this research were to evaluate and improve existing tests, in order to obtain accurate and suitable screening tests. The ultimate goal is to develop efficient hearing screening tests that will benefit overall public hearing health by preventing or delaying the onset of NIHL, and by supporting early diagnosis and treatment of NIHL. In this chapter, the value of the evaluated internet-based speech-in-noise tests for screening purposes is discussed, and the study results are interpreted. Furthermore, implications for future research and practice are presented.

## The value of Internet-based speech-in-noise tests for NIHL screening

### *Online screening of occupational NIHL*

The phased evaluation of Occupational Earcheck (OEC) for occupational NIHL screening presented in this thesis, has resulted in some important test improvements and provided important insights concerning its performance. The optimized test in terms of homogenization of the speech material, the application of a low-pass (LP) filtered masking noise instead of a broadband noise, and adjustment of the test procedure, was found conceptually sound in a laboratory setting. Reasonably valid and reliable results were attained when testing different groups of noise-exposed employees in more realistic occupational environments on the individual level instead of single ear testing. Furthermore, OEC was more reliable when performed more than once, as conditional rescreening resulted in a higher test specificity. Hence, OEC appears to be an appropriate screening test for occupational NIHL, especially for moderate NIHL, when performed in an environmental setting, and when a conditional rescreen is applied.



In order to get an impression of the concrete value of OEC for NIHL screening, a closer look is given to the last evaluation of OEC, as described in Chapter 4. This chapter presents test measures that are the closest to the intended purpose of

OEC. OEC does not seem very sensitive to detect mild high-frequency hearing loss (HFHL, defined as a PTA346 of >25 dB HL for at least one ear), and there is a risk that a significant proportion of mild cases are going to be missed. However, mild HFHL is not a severe condition, and the missing of these cases may be perceived as less serious when screening is offered on a regular base, as the missed cases may be detected in follow-up screening rounds. On the other hand, specificity for mild HFHL was quite high after automatic conditional rescreening. This means that there is a low number of false-positive cases, resulting in a small number of unnecessary referrals to further comprehensive and costly hearing assessments, which is an important advantage for the employee as well as for the employer. If the purpose is to screen for moderate HFHL (defined as a PTA346 of >35 dB HL for at least one ear), OEC incorporating automatic conditional rescreening appears to be both highly sensitive and specific, resulting in less misclassifications.

***Practical example of OEC's test properties***

In the following practical example, it is illustrated how to exactly interpret the results of Chapter 4, when a conditional rescreening is performed instead of a single screen, and for different degrees of HFHL (Figure 8.1). The prevalence of mild HFHL (HFHL25) in this specific group of occupational noise-exposed employees was 42.5%. The sensitivity and specificity of OEC for a single screen were 65% and 63%, respectively, while when a conditional rescreen was applied, these were 59% and 93%. The prevalence of moderate HFHL (HFHL35) was 22.5%, and the test sensitivity and specificity for a conditional rescreen were 94% and 90%, respectively.

For a single screen, in an average company with, for instance, 200 employees, this would theoretically result in 98 employees with a positive test result, of which 55 persons would actually have a mild HFHL. While 43 persons would not have a mild HFHL, and would therefore probably be unnecessarily referred (i.e. a positive predictive value, PPV, of 56%). Of the 102 employees with a negative test result, 72 would indeed have no HFHL, while 30 persons would actually have a mild HFHL, and would therefore be missed (i.e. a negative predictive value, NPV, of 71%). In the case of a conditional rescreen, a smaller number of employees would get a positive test result (n=58), while a larger number of employees would get a negative test result (n=142), leading to a smaller number of false-positives and a larger number of true-negatives, accompanied by a higher PPV (86%) and comparable NPP (75%). Hence, if the goal was to screen for mild HFHL, there would be eight false-positives, or over referrals, and 35 false-negatives, or employees with at least mild HFHL that would be missed. If the goal was, however, to screen for moderate HFHL using a conditional rescreen, 43 out of 58 persons with a positive result would have at least a moderate HFHL, while 15 persons would be false-positive, with a mild HFHL or no HFHL at all (i.e. a PPV of 74%). While, out of the 142 test negatives, 140 employees would be classified correctly (as having no or only mild HFHL), and only two cases would be missed (i.e. a PPV of 98%). These numbers show that OEC with a conditional rescreen of moderate HFHL is fairly accurate, and leads to less misclassifications as compared to a single screen for mild HFHL (8.5% vs. 36.5%).

Given a fixed cut-off value, the distribution of positive and negative test results is the same for both situations: the identification of mild HFHL or worse, or the identification of moderate HFHL or worse. Nonetheless, it shows that the insecurity of OEC mainly lies in the detection of mild HFHL; only 1% of all test negatives are missed cases of moderate HFHL. Thus, if one is interested in detecting early/beginning HFHL, OEC may be less accurate. On the other hand, when one is testing older noise-exposed employees populations with already established age-related hearing loss, one might not be interested in detecting smaller hearing deficits, and screening with OEC for moderate HFHL becomes more suitable. The employees with mild HFHL, classified as false positive for moderate HFHL, would still be referred for further assessment, which is positive in clinical terms. This merely indicates the importance of the intended screening aims, and the consequences of screening.

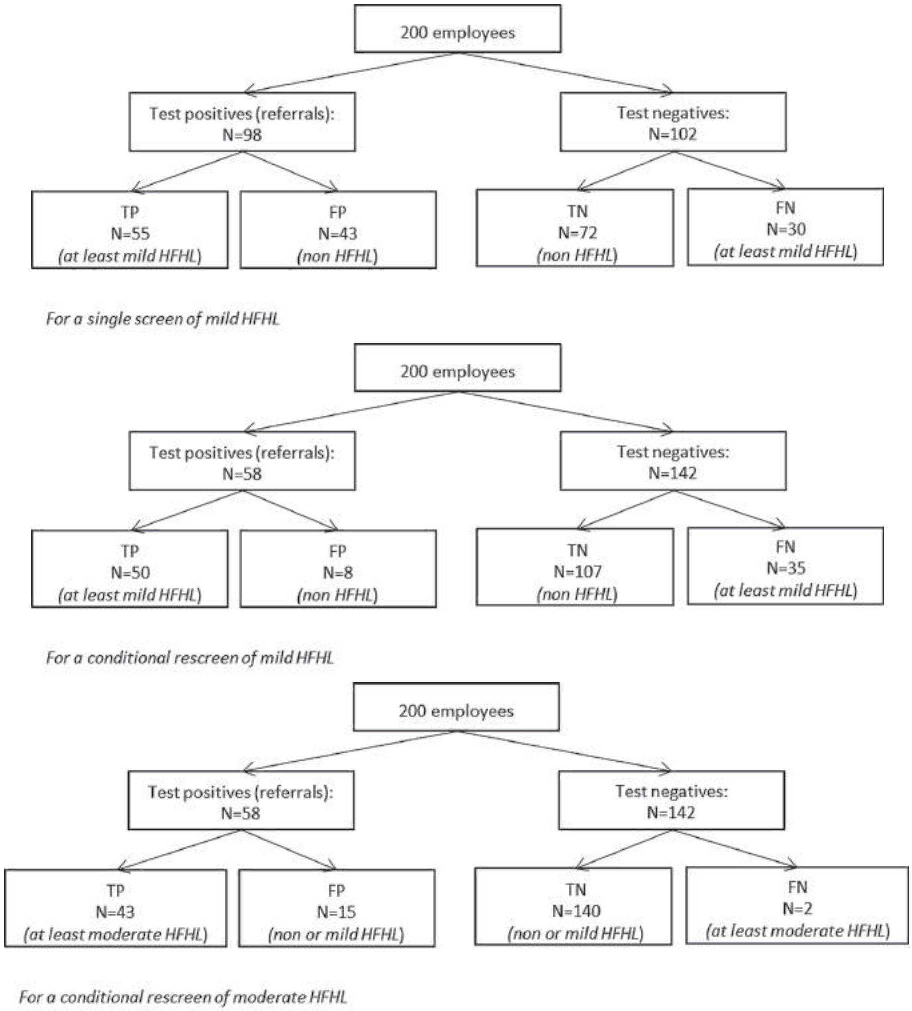


**Figure 8.1.** Practical example of OEC's test properties.

This example describes how the test properties of OEC behave in this specific situation, and although this study was performed in realistic screening circumstances, the test results cannot be generally applied (i.e. the actual numbers and percentages presented). Since test accuracy measures are not fixed and may vary due to the underlying clinical variability, such as the prevalence and severity of NIHL in certain populations, but also due to differences in test setting or test procedure (Leeflang et al., 2009). Moreover, the cut-off value for passing or failing OEC, as chosen in Chapter 3, was based on the most optimal balance of high sensitivity and specificity values, as both were initially considered important for evaluation purposes. However, in screening practices, the most optimal cut-off value should ideally be based on a proper trade-off between sensitivity and specificity, taking into account important factors such as the actual role of the test in practice, the severity of the target condition, and the burdens and costs of testing (Grobbee and Hoes, 2009). In certain circumstances a higher sensitivity may be preferred, which would consequently result in a decrease in specificity, or vice versa. It is important to note that changes in the cut-off point, and consequently in sensitivity and specificity and false positive and false negative test results, would have consequences for the actual number of missed or unnecessarily referred cases, Hence, if another cut-off point would have been chosen, the numbers illustrated in the above example would also differ.

### *Online screening of recreational NIHL*

For Earcheck (EC) and the Dutch online hearing test for children, some important milestones have been achieved towards the actual application as universal screening tests for young people. According to the results of the studies described in Chapters 5 and 6, supervised self-testing in the school setting seems more appropriate as compared to unsupervised domestic or remote self-testing, assuring proper test instructions and creating less uncertainties concerning volume settings and ambient noise levels during testing. Furthermore, tests results should be accounted for age effects, using age-appropriate correction factors, and for learning effects, by comprehensive training and multiple testing, in order to attain more reliable outcomes. For instance, an automatic rescreen for these tests should be considered as well. Furthermore, the tests can be delivered on different types of electronic devices, such as smartphones and tablets, creating more innovative opportunities to reach out to the young generation. However, it is yet unclear how sensitive both tests are for detecting NIHL (or other relevant hearing losses) in children and teenagers. Therefore, in this stage, the screening tests are suitable and recommended as an awareness tool, however the application as a valid screening test should be handled with care.



## Methodological considerations

In order to adequately interpret the findings of this research some important methodological issues should be considered.

***Effects of the reference standard***

In this research the conventional clinical diagnostic standard pure-tone audiometry was used as the reference standard. However, it is unlikely that this test is perfectly sensitive, especially when performed at a remote setting (Schlauch and Carney, 2012). In the first OEC evaluation, testing was done in a well-controlled laboratory setting (Chapter 2), in a soundproof room. This is in contrast to the second, more practice-based evaluation, in which the cut-off value for pass or fail criteria was established. In this study, reference testing was done in a quiet office room, and may have been less reliable due to these less-controlled circumstances (Chapter 3). Such variations in reference testing may consequently lead to less reliable verification of index test results. As a result, this may affect the prevalence and the test accuracy, and to some extent, explain the variation between results of different studies. This is known as the artefactual variability (Leeflang et al., 2009).

Furthermore, in the literature, there are no standardized definitions for NIHL or HFHL, and hearing loss classifications may vary per study or per purpose in terms of the test frequencies included, the degree of loss in dB HL, and ears included (uni- or bilateral hearing loss). These are all factors that influence the reported prevalence (Carter et al., 2014, Fredriksson et al., 2016, Su and Chan, 2017). For the purpose of this thesis strict definitions for HFHL were maintained throughout the studies, which may explain why relatively high HFHL prevalence numbers were found, that may have influenced test accuracy as well.

***HFHL and NIHL***

Another important methodological aspect is the main cause of hearing loss established according to the reference standard. The pure-tone audiogram is not able to distinguish between NIHL and other forms of HFHL, such as presbyacusis. In case of NIHL, a typical noise notch is not always found, and when NIHL is more severe, other frequencies can be affected as well. Especially when NIHL develops in older people, the effects of NIHL and presbyacusis may be cumulative, and it becomes hard to discriminate between the actual causes of the hearing loss. Likewise, OEC is able to discriminate HFHL from normal-hearing (or from isolated low-frequency hearing losses), however, the HFHL could be due to noise, as well as to aging, or less likely, to the use of ototoxic agents and medicines. So, equally with pure-tone audiometry, further complementary audiological assessments, such as retrospective self-reported assessments of occupational or recreational noise exposure, should be conducted in order to get an insight in whether the HFHL is related to noise exposure, so that appropriate measures can be taken.

***Test precision and sample size***

The sample sizes in this research were rather small for accurately establishing the test accuracy of a single test, which may have led to a lower test precision, and to lower generalizability of test results. In order to perform adequate sample size calculations prior to inclusion of the study sample, it is important to have good estimates of the prevalence of HFHL in certain populations. However, the noise-exposed working population is not easy to define, and reliable estimates are lacking. In table 8.1, the point estimates as well as the confidence intervals (CI) are given for test accuracy measures that were found in the second and the third practical evaluation of OEC (Chapters 3 and 4). According to a rule of thumb, 5% on both sides of the intervals suggests a good test precision. The precision of the estimates was indeed affected by the limited sample size, resulting in wider CI's for single screen results as compared to conditional rescreening. Nonetheless, the lower levels of the 95% CI for OEC with conditional rescreen still surpassed 80% (except for sensitivity for HFHL25), a value generally recognized as acceptable. However, OEC's test accuracy measures presented in this research are in the same order of previously found EC results (Leensen, 2013). Even though the studies were performed in different work-aged populations, the estimates are pointing towards the same direction. This strengthens the reliability of the findings.

