Noise induced hearing loss: Screening with pure-tone audiometry and speech-in-noise testing

Leensen, M.C.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Speech-in-noise screening tests by internet, Part 1: Test evaluation for noise-induced hearing loss identification

M.C.J. Leensen¹
J.A.P.M. de Laat²
W.A. Dreschler¹

¹ Clinical and Experimental Audiology, ENT Department, Academic Medical Centre, Amsterdam, The Netherlands
² Department of Audiology, Leiden University Medical Centre, Leiden, The Netherlands

International Journal of Audiology, 2011, 50 (11), 823 - 834
Abstract

Objective: In the Netherlands three internet-based self-screening tests for hearing have been developed; the National Hearing Test (NHT), Earcheck (EC), and Occupational Earcheck (OEC). These tests are adaptive speech-in-noise tests using either digit triplets or monosyllables, presented in stationary speech-shaped noise. These tests can be highly valuable in increasing the awareness and prevention of noise-induced hearing loss (NIHL). This study evaluates these online speech-in-noise tests and investigates their potential to detect NIHL.

Design: In a multi-centre study the results of the three online screening tests are compared to pure-tone audiometry and to the Dutch sentence SRT test (Plomp & Mimpen, 1979a), which is considered the clinical standard.

Study sample: In total, 49 normal-hearing listeners and 49 patients with different degrees of NIHL participated.

Results: The online tests show good reliability, but there is much overlap in outcomes between normal-hearing listeners and participants with early NIHL. In addition, rather low correlations of the tests’ results with both the Dutch sentence SRT test and pure-tone thresholds are found. These findings result in rather low test sensitivity: 54% (NHT) and 51% (EC), or low specificity: 49% (OEC).

Conclusions: The online screening tests in their current form are unsuitable to be used for early NIHL screening purposes.
Introduction

Noise-induced hearing loss (NIHL) is a significant social and public-health problem. In present society large groups of individuals are frequently exposed to high sound levels, either during leisure time or in occupational settings. Currently, NIHL is the most reported occupational disease in the Netherlands (Van der Molen et al, 2010). Occupational NIHL is generally detected by pure-tone air conduction audiometry. In the Netherlands, this is offered at least once every four years to all employees exposed to daily noise levels exceeding 85 dBA, and occurs on a voluntary basis.

Besides noise at workplaces, exposure to loud sounds is increasingly encountered during recreational activities. Concern is growing that over-exposure to amplified music, due to the use of personal music players or discotheque attendance, will cause NIHL in adolescents and young adults. Although evidence supporting a relationship between exposure to leisure noise and hearing damage in young people remains ambiguous (Meyer-Bisch, 1996; Mostafapour et al, 1998; Niskar et al, 2001; Biassioni et al, 2005; Shah et al, 2009; Zhao et al, 2010), any exposure to noise of significant intensity and/or duration is known to increase the risk of hearing damage. Considering the noise levels and the time spent listening to personal music players, approximately 5 to 10% of the listeners are estimated to be at risk of developing permanent hearing loss after five or more years of exposure (SCENIHR, 2008). Furthermore, the damage from chronic exposure to loud music is cumulative, so a slight hearing loss in adolescence can eventually become a substantial one in adulthood (Chung et al, 2005; SCENIHR, 2008), especially among those with higher susceptibility to noise (Biassoni et al, 2005) or those who are employed in a job with significant noise exposure.

Because of the gradual development of NIHL, persons with mild high-frequency hearing loss are often unaware of their impairment until the hearing loss reaches a certain degree (Vogel et al, 2009). Since hearing damage is irreversible, it is of major importance that it is recognized as early as possible. The earlier NIHL is detected, the earlier precautionary measures can be undertaken to prevent more impairing, permanent, hearing damage (Meyer-Bisch, 1996). One of these measures is changing young people’s personal behaviour related to the length of time and sound level at which music is played (Vogel et al, 2007). However, a Delphi study by Vogel et al. (2009) showed that adolescents first must become aware that they personally are at risk for hearing loss due to listening to high-volume music, before the promotion of protective behaviours would be useful.

An objective hearing test can be of great help in the early detection and prevention of NIHL. Access to an easily administered hearing screening test can raise awareness of possible hearing problems and reduce the risk of hearing loss after exposure to noise. Feedback of individual hearing status stimulates persons to seek audiological help.
(Smits et al, 2004; 2006a) or to change their (music listening) behaviour in order to prevent NIHL.