**Table 8.1.** Point estimates for OEC test performance measures, with lower and upper 95% CI.

| | | *HFHL Prevalence* | *Sensitivity (95% CI)* | *Specificity (95% CI)* |
|---|---|---|---|---|
| **Chapter 3 (N=87)** | | 32% | 90 (79-100) | 77 (66-87) |
| **Chapter 4 (N=80)** | *Single screen (HFHL25)* | 43% | 65 (49-81) | 63 (49-77) |
| | *Conditional rescreen (HFHL25)* | 43% | 59 (42-75) | 93 (86-100) |
| | *Single screen (HFHL35)* | 23% | 100 (100-100) | 66 (54-78) |
| | *Conditional rescreen (HFHL35)* | 23% | 94 (84-100) | 90 (83-98) |

***Test-retest reliability***

The studies in this thesis have shown that the test-retest reliability of the three evaluated online speech-in-noise tests was reasonably high. The OEC showed a learning effect between ears and/or between tests. Attempts were made to reduce the effects of the learning effect on test accuracy by introducing an individual starting level, a long run-up to the actual measurement, and finally, by introducing conditional retesting. For the hearing screening test for children a training session was presented. The last adjustment has significantly improved test specificity, resulting in a smaller number of false-positive results. According to Leensen et al. (2013b), indeed the first two tests in a session of multiple tests are mainly responsible for the large learning effect. So by introducing a conditional practice test, test-reliability is increased, without significantly increasing testing duration. However, the learning effect observed during these first tests may limit OEC's suitability for the detection of very small hearing deviations or individual hearing monitoring over time.

***The speech material***

For OEC, the words were carefully selected based on their high frequency information (containing a high proportion of HF consonants), and paired vowels. For EC, the vowels were unique. The hypothesis was that OEC would be more sensitive due to this word selection. Although it is hard to demonstrate, as there was no direct comparison study conducted and the results from the different studies could not be easily compared due to differences in testing procedure, test settings, and test populations, this does not seem to be the case. The word material of OEC did not lead to higher sensitivity for NIHL and sensitivity and specificity seemed comparable to EC values (Leensen, 2013). This was still the case in LP noise stressing the available high-frequency information. This may suggest that other factors may affect sensitivity and specificity more, such as the conditional rescreening incorporated in the testing procedure. Also, the reference standard and the population characteristics may have played a role.

## Comparison with other online hearing screening tests

Although there is much heterogeneity between studies on hearing screening tests, concerning study populations, definitions of hearing loss and cut-off values for pass and fail criteria, measurement procedures, and intended use, comparisons with existing online hearing screening strategies are important in order to place this research into context.

***Low-pass filtered noise versus broadband noise***

In Chapter 2, a better sensitivity for OEC in a LP filtered noise instead of a broadband noise was found. This is in line with results of previous research (Leensen et al., 2011b, Jansen et al., 2014b, Vlaming et al., 2014). These studies all showed better test characteristics in LP noise using either CVC words or digit triplets. However, according to a recent study on the Flemish Digit Triplet Test (DTT) hearing screening comparing the DTT in broadband noise and the DTT in a LP-filtered noise, both tests were equally sensitive and specific to detect mild or greater HFHL in a middle-aged population (Vercammen et al., 2017). This difference in findings may be explained by differences in population characteristics between the study populations, such as age, and other measurement procedures applied.

***Smartphone-based hearing screening***

In Chapter 7, the Dutch hearing screening test for children was presented online via a desktop pc and via smartphone. There were no significant differences in SRT score by type of presentation, suggesting innovative possibilities for time-efficient simultaneous group testing and screening in remote settings. Smartphone-based speech-in-noise tests are a relatively new concept, and only few studies have focused on this type of test application (Potgieter et al., 2015, Potgieter et al., 2017). These studies on the South-African English smartphone-based digits-in-noise test have similarly shown promising results. Five different smartphones and five different types of headphones were used to administer the test, and no statistically significant differences were found in SRT (Potgieter et al., 2015). Moreover, the test is shown to be suitable for use as a hearing screening test in a multilingual population (Potgieter et al., 2017).

Another interesting development is the upcoming rise of highly accessible and user-friendly smartphone-based pure-tone audiometric testing (Louw et al., 2017). This hearing screening application offers similar advantages as compared to online speech-in-noise screening tests, although the sensitivity for background noise may be expected to be more critical.

For instance, the hearScreen™ application was developed as a low cost alternative to conventional hearing screening, and is highly suitable for the early identification of hearing loss in primary healthcare clinics in developing countries and underserved regions (Swanepoel et al., 2014, Mahomed-Asmail et al., 2016, Louw et al., 2017). Though in contrast with smartphone speech-in-noise testing, this type of smartphone test application does require calibrated headphones and is not self-administered. Moreover, for NIHL awareness and screening purposes in teenagers and adults, speech-in-noise tests may be more suitable, as these are functional tests, and therefore more related to hearing abilities that are important for daily life. However, further research should confirm this assumption.

### *The cost-effectiveness of different hearing screening tests*

Studies that compare the cost-effectiveness of Internet-based speech-in-noise hearing screening tests with other types of hearing screening tests are scarce. A study by Linssen et al. (2015), is the first that assessed the cost-effectiveness of screening adults for hearing loss in the Netherlands, comparing Internet screening (i.e. the National Hearing Test) with telephone screening, screening with a handheld screening device, audiometric screening, and no screening strategy at all. The study concluded that adult hearing screening was cost-effective as compared to no screening strategy. Moreover, domestic Internet screening was the most cost-effective among all strategies, while audiometric screening was the most costly and less effective. Although this study included the assessment of older adults that might benefit from hearing aid fitting by means of the National Hearing test in a home environment, and did not include all possible types of hearing screening strategies (such as OAE), the outcomes of this study seem promising for the use of online hearing screening tests for the secondary prevention of hearing loss, including NIHL as well.

## Implications for future research

This research has contributed to the evidence base needed for the use of online speech-in-noise tests for occupational and recreational NIHL screening. However, more research is suggested, especially focusing on translational research, i.e. integrating population-based audiological research to the broader community in order to improve the quality of hearing surveillance and hearing health on the long term.

Although the focus of this research was on secondary prevention, priority must be given to the primary prevention of NIHL. Internet testing has made hearing screening more accessible and affordable, however it is still important to focus on reducing noise levels at the source (in occupational settings, as well as at music venues and through personal listening devices), and on creating awareness, changing attitudes and norms towards (recreational) noise, and tackling risky behavior through educational materials. Therefore, in general, more longitudinal follow-up studies on excessive noise exposure and the relation with NIHL and overall health are needed, both in the working-age population as well as in young people, in order to properly assess the exact magnitude of this public health risk. More insight into risky noise exposure and the prevalence of NIHL in these populations (and the effects of gender, age, socio-economic status and education level, ear infections, etc.) may be helpful in order to select effective approaches for primary prevention, as well as for secondary prevention, in the form of detecting asymptomatic cases of NIHL.

***Future research concerning Occupational Earcheck***

According to the first part of this research, OEC was found an appropriate screening test for occupational NIHL, and a broad practical implementation of OEC appears to be justified. In order to attain more reliable estimates of the performance and the effectiveness of OEC in certain circumstances, test results collected from practice should be evaluated.

Large-scale and long-term observational population-based pilot studies may be performed. By means of these studies important data from practice can be collected, making it possible to evaluate the sensitivity and specificity in specific occupationally noise-exposed populations. For instance, in current research mostly men were studied, as the study populations were selected from male-dominated industries, while it is known that women also work in noise-exposed settings with high communication demands (e.g. health care sector, educational sector). Moreover, the effects of periodic (i.e. annual) screening may be evaluated, in order to explore the potential of OEC for monitoring (i.e. assessing individual differences in test scores over time) as well.



In order to investigate the added value of OEC relative to existing screening tests such as screening audiometry, OEC should be validly compared to those tests and procedures in practice (and in combination with those tests), focusing on interesting short-term and long term outcomes, such as whether the tests leads to higher participation rates in screening procedures and more awareness that will reduce NIHL at the work floor, and whether this goes along with reduced lifetime and societal costs. In order to perform these comparisons, randomized controlled trials (RCT) of good quality are the most ideal studies, providing the most reliable evidence. However, rigorous RCTs are not commonly performed for evaluating the benefit of medical or screening tests, as they involve large study populations, and are time-consuming.

***Future research concerning Earcheck and the Dutch hearing screening test for children***

In this thesis, prior concerns, such as age and learning effects, and testing options have been explored for online-speech in noise hearing screening tests in young people. However, EC and the Dutch hearing screening test for children are less evaluated as compared to OEC, and should be handled with caution when applied for hearing screening. It is necessary to evaluate both tests in large representative groups of children and teenagers, including those with hearing loss as well, in order to assess the tests' accuracy. For these evaluations it is important to define the type and degree of hearing losses that need to be detected by means of the screening tests, such as NIHL and other sensorineural hearing losses that may affect educational performance.

Finally, the tests that were evaluated in this research were based on the Dutch language. Although the tests are easily accessible, with simple and understandable vocabulary, supported by picture identification, and therefore have small linguistic and lexical constraints, the test remains language dependent. The Dutch hearing screening test may be less suitable for young children who do not have good knowledge of the Dutch language. For pediatric screening purposes it would therefore be interesting to explore the possibilities of a more universal test. Recently, the Sound Earcheck (SEC) has been developed, a language-independent hearing screening test for young children based on the perception of non-speech sounds in noise (Denys et al., 2017). The sensitivity and specificity of the SEC to detect hearing loss, as well as its feasibility in young children, are currently being investigated.

## Considerations for policy and practice

Good quality research to provide an evidence base for test accuracy is a prerequisite to realize the implementation of adequate and effective online hearing screening tests. Therefore, this thesis has some implications for policy and practice, and provide important insights regarding current preventive measures taken by the Dutch government and the occupational sector. The online speech-in-noise screening tests are already being applied and available online, however, some aspects need further consideration. These involve practical aspects, such as uniform guidelines and creating support, but also juridical and ethical aspects, such as data handling.

### *Uniformity*

From this research it becomes clear that it is important to pay attention to various aspects concerning the intended purpose of the test in screening practice, such as the target condition and the target population, the electronic equipment and screening environment, the frequency of testing, and the referral and further hearing assessment after a positive test result. These aspects need to be clearly defined and captured in standardized protocols, so that there are no wide variations in implementation strategies, making uniform screening possible.

### *Support*

For a successful hearing screening program it is important to have consultations with all stakeholders in order to understand and address their concerns and needs. This way support may be created amongst all partners involved (Laplante-Levesque et al., 2016). Hearing health care professionals, occupational companies and physicians, public health authorities, schools, parents, and last but not least the target populations - employees and children and teenagers exposed to occupational or recreational noise - also need proper guidance and motivation to increase participation. For instance, at the moment there is still a gap between a positive attitude towards and willingness to use telehealth and the actual use of it (Eikelboom and Swanepoel, 2016).



## General conclusion

In this thesis newly developed online speech-in-noise self-tests Occupational Earcheck (OEC) and Earcheck (EC) were investigated in order to attain valid, reliable, and suitable NIHL screening tests in the working population exposed to excessive noise, and in young people exposed to recreational noise, both potentially at risk of NIHL.

Concerning OEC, several improvements have been made which have resulted in an appropriate screening test for occupational NIHL that can be used under supervision in an occupational environment. OEC is reasonably well appropriate for detecting mild hearing losses. For moderate hearing losses, the test is more accurate. However, there is still room for improvement and fine-tuning of OEC as a screening test within the working-age population exposed to occupational noise. Also, the effectiveness needs to be explored further, as well as the short- and long-term advantages on overall hearing health in noise-exposed workers.

For children and teenagers online speech-in-noise self-testing can be offered through several devices including smartphones, making innovative testing at school feasible. Age effects and learning effects have been found, and solutions have been suggested to overcome these effects in order to attain more reliable test results. EC and the Dutch hearing test for children are, however, in a less developed stadium as compared to OEC. Further practical evaluation in large and heterogeneous populations, including hearing-impaired children and teenagers, is required in order to establish the accuracy of the test for hearing loss.

In conclusion, many challenges that go along with primary test evaluation studies have been tackled. The expectation is that internet-based speech in noise testing will become increasingly important for screening and monitoring NIHL, though there is still room for further improvement and fine-tuning of these tests.

# References
# List of abbreviations
# Summary
# Samenvatting
# Dankwoord
# CV & Portfolio

# References


AKEROYD, M. A., ARLINGER, S., BENTLER, R. A., BOOTHROYD, A., DILLIER, N., DRESCHLER, W. A., GAGNE, J. P., LUTMAN, M., WOUTERS, J., WONG, L., KOLLMEIER, B. & INTERNATIONAL COLLEGIUM OF REHABILITATIVE AUDIOLOGY WORKING GROUP ON MULTILINGUAL SPEECH, T. 2015. International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests. ICRA Working Group on Multilingual Speech Tests. *Int J Audiol,* 54 Suppl 2**,** 17-22.

ALBRECHT, J., ELEWOUT, L. V., VERHAGE, L. & VERWEIJ, C. 2005. Oorcheck: De validering van een interactief screeningsinstrument. Internal report (in Dutch). Leiden, The Netherlands: LUMC.

ALTMAN, D. G. & BLAND, J. M. 1994a. Diagnostic tests 2: Predictive values. *BMJ,* 309**,** 102.

ALTMAN, D. G. & BLAND, J. M. 1994b. Diagnostic tests 3: receiver operating characteristic plots. *BMJ,* 309**,** 188.

ALTMAN, D. G. & BLAND, J. M. 1994c. Diagnostic tests. 1: Sensitivity and specificity. *BMJ,* 308**,** 1552.

ANDERSON, S. & KRAUS, N. 2010. Sensory-cognitive interaction in the neural encoding of speech in noise: a review. *J Am Acad Audiol,* 21**,** 575-85.

BAMFORD, J., FORTNUM, H., BRISTOW, K., SMITH, J., VAMVAKAS, G., DAVIES, L., TAYLOR, R., WATKIN, P., FONSECA, S., DAVIS, A. & HIND, S. 2007. Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen. *Health Technol Assess,* 11**,** 1-168, iii-iv.

BASNER, M., BABISCH, W., DAVIS, A., BRINK, M., CLARK, C., JANSSEN, S. & STANSFELD, S. 2014. Auditory and non-auditory effects of noise on health. *The Lancet,* 383**,** 1325-1332.

BEACH, E. F., GILLIVER, M. & WILLIAMS, W. 2013. Leisure noise exposure: participation trends, symptoms of hearing damage, and perception of risk. *Int J Audiol,* 52 Suppl 1**,** S20-5.

BESS, F. H., DODD-MURPHY, J. & PARKER, R. A. 1998. Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. *Ear Hear,* 19**,** 339-54.

BOOTHROYD, A. 1997. Auditory development of the hearing child. *Scand Audiol Suppl,* 46**,** 9-16.

BOSMAN, A. 1989. *Speech perception by the hearing impaired.* University of Utrecht.

BROOKHOUSER, P. E. 1994. Prevention of noise-induced hearing loss. *Preventive Medicine* 23**,** 665-669.

BROOKHOUSER, P. E., WORTHINGTON, D. W. & KELLY, W. J. 1991. Unilateral hearing loss in children. *Laryngoscope,* 101**,** 1264-72.

CARTER, L., WILLIAMS, W., BLACK, D. & BUNDY, A. 2014. The leisure-noise dilemma: hearing loss or hearsay? What does the literature tell us? *Ear Hear,* 35**,** 491-505.

CHUNG, J. H., DES ROCHES, C. M., MEUNIER, J. & EAVEY, R. D. 2005. Evaluation of noise-induced hearing loss in young people using a web-based survey technique. *Pediatrics,* 115**,** 861-7.

CORBIN, N. E., BONINO, A. Y., BUSS, E. & LEIBOLD, L. J. 2016. Development of Open-Set Word Recognition in Children: Speech-Shaped Noise and Two-Talker Speech Maskers. *Ear Hear,* 37**,** 55-63.

CULLING, J. F., ZHAO, F. & STEPHENS, D. 2005. The viability of speech-in-noise audiometric screening using domestic audio equipment. *Int J Audiol,* 44**,** 691-700.

DANIEL, E. 2007. Noise and hearing loss: a review. *J Sch Health,* 77**,** 225-31.

DAVIS, J. M., ELFENBEIN, J., SCHUM, R. & BENTLER, R. A. 1986. Effects of mild and moderate hearing impairments on language, educational, and psychosocial behavior of children. *J Speech Hear Disord,* 51**,** 53-62.

DE BOER, J. & THORNTON, A. R. 2008. Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J Neurosci,* 28**,** 4929-37.

DE LAAT, J., VAN DEELEN, L. & WIEFFERINK, K. 2016. Hearing Screening and Prevention of Hearing Loss in Adolescents. *J Adolesc Health,* 59**,** 243-245.

DENYS, S., HOFFMAN, M., WOUTERS, J., RASHID, M. S., SELS, L., DRESCHLER, W. & LAAT, J. D. 2017. Hearing screening based on sound perception in noise. KU Leuven.

DILLON, H., BEACH, E. F., SEYMOUR, J., CARTER, L. & GOLDING, M. 2016. Development of Telscreen: a telephone-based speech-in-noise hearing screening test with a novel masking noise and scoring procedure. *Int J Audiol,* 55**,** 463-71.

DOBIE, R. A. 2007. Noise-induced permanent threshold shifts in the occupational noise and hearing survey: an explanation for elevated risk estimates. *Ear Hear,* 28**,** 580-91.

EGGERMONT, J. J. & PONTON, C. W. 2003. Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. *Acta Otolaryngol,* 123**,** 249-52.

EIKELBOOM, R. H. & SWANEPOEL, W. 2016. International Survey of Audiologists' Attitudes Toward Telehealth. *Am J Audiol,* 25**,** 295-298.

EISENBERG, L. S., SHANNON, R. V., MARTINEZ, A. S., WYGONSKI, J. & BOOTHROYD, A. 2000. Speech recognition with reduced spectral cues as a function of age. *J Acoust Soc Am,* 107**,** 2704-10.

ELLIOT, L. L. 1979. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability *Journal of the Acoustical Society of America,* 66**,** 651-653.

ELLIOTT, L. L., CONNORS, S., KILLE, E., LEVIN, S., BALL, K. & KATZ, D. 1979. Children's understanding of monosyllabic nouns in quiet and in noise. *J Acoust Soc Am,* 66**,** 12-21.

ELLIS, N., KUIPERS, M., VAN DER PIJL, S. & VERBIEST, E. 2006. Ontwikkeling van een gehoorscreening voor het bedrijfsleven. Internal report (in Dutch). Leiden, The Netherlands: LUMC.

EUROPEAN PARLIAMENT, C. O. T. E. U. 2003. Seventeenth individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC. *Directive 2003/10/EC of the European Parliament and of the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise).*

EYSINK, P. E. D., BLATTER, B. M., VAN GOOL, C. H., GOMMER, A. M., VAN DEN BOSSCHE, S. N. J. & HOEYMANS, N. 2007. Ziektelast van ongunstige arbeidsomstandigheden in Nederland Bilthoven: RIVM.

FALLON, M., TREHUB, S. E. & SCHNEIDER, B. A. 2000. Children's perception of speech in multitalker babble. *J Acoust Soc Am,* 108**,** 3023-9.

FEDER, K., MARRO, L., KEITH, S. E. & MICHAUD, D. S. 2013. Audiometric thresholds and portable digital audio player user listening habits. *Int J Audiol,* 52**,** 606-16.

FLAMME, G. A., STEPHENSON, M. R., DEITERS, K. K., HESSENAUER, A., VANGESSEL, D. K., GEDA, K., WYLLYS, K. & MCGREGOR, K. D. 2014. Short-term variability of pure-tone thresholds obtained with TDH-39P earphones. *Int J Audiol,* 53 Suppl 2**,** S5-15.

FLIGOR, B. J. & COX, L. C. 2004. Output levels of commercially available portable compact disc players and the potential risk to hearing. *Ear and Hearing,* 25**,** 513-527.

FORTNUM, H. M., SUMMERFIELD, A. Q., MARSHALL, D. H., DAVIS, A. C. & BAMFORD, J. M. 2001. Prevalence of permanent childhood hearing impairment in the United Kingdom and implications for universal neonatal hearing screening: questionnaire based ascertainment study. *Bmj,* 323**,** 536-40.

FREDRIKSSON, S., HAMMAR, O., MAGNUSSON, L., KAHARI, K. & PERSSON WAYE, K. 2016. Validating self-reporting of hearing-related symptoms against pure-tone audiometry, otoacoustic emission, and speech audiometry. *Int J Audiol,* 55**,** 454-62.

GILLIVER, M., BEACH, E. F. & WILLIAMS, W. 2015. Changing beliefs about leisure noise: using health promotion models to investigate young people's engagement with, and attitudes towards, hearing health. *Int J Audiol,* 54**,** 211-9.

GOGTAY, N., GIEDD, J. N., LUSK, L., HAYASHI, K. M., GREENSTEIN, D., VAITUZIS, A. C., NUGENT, T. F., 3RD, HERMAN, D. H., CLASEN, L. S., TOGA, A. W., RAPOPORT, J. L. & THOMPSON, P. M. 2004. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A,* 101**,** 8174-9.

GROBBEE, D. E. & HOES, A. W. 2009. Diagnostic Research. *Clinical Epidemiology. Principles, Methods, and Applications for Clinical Research.* Sudbury, Massachusetts: Jones and Bartlett Publishers.

HALL, J. W., 3RD, BUSS, E. & GROSE, J. H. 2016. Factors affecting the development of speech recognition in steady and modulated noise. *J Acoust Soc Am,* 139**,** 2964.

HELLEMAN, H. W., JANSEN, E. J. & DRESCHLER, W. A. 2010. Otoacoustic emissions in a hearing conservation program: general applicability in longitudinal monitoring and the relation to changes in pure-tone thresholds. *Int J Audiol,* 49**,** 410-9.

HENDERSON, E., TESTA, M. A. & HARTNICK, C. 2011. Prevalence of noise-induced hearing-threshold shifts and hearing loss among US youths. *Pediatrics,* 127**,** e39-46.

HENRY, P. & FOOTS, A. 2012. Comparison of user volume control settings for portable music players with three earphone configurations in quiet and noisy environments. *J Am Acad Audiol,* 23**,** 182-91.

HNATH-CHISOLM, T. E., LAIPPLY, E. & BOOTHROYD, A. 1998. Age-related changes on a children's test of sensory-level speech perception capacity. *J Speech Lang Hear Res,* 41**,** 94-106.

HONG, O., KERR, M. J., POLING, G. L. & DHAR, S. 2013. Understanding and preventing noise-induced hearing loss. *Dis Mon,* 59**,** 110-8.

HONG, O., RONIS, D. L. & ANTONAKOS, C. L. 2011. Validity of self-rated hearing compared with audiometric measurement among construction workers. *Nurs Res,* 60**,** 326-32.

HORNICKEL, J., LIN, D. & KRAUS, N. 2013. Speech-evoked auditory brainstem responses reflect familial and cognitive influences. *Dev Sci,* 16**,** 101-10.

HSU, T. Y., WU, C. C., CHANG, J. G., LEE, S. Y. & HSU, C. J. 2013. Determinants of bilateral audiometric notches in noise-induced hearing loss. *Laryngoscope,* 123**,** 1005-10.

HUNTER, A. 2018. "There are more important things to worry about": attitudes and behaviours towards leisure noise and use of hearing protection in young adults. *Int J Audiol***,** 1-8.

ISO, I. O. F. S. 1990. ISO 1999:1990. *Acoustics: Determination of occupational noise exposure and estimation of noise-induced hearing impairment.* Geneva.

JANSEN, S. 2013. *Efficient and sensitive hearing assessment based on speech perception.* KU Leuven.

JANSEN, S., KONING, R., WOUTERS, J. & VAN WIERINGEN, A. 2014a. Development and validation of the Leuven intelligibility sentence test with male speaker (LIST-m). *Int J Audiol,* 53**,** 55-9.

JANSEN, S., LUTS, H., DEJONCKERE, P., VAN WIERINGEN, A. & WOUTERS, J. 2013. Efficient hearing screening in noise-exposed listeners using the digit triplet test. *Ear Hear,* 34**,** 773-8.

JANSEN, S., LUTS, H., DEJONCKERE, P., VAN WIERINGEN, A. & WOUTERS, J. 2014b. Exploring the sensitivity of speech-in-noise tests for noise-induced hearing loss. *Int J Audiol,* 53**,** 199-205.

JANSEN, S., LUTS, H., WAGENER, K. C., FRACHET, B. & WOUTERS, J. 2010. The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *Int J Audiol,* 49**,** 378-87.

JELLEMA, I. J. 2014. Preventie van gehoorschade op de werkvloer (In Dutch). Den Haag, The Netherlands: Nationale Hoorstichting.

JIN, S. H., NELSON, P. B., SCHLAUCH, R. S. & CARNEY, E. 2013. Hearing conservation program for marching band members: a risk for noise-induced hearing loss? *Am J Audiol,* 22**,** 26-39.

JOHNSON, C. E. 2000. Children's phoneme identification in reverberation and noise. *J Speech Lang Hear Res,* 43**,** 144-57.

KAANDORP, M. W., DE GROOT, A. M., FESTEN, J. M., SMITS, C. & GOVERTS, S. T. 2016. The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *Int J Audiol,* 55**,** 157-67.

KAM, A. C., GAO, H., LI, L. K., ZHAO, H., QIU, S. & TONG, M. C. 2013. Automated hearing screening for children: A pilot study in China. *Int J Audiol,* 52**,** 855-60.

KEITH, S. E., MICHAUD, D. S. & CHIU, V. 2008. Evaluating the maximum playback sound levels from portable digital audio players. *J Acoust Soc Am,* 123**,** 4227-37.

KEPPLER, H., DHOOGE, I. & VINCK, B. 2015. Hearing in young adults. Part II: The effects of recreational noise exposure. *Noise Health,* 17**,** 245-52.

KNECHT, H. A., NELSON, P. B., WHITELAW, G. M. & FETH, L. L. 2002. Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *Am J Audiol,* 11**,** 65-71.

KNOBEL, K. A. & LIMA, M. C. 2012. Knowledge, habits, preferences, and protective behavior in relation to loud sound exposures among Brazilian children. *Int J Audiol,* 51 Suppl 1**,** S12-9.