The first noticeable disability caused by noise-induced hearing loss is often a reduced ability to understand speech in a noisy environment. Therefore, the currently available Dutch internet-based speech-in-noise tests can be considered for NIHL screening purposes. The first speech-in-noise self-test for the Dutch language was developed and validated by Smits et al. (2004). This screening test, referred to as the ‘National Hearing Test’ (NHT), presents digit triplets in noise, hence it can be easily administered by telephone. Subsequently, an internet version of this test was generated, in collaboration with the Dutch National Hearing Foundation (NHF) (Smits et al, 2006a). The NHT measures the ability to understand speech in noise by determining the speech reception threshold (SRT), i.e. the signal-to-noise ratio corresponding to 50% intelligibility. The test result is presented to the participant, accompanied by a recommendation for follow-up referral, if required. For reasons of comparison, the bandwidth of the online test materials is limited to 0.3 - 3.4 kHz, to mimic the telephone network frequency band (Smits et al, 2006a). Because NIHL predominantly affects the high frequency region, an internet-based speech-in-noise test that also includes the higher frequencies was generated: ‘Earcheck’ (EC). This test, following a similar procedure as NHT but presenting nine different CVC words in a broadband noise, was developed by LUMC Leiden and NHF (Albrecht et al, 2005). The test specifically aimed at young persons between 12 and 24 years old, to raise awareness about the risks of exposure to loud music in this population. A third speech-in-noise test was developed by LUMC and NHF, specifically applicable in commercial enterprises, to monitor the hearing ability of employees in noisy occupations: ‘Occupational Earcheck’ (OEC) (Ellis et al, 2006). The procedure for this test is similar to that of Earcheck, but it was designed to have better precision by increasing the number of stimuli and by consecutive monaural testing of both ears. Also, it uses a different set of CVC words, containing matching vowels and more high-frequency consonants (Ellis et al, 2006; Kuipers, 2007).

These speech-in-noise self-tests are considered to be suitable for screening purposes. Because of their adaptive nature, they can be implemented as quick and fully automated tests (Jansen et al, 2010), measuring over a range that includes both normal and impaired hearing (Soli & Wong, 2008). The screening tests are developed to be performed in an at-home or private situation. Consequently, respondents make use of a variety of computer equipment and transducers, and both presentation level and the level of ambient noise in the room are unknown. However, measuring the ratio of speech intensity and level of masking noise makes the test relatively independent from the absolute presentation level (Plomp, 1986). Since both speech
and noise with identical spectra are played through the same playback device the signal-to-noise ratio in each frequency band is not sensitive to subtle differences in the transfer characteristics between different devices, making the test robust against possible transmission losses and variations in equipment (Smits et al, 2004; Culling et al, 2005). Finally, the tests are relatively robust against background noise (Culling et al, 2005), require little or no calibration (Jansen et al, 2010) and are insensitive to conductive hearing losses, provided that the presentation level is at the most comfortable loudness level.

When measuring the speech reception threshold in noise, higher SNRs are required for subjects with NIHL than for those without hearing loss (Chung & Mack, 1979; Smoorenburg et al, 1982; Bosman & Smoorenburg, 1995). However, the correlation between SRT in noise and the pure-tone audiogram is only modest (Smoorenburg et al, 1982). For subjects with NIHL the highest correlation coefficient was found when comparing their results of the Dutch sentence SRT test to their pure-tone average of 2 and 4 kHz (\( r = 0.72 \), Smoorenburg, 1990; Smoorenburg, 1992). When their SRT was obtained using CVC syllables this correlation was even lower (\( r = 0.59 \); Bosman & Smoorenburg, 1995).

Although subjects with a sensorineural hearing loss are more adversely affected by noise than normal-hearing subjects, studies examining patients with high-frequency hearing loss starting above 2 kHz showed similar word recognition in continuous noise as for normal-hearing listeners (Pekkarinen et al, 1990; Phillips et al, 1994). For sentence recognition in noise only small differences up to 0.5 dB were found (Festen & Smits, 2007, as reported in Rhebergen et al, 2010a). Festen & Smits (2007) used the speech intelligibility index (SII) model to predict that a sharp, infinitely deep notch of 2/3 octave width at 2 or 4 kHz would result in an increase in sentence SRT of only 2 dB. Word recognition requires only a small amount of information (Quist-Hanssen et al, 1979) and is often based on vowel recognition only (Smoorenburg, 1992). The online speech-in-noise tests contain speech materials consisting of a closed set of a small number of CVC syllables, most of which contain a unique vowel. Vowels contain speech information in the low and mid-frequencies, up to approximately 2.5 kHz. Subjects with NIHL often still exhibit normal or near-normal hearing threshold levels at these frequencies, so they may benefit from this preserved hearing for the understanding of CVC words in noise (Quist-Hanssen et al, 1979). Consequently, results of the speech-in-noise screening tests for subjects with NIHL are expected to deviate only slightly from normal performance, suggesting that the applicability of the speech-in-noise tests to detect NIHL might be low.