KOLLMEIER, B., AKEROYD, M. A., ARLINGER, S., BENTLE, R. A., BOOTHROYD, A., DILLIER, N., DRESCHLER, W. A., GAGNE, J. P., LUTMAN, M. E., WOUTERS, J. & WONG, L. 2016. Reply to the Comment on 'International Collegium of Rehabilitative Audiology (ICRA) recommendations for the constructor of multilingual speech test by Akeroyd et al' by Cas Smits. *Int J Audiol,* 55**,** 269-71.

KOOLE, A., NAGTEGAAL, A. P., HOMANS, N. C., HOFMAN, A., BAATENBURG DE JONG, R. J. & GOEDEGEBURE, A. 2016. Using the Digits-In-Noise Test to Estimate Age-Related Hearing Loss. *Ear Hear,* 37**,** 508-13.

KOOPMANS, W. G., T. & SMITS, C. 2015. Spraakverstaan in complexe luistersituaties bij basisschoolkinderen (In Dutch). Amsterdam, The Netherlands: VU Medisch Centrum.

KRAMER, S. E., KAPTEYN, T. S. & FESTEN, J. M. 1998. The self-reported handicapping effect of hearing disabilities. *Audiology,* 37**,** 302-312.

KRAUS, N., STRAIT, D. L. & PARBERY-CLARK, A. 2012. Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Ann N Y Acad Sci,* 1252**,** 100-7.

KRIZMAN, J., SKOE, E. & KRAUS, N. 2016. Bilingual enhancements have no socioeconomic boundaries. *Dev Sci,* 19**,** 881-891.

KRIZMAN, J., TIERNEY, A., FITZROY, A. B., SKOE, E., AMAR, J. & KRAUS, N. 2015. Continued maturation of auditory brainstem function during adolescence: A longitudinal approach. *Clin Neurophysiol,* 126**,** 2348-55.

LAMORÉ 2012. 7.2.2.4 (2). Patho-Anatomie en Pathofysiologie. Binnenoor 1. Perceptief gehoorverlies - Uitval van haarcellen en zenuwvezels. *Audiologieboek*.

LAPLANTE-LEVESQUE, A., LUNNER, T., ANDERSSON, G. & PREMINGER, J. E. 2016. Internet and Audiology: A Review of the Second International Meeting. *Am J Audiol,* 25**,** 257-259.

LAPSLEY MILLER, J. A., MARSHALL, L., HELLER, L. M. & HUGHES, L. M. 2006. Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss. *J Acoust Soc Am,* 120**,** 280-96.

LE, T. N., STRAATMAN, L. V., LEA, J. & WESTERBERG, B. 2017. Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. *J Otolaryngol Head Neck Surg,* 46**,** 41.

LEEFLANG, M. 2008. *Systematic reviews of diagnostic test accuracy.* Universiteit van Amsterdam.

LEEFLANG, M. M., BOSSUYT, P. M. & IRWIG, L. 2009. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol,* 62**,** 5-12.

LEENSEN, M. C. 2013. *Noise Induced Hearing Loss: Screening with Pure Tone Audiometry and Speech-in-noise Testing.*, Faculty of Medicine, University of Amsterdam.

LEENSEN, M. C., DE LAAT, J. A. & DRESCHLER, W. A. 2011a. Speech-in-noise screening tests by internet, part 1: test evaluation for noise-induced hearing loss identification. *Int J Audiol,* 50**,** 823-34.

LEENSEN, M. C., DE LAAT, J. A., SNIK, A. F. & DRESCHLER, W. A. 2011b. Speech-in-noise screening tests by internet, part 2: improving test sensitivity for noise-induced hearing loss. *Int J Audiol,* 50**,** 835-48.

LEENSEN, M. C. & DRESCHLER, W. A. 2013a. The applicability of a speech-in-noise screening test in occupational hearing conservation. *Int J Audiol,* 52**,** 455-65.

LEENSEN, M. C. & DRESCHLER, W. A. 2013b. Speech-in-noise screening tests by internet, part 3: test sensitivity for uncontrolled parameters in domestic usage. *Int J Audiol,* 52**,** 658-69.

LEVEY, S., LEVEY, T. & FLIGOR, B. J. 2011. Noise exposure estimates of urban MP3 player users. *J Speech Lang Hear Res,* 54**,** 263-77.

LINSSEN, A. M., ANTEUNIS, L. J. & JOORE, M. A. 2015. The Cost-Effectiveness of Different Hearing Screening Strategies for 50- to 70-Year-Old Adults: A Markov Model. *Value Health,* 18**,** 560-9.

LITOVSKY, R. 2015. Development of the auditory system. *Handb Clin Neurol,* 129**,** 55-72.

LOUW, C., SWANEPOEL, W., EIKELBOOM, R. H. & MYBURGH, H. C. 2017. Smartphone-Based Hearing Screening at Primary Health Care Clinics. *Ear Hear,* 38**,** e93-e100.

LU, J., HUANG, Z., MA, Y., LI, Y., MEI, L., YAO, G., WANG, Y., SHEN, X. & WU, H. 2014. Comparison between hearing screening-detected cases and sporadic cases of delayed-onset hearing loss in preschool-age children. *Int J Audiol,* 53**,** 229-34.

MACKENBACH, J. P. & VAN DER MAAS, P. J. 2005. Deel 2: Preventie, 3. Theorie en methoden van preventie. *Volksgezondheid en gezondheidszorg.* Maarssen: Elsevier Gezondheidszorg.

MAHAJAN, Y. & MCARTHUR, G. 2012. Maturation of auditory event-related potentials across adolescence. *Hear Res,* 294**,** 82-94.

MAHOMED-ASMAIL, F., SWANEPOEL DE, W., EIKELBOOM, R. H., MYBURGH, H. C. & HALL, J., 3RD 2016. Clinical Validity of hearScreen Smartphone Hearing Screening for School Children. *Ear Hear,* 37**,** e11-7.

MAY, J. J. 2000. Occupational hearing loss. *Am J Ind Med,* 37**,** 112-20.

MCNEALY, K., MAZZIOTTA, J. C. & DAPRETTO, M. 2010. The neural basis of speech parsing in children and adults. *Dev Sci,* 13**,** 385-406.

MEDINA, V., HOONHORST, I., BOGLIOTTI, C. & SERNICLAES, W. 2010. Development of voicing perception in French Comparing adults, adolescents, and children. *Journal of Phonetics,* 38**,** 493-503.

MEYER-BISCH, C. 1996. Epidemiological evaluation of hearing damage related to strongly amplified music (personal cassette players, discotheques, rock concerts)--high-definition audiometric survey on 1364 subjects. *Audiology,* 35**,** 121-42.

MILLER, J. D. 1974. Effects of noise on people. *J Acoust Soc Am,* 56**,** 729-64.

MOLANDER, P., NORDQVIST, P., OBERG, M., LUNNER, T., LYXELL, B. & ANDERSSON, G. 2013. Internet-based hearing screening using speech-in-noise: validation and comparisons of self-reported hearing problems, quality of life and phonological representation. *BMJ Open,* 3**,** e003223.

MOON, I. J., WON, J. H., PARK, M. H., IVES, D. T., NIE, K., HEINZ, M. G., LORENZI, C. & RUBINSTEIN, J. T. 2014. Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. *J Neurosci,* 34**,** 12145-54.

MOONS, K. G., VAN ES, G. A., DECKERS, J. W., HABBEMA, J. D. & GROBBEE, D. E. 1997. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology,* 8**,** 12-7.

MOONS, K. G. M., BIESHEUVEL, C. J. & GROBBEE, D. E. 2004. Test research versus diagnostic research. *Clinical Chemistry,* 50**,** 473-476.

MOSITES, E., NEITZEL, R., GALUSHA, D., TRUFAN, S., DIXON-ERNST, C. & RABINOWITZ, P. 2016. A comparison of an audiometric screening survey with an in-depth research questionnaire for hearing loss and hearing loss risk factors. *Int J Audiol,* 55**,** 782-786.

MOSTAFAPOUR, S. P., LAHARGOUE, K. & GATES, G. A. 1998. Noise-induced hearing loss in young adults: the role of personal listening devices and other sources of leisure noise. *Laryngoscope,* 108**,** 1832-9.

MUCHNIK, C., AMIR, N., SHABTAI, E. & KAPLAN-NEEMAN, R. 2012. Preferred listening levels of personal listening devices in young teenagers: self reports and physical measurements. *Int J Audiol,* 51**,** 287-93.

NACHTEGAAL, J., SMIT, J. H., SMITS, C., BEZEMER, P. D., VAN BEEK, J. H., FESTEN, J. M. & KRAMER, S. E. 2009. The association between hearing status and psychosocial health before the age of 70 years: results from an internet-based national survey on hearing. *Ear Hear,* 30**,** 302-12.

NCJ, N. C. J. 2016. Richtlijn Vroegtijdige opsporing van gehoorverlies bij kinderen en jongeren (0-18 jaar). Utrecht.

NELSON, D. I., NELSON, R. Y., CONCHA-BARRIENTOS, M. & FINGERHUT, M. 2005. The global burden of occupational noise-induced hearing loss. *Am J Ind Med,* 48**,** 446-58.

NEUMAN, A. C., WROBLEWSKI, M., HAJICEK, J. & RUBINSTEIN, A. 2010. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear Hear,* 31**,** 336-44.

NIOSH, N. I. F. O. S. A. H. 1998. Criteria for a recommended standard: occupational noise exposure, revised criteria. Cincinatti, Ohio.

NISKAR, A. S., KIESZAK, S. M., HOLMES, A., ESTEBAN, E., RUBIN, C. & BRODY, D. J. 1998. Prevalence of hearing loss among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey. *JAMA,* 279**,** 1071-5.

NISKAR, A. S., KIESZAK, S. M., HOLMES, A. E., ESTEBAN, E., RUBIN, C. & BRODY, D. J. 2001a. Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: The third national health and nutrition examination survey, 1988-1994, United States. *Pediatrics* 108**,** 40-43.

NISKAR, A. S., KIESZAK, S. M., HOLMES, A. E., ESTEBAN, E., RUBIN, C. & BRODY, D. J. 2001b. Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey, 1988-1994, United States. *Pediatrics,* 108**,** 40-3.

PAGLIALONGA, A., PINCIROLI, F. & TOGNOLA, G. 2017. The ALFA4Hearing Model (At-a-Glance Labeling for Features of Apps for Hearing Health Care) to Characterize Mobile Apps for Hearing Health Care. *Am J Audiol,* 26**,** 408-425.

PAGLIALONGA, A., TOGNOLA, G. & GRANDORI, F. 2014. A user-operated test of suprathreshold acuity in noise for adult hearing screening: The SUN (Speech Understanding in Noise) test. *Comput Biol Med,* 52**,** 66-72.

PARBERY-CLARK, A., MARMEL, F., BAIR, J. & KRAUS, N. 2011. What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci,* 33**,** 549-57.

PARIKH, R., PARIKH, S., ARUN, E. & THOMAS, R. 2009. Likelihood ratios: clinical application in day-to-day practice. *Indian J Ophthalmol,* 57**,** 217-21.

PLACK, C. J., LEGER, A., PRENDERGAST, G., KLUK, K., GUEST, H. & MUNRO, K. J. 2016. Toward a Diagnostic Test for Hidden Hearing Loss. *Trends Hear,* 20**,** 1-9.

PLOMP, R. 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Hear Res,* 29**,** 146-54.

PLOMP, R. & MIMPEN, A. M. 1979a. Improving the reliability of testing the speech reception threshold for sentences. *Audiology,* 18**,** 43-52.

PLOMP, R. & MIMPEN, A. M. 1979b. Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am,* 66**,** 1333-42.

PONTON, C. W., EGGERMONT, J. J., KWONG, B. & DON, M. 2000. Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clin Neurophysiol,* 111**,** 220-36.

PORTNUFF, C. D., FLIGOR, B. J. & AREHART, K. H. 2011. Teenage use of portable listening devices: a hazard to hearing? *J Am Acad Audiol,* 22**,** 663-77.

PORTNUFF, C. D., FLIGOR, B. J. & AREHART, K. H. 2013. Self-report and long-term field measures of MP3 player use: how accurate is self-report? *Int J Audiol,* 52 Suppl 1**,** S33-40.

POTGIETER, J. M., SWANEPOEL DE, W., MYBURGH, H. C., HOPPER, T. C. & SMITS, C. 2015. Development and validation of a smartphone-based digits-in-noise hearing test in South African English. *Int J Audiol,* 55**,** 405-411.

POTGIETER, J. M., SWANEPOEL, W., MYBURGH, H. C. & SMITS, C. 2017. The South African English Smartphone Digits-in-Noise Hearing Test: Effect of Age, Hearing Loss, and Speaking Competence. *Ear Hear***,** 1-8.

PRIEVE, B. A., SCHOOLING, T., VENEDIKTOV, R. & FRANCESCHINI, N. 2015. An Evidence-Based Systematic Review on the Diagnostic Accuracy of Hearing Screening Instruments for Preschool- and School-Age Children. *Am J Audiol,* 24**,** 250-67.

PUNCH, J. L., ELFENBEIN, J. L. & JAMES, R. R. 2011. Targeting hearing health messages for users of personal listening devices. *Am J Audiol,* 20**,** 69-82.

QUINTANILLA-DIECK MDE, L., ARTUNDUAGA, M. A. & EAVEY, R. D. 2009. Intentional exposure to loud music: the second MTV.com survey reveals an opportunity to educate. *J Pediatr,* 155**,** 550-5.

RABINOWITZ, P., CANTLEY, L. F., GALUSHA, D., TRUFAN, S., SWERSEY, A., DIXON-ERNST, C., RAMIREZ, V. & NEITZEL, R. 2018. Assessing Hearing Conservation Program Effectiveness: Results of a Multisite Assessment. *J Occup Environ Med,* 60**,** 29-35.

RABINOWITZ, P. M., GALUSHA, D., SLADE, M. D., DIXON-ERNST, C., SIRCAR, K. D. & DOBIE, R. A. 2006a. Audiogram notches in noise-exposed workers. *Ear Hear,* 27**,** 742-50.

RABINOWITZ, P. M., SLADE, M. D., GALUSHA, D., DIXON-ERNST, C. & CULLEN, M. R. 2006b. Trends in the prevalence of hearing loss among young adults entering an industrial workforce 1985 to 2004. *Ear Hear,* 27**,** 369-75.

RASHID, M. S., LEENSEN, M. C. & DRESCHLER, W. A. 2016. Application of the online hearing screening test "Earcheck": Speech intelligibility in noise in teenagers and young adults. *Noise Health,* 18**,** 312-318.

ROSE, G. 2001. Sick individuals and sick populations. *Int J Epidemiol,* 30**,** 427-32; discussion 433-4.

ROSE, G. 2008. Commentary by Kay-Tee Khaw and Mchael Marmot. *Rose's strategy of preventive medicine.* New York: Oxford University Press Inc.

RYAN, A. F., KUJAWA, S. G., HAMMILL, T., LE PRELL, C. & KIL, J. 2016. Temporary and Permanent Noise-induced Threshold Shifts: A Review of Basic and Clinical Observations. *Otol Neurotol,* 37**,** e271-5.

SAMELLI, A. G., RABELO, C. M., PEREIRA, M. B., PORTELA, M. N., SANCHES, S. G. & NEVES-LOBO, I. F. 2012. Comparison of screening methods for conductive hearing loss identification in children: low-cost proposal. *J Med Screen,* 19**,** 1-7.

SCHAFER, E. C., BEELER, S., RAMOS, H., MORAIS, M., MONZINGO, J. & ALGIER, K. 2012. Developmental effects and spatial hearing in young children with normal-hearing sensitivity. *Ear Hear,* 33**,** e32-43.

SCHLAUCH, R. S. & CARNEY, E. 2011. Are false-positive rates leading to an overestimation of noise-induced hearing loss? *J Speech Lang Hear Res,* 54**,** 679-92.

SCHLAUCH, R. S. & CARNEY, E. 2012. The challenge of detecting minimal hearing loss in audiometric surveys. *Am J Audiol,* 21**,** 106-19.

SEKHAR, D. L., ZALEWSKI, T. R., BEILER, J. S., CZARNECKI, B., BARR, A. L., KING, T. S. & PAUL, I. M. 2016. The Sensitivity of Adolescent School-Based Hearing Screens Is Significantly Improved by Adding High Frequencies. *J Sch Nurs,* 32**,** 416-422.

SERRA, M. R., BIASSONI, E. C., RICHTER, U., MINOLDO, G., FRANCO, G., ABRAHAM, S., CARIGNANI, J. A., JOEKES, S. & YACCI, M. R. 2005. Recreational noise exposure and its effects on the hearing of adolescents. Part I: an interdisciplinary long-term study. *Int J Audiol,* 44**,** 65-73.

SHARGORODSKY, J., CURHAN, S. G., CURHAN, G. C. & EAVEY, R. 2010. Change in prevalence of hearing loss in US adolescents. *JAMA,* 304**,** 772-8.

SHEIKH RASHID, M. & DRESCHLER, W. A. 2014. Evaluatie Kinderhoortest (In Dutch). Amsterdam, The Netherlands: AMC.

SHEIKH RASHID, M., DRESCHLER, W. A. & DE LAAT, J. 2017a. Evaluation of an internet-based speech-in-noise screening test for school-age children. *Int J Audiol,* 56**,** 967-975.

SHEIKH RASHID, M., LEENSEN, M. C., DE LAAT, J. A. & DRESCHLER, W. A. 2017b. Cross-sectional evaluation of an internet-based hearing screening test in an occupational setting. *Scand J Work Environ Health,* 43**,** 279-286.

SHEIKH RASHID, M., LEENSEN, M. C. J., DE LAAT, J. & DRESCHLER, W. A. 2017c. Laboratory evaluation of an optimised internet-based speech-in-noise test for occupational high-frequency hearing loss screening: Occupational Earcheck. *Int J Audiol,* 56**,** 844-853.

SKARZYNSKI, H. & PIOTROWSKA, A. 2012. Screening for pre-school and school-age hearing problems: European Consensus Statement. *Int J Pediatr Otorhinolaryngol,* 76**,** 120-1.

SKOE, E., KRIZMAN, J., ANDERSON, S. & KRAUS, N. 2015. Stability and plasticity of auditory brainstem function across the lifespan. *Cereb Cortex,* 25**,** 1415-26.

SKOE, E., KRIZMAN, J., SPITZER, E. & KRAUS, N. 2013. The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience,* 243**,** 104-14.

SLATER, J., SKOE, E., STRAIT, D. L., O'CONNELL, S., THOMPSON, E. & KRAUS, N. 2015. Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behav Brain Res,* 291**,** 244-52.

SLIWINSKA-KOWALSKA, M. & DAVIS, A. 2012. Noise-induced hearing loss. *Noise Health,* 14**,** 274-80.

ŚLIWIŃSKA-KOWALSKA, M., DUDAREWICZ, A., KOTYŁO, P., ZAMYSŁOWSKA-SZMYTKE, E., PAWLACZYK-ŁUSZCZYŃSKA, M. & GAJDA-SZADKOWSKA, A. 2006. Individual Susceptibility to Noise-Induced Hearing Loss: Choosing an Optimal Method of Retrospective Classification of Workers into Noise-Susceptible and Noise-Resistant Groups. *International Journal of Occupational Medicine and Environmental Health,* 19.

SMITS, C. 2005. *Hearing screening by telephone. Fundamentals & applications.* VU University Medical Center.

SMITS, C. & HOUTGAST, T. 2005. Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear,* 26**,** 89-95.

SMITS, C., KAPTEYN, T. S. & HOUTGAST, T. 2004. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol,* 43**,** 15-28.

SMITS, C., MERKUS, P. & HOUTGAST, T. 2006. How we do it: The Dutch functional hearing-screening tests by telephone and internet. *Clin Otolaryngol,* 31**,** 436-40.

SMITS, C., THEO GOVERTS, S. & FESTEN, J. M. 2013. The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J Acoust Soc Am,* 133**,** 1693-706.

SMITS, C., WATSON, C. S., KIDD, G. R., MOORE, D. R. & GOVERTS, S. T. 2016. A comparison between the Dutch and American-English digits-in-noise (DIN) tests in normal-hearing listeners. *Int J Audiol,* 55**,** 358-65.

SMOORENBURG, G. F. 1990. On the limited transfer of information with noise-induced hearing loss. *Acta Otolaryngol Suppl,* 469**,** 38-46.

SMOORENBURG, G. F. 1992. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *Journal of the Acoustical Society of America,* 91**,** 421-437.

SORGDRAGER, B., DOORNBUSCH, J., VAN GAALEN-MULDER, P., HOORENMAN, T. A. M., DE JAGER, H. J., MARINUS, E., NIJBOER, A. P., PUTMNAN, J. & TIESSIENK, W. 2006. Multidisciplinaire richtlijn: Preventie beroepsslechthorendheid (In Dutch). Utrecht, The Netherlands: NVAB.

STAM, M., SMITS, C., TWISK, J. W., LEMKE, U., FESTEN, J. M. & KRAMER, S. E. 2015. Deterioration of Speech Recognition Ability Over a Period of 5 Years in Adults Ages 18 to 70 Years: Results of the Dutch Online Speech-in-Noise Test. *Ear Hear,* 36**,** e129-37.

STENFELT, S., JANSSEN, T., SCHIRKONYER, V. & GRANDORI, F. 2011. e-Health Technologies for Adult Hearing Screening. *Audiol Res,* 1**,** e14.

STRAIT, D. L., SLATER, J., ABECASSIS, V. & KRAUS, N. 2014. Cortical response variability as a developmental index of selective auditory attention. *Dev Sci,* 17**,** 175-86.

STUART, A. 2005. Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear Hear,* 26**,** 78-88.

SU, B. M. & CHAN, D. K. 2017. Prevalence of Hearing Loss in US Children and Adolescents: Findings From NHANES 1988-2010. *JAMA Otolaryngol Head Neck Surg,* 143**,** 920-927.

SUSSMAN, E. & STEINSCHNEIDER, M. 2009. Attention effects on auditory scene analysis in children. *Neuropsychologia,* 47**,** 771-85.

SUSSMAN, J. E. 1993. Auditory processing in children's speech perception: results of selective adaptation and discrimination tasks. *J Speech Hear Res,* 36**,** 380-95.

SWANEPOEL, D., W., MYBURGH, H. C., HOWE, D. M., MAHOMED, F. & EIKELBOOM, R. H. 2014. Smartphone hearing screening with integrated quality control and data management. *Int J Audiol,* 53**,** 841-9.

TALARICO, M., ABDILLA, G., ALIFERIS, M., BALAZIC, I., GIAPRAKIS, I., STEFANAKIS, T., FOENANDER, K., GRAYDEN, D. B. & PAOLINI, A. G. 2007. Effect of age and cognition on childhood speech in noise perception abilities. *Audiol Neurootol,* 12**,** 13-9.

TIERNEY, A. T., KRIZMAN, J. & KRAUS, N. 2015. Music training alters the course of adolescent auditory development. *Proc Natl Acad Sci U S A,* 112**,** 10062-7.

TWARDELLA, D., RAAB, U., PEREZ-ALVAREZ, C., STEFFENS, T., BOLTE, G. & FROMME, H. 2017. Usage of personal music players in adolescents and its association with noise-induced hearing loss: A cross-sectional analysis of Ohrkan cohort study data. *Int J Audiol,* 56**,** 38-45.

VAEZ, N., DESGUALDO-PEREIRA, L. & PAGLIALONGA, A. 2014. Development of a test of suprathreshold acuity in noise in Brazilian Portuguese: a new method for hearing screening and surveillance. *Biomed Res Int,* 2014**,** 652838.

VAILLANCOURT, V., LAROCHE, C., GIGUERE, C. & SOLI, S. D. 2008. Establishment of age-specific normative data for the canadian French version of the hearing in noise test for children. *Ear Hear,* 29**,** 453-66.

VAN DER MOLEN, H., KUIJER, P., DE GROENE, G., BAKKER, J., SORGDRAGER, B., LENDERINK, A., MAAS, J., VAN VALEN, E. & BRAND, T. 2014. Beroepsziekten in cijfers 2014 (in Dutch). Amsterdam, The Netherlands: Nederlands Centrum voor Beroepsziekten, Coronel Instituut voor Arbeid en Gezondheid, AMC, UvA.

VAN DER MOLEN, H., KUIJER, P., DE GROENE, G., BAKKER, J., SORGDRAGER, B., LENDERINK, A., MAAS, J., VAN VALEN, E. & BRAND, T. 2016a. Beroepsziekten in Cijfers 2016. Amsterdam, The Netherlands: Nederlands Centrum voor Beroepsziekten, Coronel Instituut voor Arbeid en Gezondheid, AMC, UvA.

VAN DER MOLEN, H. F., DE VRIES, S. C., STOCKS, S. J., WARNING, J. & FRINGS-DRESEN, M. H. 2016b. Incidence rates of occupational diseases in the Dutch construction sector, 2010-2014. *Occup Environ Med,* 73**,** 350-2.

VAN RIJN, M. J. 2017. Tweede voortgangsrapportage Actieplan Preventie Gehoorschade. *In:* MINISTERIE VAN VOLKSGEZONDHEID, W. E. S. (ed.). Den Haag.

VERBEEK, J. H., KATEMAN, E., MORATA, T. C., DRESCHLER, W. A. & MISCHKE, C. 2014. Interventions to prevent occupational noise-induced hearing loss: a Cochrane systematic review. *Int J Audiol,* 53 Suppl 2**,** S84-96.

VERCAMMEN, C., GOOSSENS, T., WOUTERS, J. & VAN WIERINGEN, A. 2017. Digit Triplet Test Hearing Screening With Broadband and Low-Pass Filtered Noise in a Middle-Aged Population. *Ear Hear***,** 1-4.

VINETTE, S. A. & BRAY, S. 2015. Variation in functional connectivity along anterior-to-posterior intraparietal sulcus, and relationship with age across late childhood and adolescence. *Dev Cogn Neurosci,* 13**,** 32-42.

VLAMING, M. S., MACKINNON, R. C., JANSEN, M. & MOORE, D. R. 2014. Automated screening for high-frequency hearing loss. *Ear Hear,* 35**,** 667-79.

VOGEL, I. 2009. *Music-Listening Behavior of Adolescents and Hearing Conservation: many risks, few precautions*.

VOGEL, I., BRUG, J., HOSLI, E. J., VAN DER PLOEG, C. P. & RAAT, H. 2008. MP3 players and hearing loss: adolescents' perceptions of loud music and hearing conservation. *J Pediatr,* 152**,** 400-4.