This study was designed to examine the performance of normal-hearing listeners and patients with different degrees of NIHL on the three Dutch online speech-in-noise tests, in order to investigate their potential to discover NIHL. For this purpose, a
multi-centre study was conducted, in which the different online screening tests were compared to pure-tone thresholds and the Dutch sentence SRT test developed by Plomp and Mimpen (1979a). The test-retest reliability and validity of the three online tests are evaluated and the sensitivity of the three tests for NIHL is determined.

**Methods**

**Participants**
The number of normal-hearing listeners and hearing-impaired participants to be tested was based on a power analysis. This showed that a sample size of 44 in each group had 80% power to detect a difference between means of 2.0 dB, using a two group t-test with a 0.05 two-sided significance level and a common standard deviation of 3.3 (Kuipers, 2006; Jongmans et al, 2008). In this calculation the following assumptions were used: the normal-hearing (NH) group and the hearing-impaired (HI) group are expected to result in different outcome categories, so the difference between groups will be at least an interval width apart. A difference of 2.0 dB, corresponding to the smallest interval width found in Occupational Earcheck, was thus considered relevant. We anticipated that only 90% of included patients will have valid measurements, therefore in total 100 participants were included, 50 in each group.

Hence, data were collected from 50 participants with normal-hearing and 50 hearing-impaired subjects, measured at three different audiology departments; LUMC Leiden, UMCN St Radboud Nijmegen and AMC Amsterdam. Only participants who were native speakers of the Dutch language were included. The normal-hearing group (33 female, 17 male) consisted of college undergraduates, recruited from the universities allied to the three university hospitals, complemented by a small number of lab workers. On the day of testing, all subjects had pure-tone thresholds of 15 dB HL or better at the octave frequencies from 0.125 to 8 kHz (including 3 and 6 kHz), except for one who was excluded from further analysis, leaving 49 subjects in this group. Their ages ranged from 18 to 50 years, with a mean age of 27.0 years (SD = 8.4 years).

Subjects in the hearing-impaired group (3 female, 47 male) were patients of one of the three ENT departments who had recently received audiological evaluation. In addition, a small number of subjects that participated in previous research concerning occupational noise exposure and hearing loss of AMC Amsterdam completed the group. Patients with normal or near-normal low-frequency hearing (pure-tone thresholds at 0.125 to 1 kHz of 20 dB HL or better) and high-frequency hearing loss (one or more pure-tone thresholds at 2 to 6 kHz greater than 25 dB HL) were selected. The included subjects had a history of noise exposure, although it is impossible to
prove a direct relationship between the hearing loss and the exposure to noise. Exclusion criterion was an air-bone gap greater than 15 dB in the tested ear. One patient did not meet the defined criteria and was excluded from further analysis, leaving 49 HI subjects for analysis with a mean age of 56.1 years (range 36-72 years, SD = 8.6 years).

A Student’s t test showed that the hearing-impaired patients were significantly older than the normal-hearing students (p < 0.001). Across the three centres only small variation in the participants’ age was observed and the NH participants tested in LUMC turned out significantly younger than the NH subjects tested in the UMCN. For more details on sub-group demographics see Table 4.1.

### Table 4.1. Demographics of the population per centre, displayed for both the normal-hearing (NH) and hearing-impaired (HI) group.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Sex (M) n (%)</th>
<th>Age mean (SD)</th>
<th>Right ear tested n (%)</th>
<th>SRT&lt;sub&gt;q&lt;/sub&gt; mean (SD)</th>
<th>SRT&lt;sub&gt;n&lt;/sub&gt; mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>AMC</td>
<td>22</td>
<td>8 (36%)</td>
<td>26.5 (6.0)</td>
<td>14 (64%)</td>
<td>26.8 (2.3)</td>
</tr>
<tr>
<td></td>
<td>LUMC</td>
<td>12</td>
<td>2 (17%)</td>
<td>21.9 (4.4)</td>
<td>8 (67%)</td>
<td>31.2 (2.6)</td>
</tr>
<tr>
<td></td>
<td>UMCN</td>
<td>15</td>
<td>6 (40%)</td>
<td>32.1 (11.1)</td>
<td>9 (60%)</td>
<td>34.6 (2.2)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>49</td>
<td>16 (33%)</td>
<td>27.0 (8.4)</td>
<td>31 (63%)</td>
<td>30.3 (4.1)</td>
</tr>
<tr>
<td>HI</td>
<td>AMC</td>
<td>10</td>
<td>9 (90%)</td>
<td>49.0 (7.8)</td>
<td>4 (40%)</td>
<td>30.7 (3.1)</td>
</tr>
<tr>
<td></td>
<td>LUMC</td>
<td>25</td>
<td>24 (96%)</td>
<td>58.9 (8.9)</td>
<td>12 (48%)</td>
<td>44.5 (8.1)</td>
</tr>
<tr>
<td></td>
<td>UMCN</td>
<td>14</td>
<td>14 (100%)</td>
<td>56.2 (6.5)</td>
<td>8 (57%)</td>
<td>42.0 (5.4)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>49</td>
<td>47 (96%)</td>
<td>56.1 (8.6)</td>
<td>25 (51%)</td>
<td>41.0 (8.4)</td>
</tr>
</tbody>
</table>