WAGENER, K. C. & BRAND, T. 2005. Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. *Int J Audiol,* 44**,** 144-56.

WATSON, C. S., KIDD, G. R., MILLER, J. D., SMITS, C. & HUMES, L. E. 2012. Telephone screening tests for functionally impaired hearing: current use in seven countries and development of a US version. *J Am Acad Audiol,* 23**,** 757-67.

WESTERHAUSEN, R., BLESS, J. J., PASSOW, S., KOMPUS, K. & HUGDAHL, K. 2015. Cognitive control of speech perception across the lifespan: A large-scale cross-sectional dichotic listening study. *Dev Psychol,* 51**,** 806-15.

WHO, W. H. O. 2012. *WHO global estimates on prevalence of hearing loss* [Online]. Geneva. Available: http://www.who.int/pbd/deafness/WHO_GE_HL.pdf [Accessed].

WHO, W. H. O. March 2015. *Deafness and hearing loss; Fact Sheet No 300* [Online]. Geneva. Available: http://www.who.int/mediacentre/factsheets/fs300/en/ [Accessed].

WIGHTMAN, F., KISTLER, D. & BRUNGART, D. 2006. Informational masking of speech in children: auditory-visual integration. *J Acoust Soc Am,* 119**,** 3940-9.

WIGHTMAN, F. L. & KISTLER, D. J. 2005. Informational masking of speech in children: effects of ipsilateral and contralateral distracters. *J Acoust Soc Am,* 118**,** 3164-76.

WIGHTMAN, F. L., KISTLER, D. J. & O'BRYAN, A. 2010. Individual differences and age effects in a dichotic informational masking paradigm. *J Acoust Soc Am,* 128**,** 270-9.

WILLIAMS-SANCHEZ, V., MCARDLE, R. A., WILSON, R. H., KIDD, G. R., WATSON, C. S. & BOURNE, A. L. 2014. Validation of a screening test of auditory function using the telephone. *J Am Acad Audiol,* 25**,** 937-51.

WILLIAMS, W. & CARTER, L. 2017. Tinnitus and leisure noise. *Int J Audiol,* 56**,** 219-225.

WINSTON-GERSON, R. & SABO, D. L. 2016. Hearing Loss Detection in Schools and Early Child Care Settings: An Overview of School-Age Hearing Screening Practices. *NASN Sch Nurse,* 31**,** 257-62.

WUNDERLICH, J. L. & CONE-WESSON, B. K. 2006. Maturation of CAEP in infants and children: a review. *Hear Res,* 212**,** 212-23.

YOSHINAGA-ITANO, C., SEDEY, A. L., COULTER, D. K. & MEHL, A. L. 1998. Language of early- and later-identified children with hearing loss. *Pediatrics,* 102**,** 1161-71.

ZHAO, F., MANCHAIAH, V. K., FRENCH, D. & PRICE, S. M. 2010. Music exposure and hearing disorders: an overview. *Int J Audiol,* 49**,** 54-64.

ZOKOLL, M. A., HOCHMUTH, S., WARZYBOK, A., WAGENER, K. C., BUSCHERMOHLE, M. & KOLLMEIER, B. 2013. Speech-in-noise tests for multilingual hearing screening and diagnostics1. *Am J Audiol,* 22**,** 175-8.

# List of abbreviations

| | |
|---|---|
| AABR | automated auditory brainstem response |
| AUC | area under the curve |
| CVC | consonant-vowel-consonant |
| CI | confidence interval |
| dB | decibel |
| dBA | decibel A-weighted |
| DIN | Digits-in-noise test |
| DTT | Digit Triplet Test |
| EC | Earcheck |
| FN | false-negatives |
| HF | high-frequency |
| HFHL | high-frequency hearing loss |
| HI | hearing-impaired |
| HL | hearing level |
| Hz | hertz |
| ICC | intra-class correlation coefficient |
| ISO | International Organization for Standardization |
| kHz | kilohertz |
| LP | low-pass |
| LR+ | positive likelihood ratio |
| LR- | negative likelihood ratio |
| LTASS | long-term average speech spectrum |
| N | number |
| NHANES | National Health and Nutrition Examination Surveys |
| NIHL | noise-induced hearing loss |
| NH | normal-hearing |
| M | mean |
| OAE | otoacoustic emissions |
| OEC | Occupational Earcheck |
| PTA | pure-tone average |
| PV+ | positive predictive value |
| PV- | negative predictive value |
| RCT | randomized controlled trial |
| ROC | receiver operating characteristics |
| SD | standard deviation |
| SEC | Sound Earcheck |
| SI(I) | speech intelligibility (index) |
| SNR | signal-to-noise ratio |
| SRT | speech reception threshold |
| WHO | World Health Organization |

# Summary

Speech-in-noise testing provides a sensitive approach for screening for noise-induced hearing loss (NIHL), as one of the first symptoms of NIHL is a difficulty in understanding speech in background noise. It's possibility for internet presentation offers easily and broadly accessible hearing screening, however, the test's accuracy should be adequate enough for an accurate and efficient application.

As described in **Chapter 1**, the main theme of this thesis is the evaluation of internet-based speech-in-noise hearing screening tests for NIHL in two distinct populations: the working-age population exposed to occupational noise on the one hand, and the general teenage population exposed to recreational noise on the other. These two populations are specifically known to be at risk for NIHL. The main objectives were to evaluate and improve the existing tests in order to obtain valid, reliable, and suitable hearing screening tests.

The first part of this thesis (Chapters 2, 3, and 4) focusses on the evaluation of an internet-based speech-in-noise test "Occupational Earcheck" (OEC; In Dutch: "Bedrijfsoorcheck") for NIHL screening in adults exposed to occupational noise. The test accuracy and precision were investigated by means of a phased approach. First, the main recommendations of previous research were incorporated in OEC in order to improve the test. This was followed by an evaluation in a well-controlled laboratory setting, and subsequently in more realistic and less-controlled occupational settings in representative noise-exposed populations.

**Chapter 2** describes a preliminary study in which OEC was evaluated. A laboratory-based cross-sectional study was performed, in which the optimized test, in terms of homogenization of the speech material and shortening the test length, was evaluated in alternative low-pass filtered masking noise conditions, in order to improve the discriminative power of the test. The tests were performed by adult normal-hearing (NH) listeners and middle-aged listeners with different degrees of high-frequency hearing loss (HFHL) probably (partly) caused by noise exposure. OEC in a low-pass (LP) filtered stationary background noise (test version LP 3) was found to be the most accurate in differentiating

between NH and HFHL listeners, while remaining sufficiently precise. Although this evaluation was performed in a controlled setting, and in a small study sample (N=33) using a two-gate design, the study provides insight into the best option for an alternative masking noise and results indicate the potential of the improved OEC for screening purposes in an occupational environment.

In **Chapter 3** the improved OEC was further evaluated in a representative population of noise-exposed workers in more realistic occupational environments as compared to the laboratory setting. Based on results of 94 workers, a cut-off value for a dichotomous pass or fail classification was chosen, resulting in a proper trade-off between test sensitivity and test specificity on the ear level. Then, a transition from single ear-level test results to test results for the individual was realized. The sensitivity and specificity of OEC for HFHL (defined as a pure-tone average of the higher frequencies 3, 4, and 6 kHz (PTA346) of >25 dB HL for at least one ear) on the individual level were 90% and 77%, respectively. A possible learning effect may have resulted in a lower test specificity, and should therefore be accounted for. Furthermore, the cut-off value for pass and fail was established post-hoc in the same study sample as was used to obtain test characteristics, possibly overrating sensitivity and specificity for HFHL in the target group. The set cut-off value should therefore be confirmed in other, independent, noise-exposed populations.

This is described in **Chapter 4**; the OEC was validated in another noise-exposed population of 80 workers in order to assess the sensitivity and specificity for the test using the cut-off value established in an earlier study, and incorporating automatic conditional rescreening (i.e. a retest was provided for the ear(s) of a subject with a positive test result). This was done in order to account for possible learning effects, as it is known that test accuracy may vary according to population characteristics. A secondary objective was to establish the test accuracy for different degrees of HFHL, and for different age groups. As expected, since this is a different independent population, a lower test sensitivity and specificity for mild HFHL (defined as a PTA346 of >25 dB HL for at least one ear) after a single test was found, namely 65% and 63%, respectively. After the automatic conditional rescreen, the specificity considerably improved to 93%, while the sensitivity slightly decreased to 59%. Test sensitivity and specificity including automatic conditional rescreening for a moderate HFHL (defined as a PTA346 of >35 dB HL for at least one ear) was better; 94% and 90%, respectively. OEC appears to be suitable for ruling out HFHL, preventing unnecessary referrals, however, it is less sensitive for detecting mild HFHL. The sensitivity and specificity of OEC varied according to age group. However the age subgroups were small, therefore the test accuracy should be further explored in various working-age populations.

The second part of this thesis (Chapters 5, 6 and 7) focusses on the application and the suitability of internet-based screening tests in young people. The main approach was to explore online test results, focusing on the effects of age, test-retest reliability, and type of test presentation.

**Chapter 5** describes the results of speech-in-noise testing by means of the Dutch online speech-in-noise test "Earcheck" (EC; In Dutch: "Oorcheck"), in teenagers and young adults. In total, 96,803 valid online EC responses were collected nation-wide for over five years. A significant proportion of users obtained insufficient or poor test scores, implying that a substantial number had some difficulty in understanding speech in noise potentially related to NIHL. Furthermore, mean SRT scores improved with increasing age, especially among teenagers between the age of 12 to 18 years. This study was conducted on a large internet convenience sample of users, and due to its limitations (such as inclusion bias, poorly controlled test conditions, and uncertainties concerning the users' hearing threshold levels and age specifications) the results are not generalizable to the Dutch teenage population. Therefore, online speech-in-noise performance in this population should be studied further in a more controlled study focusing on age effects.

Consequently, EC performance and age effects were further evaluated in 72 normal-hearing teenagers aged 12-17 years, in well-controlled school settings. This study is described in **Chapter 6**. The primary aim was to investigate the effects of age, level of education, gender, and test repetition on EC's speech-reception threshold (SRT) performance in this teenage sample. Regression analyses of the SRT scores within the teenage sample revealed an effect of age between the ages of 12-17 years. The scores of the 17 year-olds were comparable to young adults' SRT scores, whereas younger participants yielded poorer results. Effects of level of education and gender were not significant.

Repeated testing revealed a small, but consistent learning effect. Age correction factors are suggested to compensate for the maturation effect on EC results. Additional research is required in larger populations, including hearing-impaired (HI) teenagers, to verify the findings and proposed age-corrections, and in order to assess the tests' ability to detect relevant hearing loss in this specific age group.

Additionally, in **Chapter 7**, the suitability of online and smartphone speech-in-noise testing at school was investigated in primary school-age children by means of the "Dutch hearing screening test for children" (In Dutch: "Kinderhoortest"). The main objectives were to evaluate test-retest reliability, the effect of presentation type on a desktop computer or smartphone, and to assess age effects in a group of 75 normal-hearing children. Overall, two test conditions were performed better as compared to a training test, and the second test condition was performed better as compared to the first test condition, indicating an additional learning effect after training in school-age children. The study further suggested that the test can be performed either on a desktop computer or on a smartphone, as there were no significant differences in SRT score by type of presentation. Furthermore, in this study age-related effects were present as well; the older children outperformed the younger children in all test conditions. The test should be evaluated further in a larger population of school-age children, including HI children, in order to further assess learning effects and age-effects, and to evaluate its sensitivity and specificity for childhood hearing loss in realistic screening conditions.

Finally, in **Chapter 8**, the value of the internet-based speech-in-noise tests for screening purposes is discussed, and the study results are interpreted while taking into account important methodological considerations. Furthermore, implications for future research and practice are presented.

In conclusion, this thesis contributes to the evidence base for the accuracy of internet-based speech-in-noise screening tests for NIHL. Several test improvements have been proposed and implemented, resulting in appropriate screening tests for occupational and recreational NIHL, however, there is still room for further improvement and fine-tuning.

# Samenvatting

Eén van de eerste en belangrijkste kenmerken van lawaaislechthorendheid is het optreden van problemen bij het verstaan van spraak in een rumoerige omgeving. De functionele spraak-in-ruis test die het spraak verstaan in achtergrondruis meet, is om deze reden een geschikte test voor screening, een methode om lawaaislechthorendheid vroegtijdig op te sporen. Een simpele woorden-in-ruis test kan via het internet aangeboden worden en door de afnemer zelf worden uitgevoerd, waardoor het gehoor op een laagdrempelige, snelle en goedkope manier op afstand getest kan worden. Echter, voor screeningsdoeleinden is het van groot belang dat de test voldoende accuraat is, zodat efficiënte, valide en betrouwbare metingen uitgevoerd kunnen worden.

Het thema van dit proefschrift, zoals beschreven in **Hoofdstuk 1**, is de evaluatie van online spraak-in-ruis screeningstesten voor lawaaislechthorendheid in twee specifieke populaties, namelijk werknemers die beroepsmatig aan lawaai worden blootgesteld en jongeren die in hun vrije tijd langdurig aan harde geluiden worden blootgesteld, met name luide muziek. Hierbij staat het evalueren van de validiteit, betrouwbaarheid en geschiktheid van de testen centraal, om de testen te verbeteren voor screeningsdoeleinden, met het uiteindelijke doel het efficiënter opsporen van (beginnende) lawaaislechthorendheid om verergering te voorkomen. Meer algemeen wordt gestreefd naar een gezondheidswinst met betrekking tot het gehoor op publiek niveau. Verder wordt in dit hoofdstuk een algemene introductie gegeven met betrekking tot diverse gerelateerde achtergrondthema's, zoals lawaaislechthorendheid, preventie, screening door middel van online spraak-in-ruis testen, bestaande online screening testen en de algemene methodologie die is toegepast in dit proefschrift.

Het eerste deel van het proefschrift (de hoofdstukken 2, 3 en 4) richt zich op de evaluatie van de online spraak-in-ruis test "Bedrijfsoorcheck"(BOC; in het Engels: "Occupational Earcheck") voor het screenen van volwassenen die beroepsmatig worden blootgesteld aan lawaai. De validiteit en precisie van de test worden bestudeerd aan de hand van een gefaseerde aanpak. Als eerste wordt in een goed-gecontroleerde klinische test omgeving aangetoond dat de test een onderscheid kan maken tussen lawaaislechthorendheid en normaal gehoor en wordt de test verder aangepast en verbeterd. Vervolgens wordt de test in een realistische en minder gecontroleerde bedrijfsomgeving geëvalueerd.

**Hoofdstuk 2** beschrijft een studie waarin de BOC wordt geëvalueerd in een gecontroleerde klinische omgeving. De test wordt verbeterd, door het spraakmateriaal te homogeniseren en de test in te korten, en vervolgens geëvalueerd in vijf verschillende maskeerruiscondities. De testen worden uitgevoerd door, enerzijds, normaalhorende volwassenen en, anderzijds, volwassenen met een hoogfrequent gehoorverlies, hoogstwaarschijnlijk veroorzaakt door lawaaiblootstelling. Uit het onderzoek blijkt dat de BOC met een laag-gefilterde stationaire maskeerruis (test versie 3) het beste kan discrimineren tussen normaal gehoor en lawaaislechthorendheid en tevens voldoende betrouwbaar is. Ondanks dat de evaluatie is uitgevoerd in een gecontroleerde omgeving en de groepen klein waren (N=33) en niet gerandomiseerd, toont deze studie aan dat de verbeterde BOC potentieel geschikt is als een screeningstest in de arbeidsgeneeskunde.

In **Hoofdstuk 3** wordt de verbeterde BOC verder geëvalueerd in een representatieve populatie van aan lawaai blootgestelde werknemers (N=94) op locatie in een realistische bedrijfsomgeving. Er wordt een afkappunt bepaald voor een dichotome classificatie (i.e. een positieve of een negatieve uitslag) op oor-niveau, waarbij er een afweging wordt gemaakt tussen de sensitiviteit en de specificiteit van de test. Vervolgens wordt een vertaalslag gemaakt van oor-niveau naar individueel niveau. De sensitiviteit en specificiteit van de BOC voor lawaaislechthorendheid (gedefinieerd als een gemiddelde gehoordrempel van de hogere frequenties 3, 4 en 6 kHz (PTA346) van >25 dB HL, voor tenminste 1 oor) op individueel niveau zijn 90% en 77%. De specificiteit is waarschijnlijk lager uitgevallen door de aanwezigheid van een leereffect. Verder is het afkappunt post hoc bepaald voor deze studiepopulatie. Dit zou een overschatting van de gevoeligheid van de test kunnen betekenen in de beoogde populatie. Om deze reden moet de test met het gekozen afkappunt ook nog in een onafhankelijke groep lawaai-blootgestelde werknemers worden gevalideerd.

Deze studie wordt beschreven in **Hoofdstuk 4**; de BOC wordt gevalideerd in een andere populatie lawaai blootgestelde werknemers (N=80) om zodoende de sensitiviteit en specificiteit te bepalen voor de BOC met het gekozen afkappunt. Daarnaast worden ook de mogelijkheden van een automatische conditionele re-screening onderzocht. Dit is een herhaling van de test voor het oor met een positieve testuitslag om zodoende te corrigeren voor een mogelijk leereffect dat ongewenst een hogere SRT score veroorzaakt. Overige subdoelen van de studie zijn het bepalen van de testgevoeligheid voor verschillende maten van lawaaislechthorendheid en voor verschillende leeftijdsgroepen. Zoals verwacht, wordt er in deze studie een lagere gevoeligheid behaald voor milde lawaaislechthorendheid (gedefinieerd als een gemiddelde gehoordrempel van de hogere frequenties 3, 4 en 6 kHz (PTA346) van >25 dB HL, voor tenminste 1 oor) na een enkele test, namelijk 65% en 63%. De automatische conditionele re-screening leidt tot een aanzienlijke verbetering in de specificiteit (93%), terwijl de sensitiviteit verder afneemt naar 59%. De gevoeligheid van de test voor een matig gehoorverlies (gedefinieerd als een gemiddelde gehoordrempel van de hogere frequenties 3, 4 en 6 kHz (PTA346) van >35 dB HL voor tenminste 1 oor) na een automatische conditionele re-screening is groter, met een sensitiviteit van 94% en een specificiteit van 90%. Verder verschilt de test gevoeligheid per leeftijdsgroep, echter waren de groepen te klein om een significante verschillen aan te kunnen tonen.

De BOC lijkt een geschikte test om lawaaislechthorendheid uit te sluiten in een realistische werkomgeving en kan daarmee onnodige doorverwijzingen voorkomen. De test is minder gevoelig voor het opsporen van milde gevallen, waardoor de kans bestaat dat deze gemist zullen worden in de praktijk. Gemiste gevallen kunnen alsnog opgespoord worden door de test regelmatig aan te bieden.

Het tweede deel van deze thesis (hoofdstukken 5, 6 en 7) richt zich op de toepassing en geschiktheid van online spraak-in-ruis screeningstesten bij jonge mensen die blootgesteld worden aan lawaai in hun vrije tijd. De aanpak behelst het onderzoeken van online test resultaten, met een focus op de effecten van leeftijd en presentatie en op de test-retest betrouwbaarheid.

**Hoofdstuk 5** geeft de beschrijving van testuitslagen weer van jongeren van 12-24 jaar die de Nederlandse online spraak-in-ruis screening test “Oorcheck” (In het Engels: “Earcheck”) online hebben verricht. In totaal zijn er in een periode van vijf jaar, 96 803 valide online test responses verzameld op nationaal niveau. Een significant percentage Oorcheck gebruikers in die leeftijdsgroep kreeg een onvoldoende of slechte testuitslag, wat mogelijk impliceert dat een substantieel aantal moeite heeft met het verstaan van spraak in ruis. Hoewel de onderliggende oorzaak onbekend is, kan dit gerelateerd zijn aan lawaaislechthorendheid. Verder scoorden oudere jongeren beter op de test dan jongere tieners (tussen de 12 en 18 jaar). Omdat deze studie online is uitgevoerd bij een grote groep jongeren, waardoor de studie een aantal beperkingen heeft (zoals inclusie bias, ongecontroleerde testomstandigheden en onzekerheden met betrekking tot de gehoordrempels en leeftijd van de gebruikers), kunnen er op basis van de resultaten geen uitspraken worden gedaan die representatief zijn voor de algemene Nederlandse jongerenpopulatie. Om meer inzicht in online spraak-in-ruis prestaties van jongeren te krijgen is verder onderzoek nodig door middel van een gecontroleerde studie.

Zodoende is in **Hoofdstuk 6** een studie uitgevoerd bij 72 normaalhorende jongeren van 12 tot 17 jaar. Het doel van deze studie is om de effecten van leeftijd, geslacht, opleidingsniveau en test herhaling op Oorcheck uitslagen te onderzoeken in een realistische screeningsomgeving, namelijk de middelbare school. Regressie analyses wijzen op het bestaan van een leeftijdseffect met betrekking tot de testscore, weergegeven als de spraakverstaanvaardigheids drempel ("speech reception threshold" SRT). Jongere tieners hebben slechtere uitslagen dan oudere tieners. De SRT-uitslagen van 17-jarigen zijn vergelijkbaar met de uitslagen van jonge volwassenen. Er is bovendien sprake van een leereffect bij herhaling van de testen. Effecten van geslacht en opleidingsniveau zijn niet significant. Vervolgens worden leeftijdscorrectiefactoren voorgesteld om te compenseren voor rijpingseffecten met betrekking tot de Oorcheck scores. Verder onderzoek in grote jongerenpopulaties inclusief jongeren met een gehoorafwijking is nodig om de bevindingen van deze studie te verifiëren en de correctiefactoren te valideren en tevens om de test accuratesse van Oorcheck voor gehoorverlies te onderzoeken.

Verder wordt in **Hoofdstuk 7** de geschiktheid van een smartphone spraak-in-ruis screeningstest onderzocht, de "Kinderhoortest" (In het Engels: "Dutch hearing screening test for children") bij 75 kinderen in een schoolomgeving. Test-retest betrouwbaarheid, de effecten van presentatie type (pc en smartphone) en leeftijdseffecten worden onderzocht. Over het algemeen worden de twee testen beter verricht dan een training test die aan de daadwerkelijke metingen vooraf ging, en is de uitslag van de tweede test beter dan de eerste test, wat een additioneel leereffect aanduidt na de training. Verder blijkt uit de studie dat de test zowel op een computer als op een smartphone aangeboden kan worden, aangezien er geen significante verschillen in SRT uitslagen zijn. Verder blijkt dat bij de kinderen tevens sprake is van een leeftijdseffect: de oudere kinderen hebben betere uitslagen dan de jongere kinderen in elke test conditie. De Kinderhoortest zou verder geëvalueerd moeten worden in een grotere populatie van schoolgaande kinderen, inclusief kinderen met een gehoorverlies, om de gevonden leeftijds- en leereffecten verder in kaart te brengen en bovendien om de sensitiviteit en specificiteit voor gehoorverlies bij kinderen in realistische screeningsomstandigheden te evalueren.

Ten slotte wordt in **Hoofdstuk 8** de waarde van de onderzochte online-spraak-in-ruis testen bediscussieerd. De onderzoeksresultaten van de testen worden geïnterpreteerd en beoordeeld. Er worden belangrijke methodologische aspecten benoemd en aanbevelingen gedaan voor toekomstig onderzoek en de implementatie in de praktijk.

Dit proefschrift draagt bij aan de evidence-base voor de accuratesse van online spraak-in-ruis testen voor gehoorscreening. Diverse testaanpassingen hebben geleid tot een verbetering van de bestaande screeningstesten voor de praktische toepassing in zowel de arbeidsgeneeskunde, als bij jongeren. De testen kunnen een zinvolle bijdrage leveren aan screening en monitoring van gehoorverliezen die door lawaai zijn ontstaan, echter is er nog ruimte voor verdere evaluatie en fine-tuning van de testen.

# Dankwoord

Graag wil ik alle mensen heel hartelijk bedanken die mij waardevolle inzichten hebben gegeven, zowel op wetenschappelijk als persoonlijk gebied. De mensen die mij hebben geholpen te groeien tot zelfstandig onderzoeker, maar ook als persoon. Zonder jullie hulp was dit proefschrift niet tot stand gekomen!

Allereerst mijn promotor **Wouter**. Dank voor al het vertrouwen in mij en voor alle kansen die je mij geboden hebt. Ik vond het geweldig om deel uit te maken van het preventie-team en een bijdrage te mogen leveren aan de testen. De afgelopen jaren heb ik ontzettend veel van je geleerd en jouw begeleiding heb ik ook als zeer plezierig ervaren. Dank voor alle steun, geduld, ruimte en flexibiliteit die je me geboden hebt om mezelf te kunnen ontwikkelen.

Mijn copromotor **Monique**, wat fijn dat ik jou als voorganger, mede-gezondheidswetenschapper en -preventist en uiteindelijk begeleider had. Ik heb veel van je geleerd en je stond altijd voor me klaar. Ook toen je niet meer bij ons op de afdeling werkte, heb je me op afstand toch altijd flink op weg kunnen helpen en kunnen steunen bij de afronding van mijn boekje. Onze (telefonische) afspraken de afgelopen jaren waren niet alleen zeer nuttig, maar ook heel erg gezellig en daardoor vaak ook net iets te lang..

Leden van de leescommissie, **Prof. dr. F.G. Dikkers, Prof. dr. M.H.W. Frings-Dresen, Dr. B. Sorgdrager, Prof. dr. ir. A.F.M. Snik, Prof. dr. A. van Wieringen en Dr. J.C.M. Smits**. Dank voor het lezen en beoordelen van mijn proefschrift.

**Laura van Deelen**, dankjewel voor het realiseren van alle kansen en het vertrouwen in mijn werk. Zonder alle interessante vragen en projecten vanuit de Nationale Hoorstichting, was er geen onderzoek geweest om dit boekje mee te vullen. **Lisanne en Kelly**, het was erg leuk om met jullie samen te werken aan alle onderzoeksprojecten.

**Jan de Laat,** dankjewel voor het meeschrijven aan de papers. Ik heb veel bewondering voor jouw onuitputtelijke enthousiasme en passie voor het vak. Ik vond onze samenwerking dan ook zeer inspirerend en plezierig.

**Irene**, wat leuk dat we samen nog hebben kunnen werken aan het scholen onderzoek en het schrijven van een artikel. Dank voor alle gezelligheid en promotie tips!

**Mariska Leeflang**, bedankt voor de methodologische input. Het heeft me enorm geholpen bij de afronding van dit proefschrift.

**Ilja,** jij bracht zoveel gezelligheid en vrolijkheid mee op de afdeling. In de laatste fase van mijn promotie heb je me vaak een luisterend oor geboden. Ik ben dan ook heel blij dat je mijn paranimf wil zijn. En we moeten nog steeds een keer naar Artis!

**Hiske,** je gaf mij een warm welkom op de afdeling en je hebt je daarna altijd over mij 'ontfermd'. Ik vond het heel gezellig om jouw kamergenoot te zijn, samen met jou te sparren over onderzoek en samen cursussen te volgen. Dank voor alle fijne gesprekken die we hebben gehad. Ik wens je heel veel succes met jouw onderzoek!

**Inge**, vanaf het begin stond je altijd voor me klaar. Je hebt me wegwijs gemaakt binnen de afdeling en de wereld van de audiologie. Dankjewel voor alle hulp die je me geboden hebt, van het meedenken bij ingewikkelde onderzoeksvragen en het maken van Matlab figuren, tot aan het audiometreren op locatie. En natuurlijk dank voor alle gezelligheid op de afdeling.

**László**, dank voor alle steun bij het meten en het oplossen van technische problemen. En uiteraard, dank voor alle nodige afleiding en inspiratie.

**Sabine**, dank voor het nakijken van mijn werk op het Engels en je interesse in mijn onderzoek.

**Koen,** ondanks dat ik je al een tijd niet meer heb gezien, dank voor de interesse in mijn werk en het destijds bewerken van het spraak- en ruis materiaal voor de Bedrijfsoorcheck.

Dank aan alle andere (research) collega's voor de goede sfeer op de afdeling, op het lab, en in mijn kamer: **Femke, Rolph, Jelmer, Monique Boymans, Maaike de**

**Vrijer, Thamar, Maj, Mirjam, Gijs, Hilde, Simon, Ingrid, Joke, Miranda, Thirza, Hieke, Patrick, Tim en alle anderen**, dank jullie wel voor alle hulp, interesse en gezelligheid.

Daarnaast wil ik ook mijn lieve vriendinnen bedanken: **Maha, Asiana, Sana, Shamaila, Hakima, Somia, Rachida, Roha en Yasemin**. Ondanks dat we elkaar niet zo vaak zien als dat ik zou willen, is het toch altijd weer ontzettend gezellig, vertrouwd en als vanouds. Ik kijk uit naar nog veel meer gezellige etentjes, shop-uitjes, spa bezoekjes, filmavonden, etc. **Maha, Hakima, Monika en Amara,** het was erg leuk dat jullie naast vriendinnen en zus, ook AMC collega's zijn geweest en dat we weleens samen hebben kunnen pauzeren.

Mijn lieve (schoon)familie, ik wil jullie bedanken voor alle interesse, ontspanning en gezelligheid: **Zainab & Mohamed, Naoual & Mustafa, Nabila & Hicham, Imad, Amara & Ashraf en Sobana:** dank jullie wel voor jullie gezelschap. Mijn lieve zussen **Amara en Sobana,** jullie kennen mij waarschijnlijk als de beste. Ik ben blij dat we al zoveel samen hebben meegemaakt en dat ik nog steeds voor van alles bij jullie terecht kan. Ik ben trots op jullie. **Amara**, wat fijn dat je ook mijn paranimf wil zijn.

En al onze kanjers en Adam en Nora's beste vriendjes: **Wail, Naoufal, Junaid, Amani, Anwar, Tahir en Sarah**: Ik ben er trots op jullie tante te zijn!

Mijn lieve **ouders en schoonouders**, dank jullie wel voor jullie onvoorwaardelijke liefde, zorgen, steun en geduld. Ik ben blij dat ik altijd bij jullie terecht kan. Terwijl jullie voor de kindjes zorgden (en ook voor mij), kon ik zonder zorgen werken aan mijn promotie. Lieve **mam en abu**, ik ben erg dankbaar dat jullie mij hebben grootgebracht. Dankzij jullie opvoeding en goede zorgen ben ik geworden wie ik nu ben. Ik hoop op een mooie toekomst samen, vol geluk, gezondheid en zegeningen.

Lieve **Mohamed.** Ik ben erg gezegend met jou aan mijn zijde. Jij zorgt voor geluk, focus en energie. Dankzij jou besef ik me altijd weer waar het allemaal echt om draait in het leven. En wanneer ik er zelf even niet meer in geloofde, sprak jij me de moed weer in die ik nodig had om mijn promotie tot een goed einde te brengen. Dank je wel voor al je liefde, steun en geduld.

Mijn allerliefste **Adam en Nora**, ook al zijn jullie nu nog te klein om dit te kunnen lezen, ik hou zo veel van jullie. Door mijn werk ben ik me er nog meer van bewust geworden hoe belangrijk het gehoor is en hoe blij ik mag zijn dat ik elke dag weer kan genieten van jullie aanstekelijke lach en heerlijke praatjes. Ik heb hard gewerkt om de titel dr. te behalen, maar de waarheid is dat jullie mij al de beste en de mooiste titel hebben gegeven die er is en dat is 'mama' (of zoals Adam mij ook wel noemt: 'mams' of 'mamietje').

## Curriculum vitae

Marya Sheikh Rashid was born on August 12th 1987 in Haarlem. She finished higher secondary school at Linnaeus College Haarlem in 2005. Subsequently, she obtained her Bachelor Health Sciences at the VU University in Amsterdam in 2008. During the Master Health Sciences at the same university she specialized in 'Policy and Organisations in Healthcare' and obtained her Master of Science degree in 2009. She finished her MSc thesis at the hospital pharmacy of OLVG-West Amsterdam, which included a systematic review on discharge medication-related interventions. In 2011, she obtained her second Master of Science degree in Epidemiology (Research Master Biomedical Sciences, with a specialization in Clinical Epidemiology) at the University of Utrecht. She performed her final internship at the department of Public Health of the UMC Utrecht Julius Center where she studied the effectiveness of public smoking bans. Since April 2012, she has been working as a scientific researcher at the department of Clinical and Experimental Audiology of the Academic Medical Center Amsterdam. There, she conducted research on the (secondary) prevention of occupational and recreational noise-induced hearing loss under supervision of Prof. dr. ir. W.A. Dreschler and Dr. M.C.J. Koks-Leensen, which resulted in this PhD project.

# Publications

## Journal articles (included in this thesis)

- Sheikh Rashid M, Dreschler WA. Accuracy of an internet-based speech-in-noise hearing screening test for high-frequency hearing loss: Incorporating automatic conditional rescreening. *Revised manuscript submitted to International Archives of Occupational and Environmental Health*.

- Jacobi I, Sheikh Rashid M, de Laat JAPM, Dreschler WA. Earcheck. Age dependence of thresholds for speech in noise in normal-hearing adolescents. *Trends Hear*, 2017; 21:1-9.

- Sheikh Rashid M, Dreschler WA, de Laat JAPM. Evaluation of an internet-based speech-in-noise screening test for school-age children. *Int J Audiol*, 2017; 56(12):967-975.

- Sheikh Rashid M, Leensen MCJ, de Laat JAPM, Dreschler WA. Laboratory evaluation of an optimized internet-based speech-in-noise test for occupational high-frequency hearing loss screening: Occupational Earcheck. *Int J Audiol*, 2017; 56(11):844-853.

- Sheikh Rashid M, Leensen MCJ, de Laat JAPM, Dreschler WA. Cross-sectional evaluation of an internet-based hearing screening test in an occupational setting. *Scand J Work Environ Health,* 2017; 43(3):279-286.

- Rashid MS, Leensen MCJ, Dreschler WA. Application of the online hearing screening test 'Earcheck': Speech intelligibility in noise in teenagers and young adults. *Noise Health*, 2016; 18(85):312-318.

## Other publications (not included in this thesis)

- Ting JW, Sheikh Rashid M, Dreschler WA, Brand HS. Tandartsen en beroepsgerelateerde lawaaislechthorendheid. Het tandheelkundig jaar 2016 (Hoofdstuk 2). Uitgeverij: Bohn Stafleu van Loghum.

- Sheikh Rashid M, de Laat JAPM, Dreschler WA. Onderzoek naar de toepasbaarheid van de Bedrijfsoorcheck (BOC). Deel B: Evaluatie in de praktijk. Amsterdam, 2016.

- Sheikh Rashid M, de Laat JAPM, Dreschler WA. Evaluatie Kinderhoortest II. Amsterdam, 2016.
- Sheikh Rashid M, Jacobi I, de Laat JAPM, Dreschler WA. Application of the online hearing screening test Earcheck: Online speech-in-noise testing and SRT performance among teenage students. Amsterdam, 2016.
- Sheikh Rashid M, Leensen MCJ, Dreschler WA. Report online hearing tests 2010-2014: Five years of Oorcheck & Hoorscan. Amsterdam, 2015.
- Adluni-Sheikh Rashid M, Dreschler WA. Evaluatie Kinderhoortest I. Amsterdam, 2014.
- Adluni-Sheikh Rashid M, Leensen MCJ, Dreschler WA. De toepasbaarheid van een internetscreeningstest om lawaaislechthorendheid (NIHL) te detecteren. Amsterdam, 2014.

# Portfolio - PhD training, AMC Graduate School for Medical Sciences

| Courses | Year | ECTS |
| --- | --- | --- |
| Basic Course in Legislation and Organization for Clinical Researchers (BROK) | 2013 | 1.0 |
| Clinical data management | 2015 | 0.2 |
| Clinical epidemiology: Evaluation of medical tests | 2015 | 0.9 |
| Project management | 2015 | 0.6 |
| Recertification BROK | 2016 | 0.1 |
| Didactical skills training | 2017 | 0.2 |

| Seminars and meetings | | |
| --- | --- | --- |
| Weekly department research meetings | 2012-2017 | 5.5 |
| Scientific meetings Werkgemeenschap Auditief Systeem (WAS) | 2012-2017 | 0.5 |
| Meetings Nederlandse Vereniging van Audiologie (NVA) | 2012-2017 | 0.5 |
| Annual scientific research days ENT department, AMC | 2012-2017 | 3 |
| Meetings with the Netherlands Hearing Health Foundation (Nationale Hoorstichting, NHS) | 2012-2016 | 3 |

| (Inter)national conferences | | |
| --- | --- | --- |
| ARCHES meeting, Paris, France | 2013 | 0.5 |
| Hearing across the lifespan (HEAL) , Cernobbio, Italy | 2014 | 0.8 |
| ARCHES meeting, Oldenburg, Germany | 2014 | 0.5 |
| The 2nd International Meeting on Internet & Audiology, Snekkersten, Denmark | 2015 | 0.5 |
| ARCHES meeting, Groningen, the Netherlands | 2015 | 0.5 |
| The 41th Annual conference National Hearing Conservation Association (NHCA), San Diego, USA | 2016 | 0.8 |
| The 3rd International Meeting on Internet and Audiology, Louisville, USA (remote attendance) | 2017 | 0.2 |
| ARCHES meeting, Leuven, Belgium | 2017 | 0.5 |

| Presentations | | |
| --- | --- | --- |
| *'MP3-check results',* meeting with NHS, LUMC. | 2012 | 0.5 |
| *'The challenge of detecting minimal hearing loss in audiometric surveys'*, literature club, AMC. | 2013 | 0.5 |
| *'OEC adjustments'*, meeting with NHS, LUMC. | 2013 | 0.5 |
| *'The risk of music-induced hearing loss among Dutch adolescents due to MP3-player use'* (poster), ARCHES meeting, Paris, France. | 2013 | 0.5 |
| *'The risk of music-induced hearing loss among Dutch adolescents due to MP3-player use'* (poster), NHCA, Las Vegas, USA. | 2014 | 0.5 |
| *'The risk of music-induced hearing loss among Dutch adolescents due to MP3-player use'* (poster), HEAL, Cernobbio, Italy. | 2014 | 0.5 |
| *'Evaluation and validation of the Occupational Earcheck'*, ARCHES meeting, Oldenburg, Germany. | 2014 | 0.5 |

| | | |
|---|---|---|
| '*Noise-induced hearing loss screening by internet'*, scientific research day ENT, AMC. | 2014 | 0.5 |
| *'Results hearing tests 2013 and 2014'*, meeting with NHS, LUMC. | 2015 | 0.5 |
| *'Evaluation and validation of the Occupational Earcheck'*, meeting on Internet & Audiology. | 2015 | 0.5 |
| *'Validation of a Dutch online speech in noise test for NIHL screening among noise-exposed workers'*, ARCHES meeting, Groningen, the Netherlands. | 2015 | 0.5 |
| *'Evaluation and validation of the Occupational Earcheck'*, NHCA, San Diego, USA. | 2016 | 0.5 |
| *'Online speech-in-noise test for hearing screening'*, scientific research day ENT, AMC. | 2016 | 0.5 |
| *'Effects on speech-in-noise performance in normal-hearing children and teenagers in school settings'*, ARCHES meeting, Leuven, Belgium. | 2017 | 0.5 |

| **Supervising** | | |
|---|---|---|
| L. Ruhaak, Master student Speech Therapy, scientific internship | 2013 | 1 |
| L. Ting, Master student Dentistry, scientific internship | 2014-2015 | 4 |
| L. Sels, Master student Medical Physics, scientific internship | 2016-2017 | 1 |

P