The number of males and the number of participants whose right ear was tested are displayed, percentages are in parenthesis. Also, SRTs measured by Dutch sentence SRT test in quiet (SRT<sub>q</sub>) and in stationary noise (SRT<sub>n</sub>) are shown, averaged over test and retest sessions.

Audiology departments: LUMC Leiden, UMCN St Radboud Nijmegen, and AMC Amsterdam.

The group of hearing-impaired listeners was divided into two groups of participants having either a narrow audiometric dip (HI-ND), corresponding with early NIHL, or a broad dip (HI-BD) corresponding with more severe hearing loss. Distinction was made based on whether or not their hearing threshold level at 2 kHz was affected; when hearing threshold at 2 kHz was more than 15 dB HL poorer than the pure-tone average
of the lower frequencies 0.5 and 1 kHz, the patient was classified as having a broad dip (n = 24, mean age 58.5 years, SD = 10.2 years). If not, the participant was classified as having a narrow dip (n = 25, mean age 54.1 years, SD = 7.6 years).

A power analysis, using the input as described above, shows that a sample size of 15 in each group is required to have 80% power to detect a difference between the three subject groups with a one-way ANOVA. For each of the three subject groups, mean audiometric hearing thresholds of the ears selected for monaural testing are displayed in Figure 4.1. See ‘procedure’ for more details on the criteria used to select the tested ear.

![Figure 4.1](image_url)

**Figure 4.1.** Audiometric thresholds of the ear selected for monaural testing, averaged for each of the three subject groups. Error bars represent one SD.

**Procedure**

All participants signed informed consent forms before starting the experiment. The experimental protocol and all procedures in this study were approved by the ethics committee of the University of Amsterdam (approval number: 08/049).

At the beginning of the experiment, a pure-tone audiogram was recorded at the octave frequencies of 0.125 - 8 kHz, including also 3 and 6 kHz. In addition, bone conduction was measured at 0.25, 0.5, 1, 2 and 4 kHz. The different speech-in-noise tests will be compared in case of monaural signal presentation, since OEC is designed to measure both ears separately. The test ear was chosen based on audiogram
configuration. For the normal-hearing listeners, this was either the subject’s best ear, or, in the case of no difference, the right ear. For the NIHL subjects, the ear showing the most pronounced audiometric dip was selected.

Following pure-tone threshold testing, participants performed the Dutch sentence SRT test developed by Plomp and Mimpèn (1979a). This test applies open sentences, has a very high test-retest reliability of 0.9 dB, and is used in most clinics in the Netherlands. Therefore, this Dutch sentence SRT test is considered as the clinical (or ‘gold’) standard. The results of the online tests will be compared relative to the performance on this clinical standard, in order to assess these tests’ validity. This Dutch sentence SRT test was first performed in quiet, using lists 1 and 2. The presentation level of all speech-in-noise tests needs to be approximately 20 dB higher than this SRT obtained in quiet (SRT_q), to ensure that both speech and noise are well above threshold. The noise level for all consecutive speech-in-noise tests was fixed at 65 dBA or at SRT_q + 20 dBA in cases of highly elevated SRT in quiet. Next, two lists in stationary noise were conducted. The speech stimuli recorded by a female speaker were used, and the order of the lists in noise was counterbalanced.

After finishing the Dutch sentence SRT test lists, the subject performed the three different online speech-in-noise tests. A retest was conducted, with an intermediate period of approximately 45 minutes. The sequence of test conditions was counterbalanced according to a digram-balanced Latin square, to avoid order effects. All test outcomes will be described by the term ‘speech reception threshold’ (SRT). These outcomes concern speech-in-noise test results, and therefore this SRT is defined as the SNR (in dB) required to correctly recognize 50% of the presented speech stimuli rather than an absolute threshold level.

**Equipment and set-up**

All audiometric testing and the majority of the speech-in-noise tests were carried out in a sound-insulated booth. However, no internet access was available in the audiologic booth in LUMC and online speech-in-noise measurements were carried out in a quiet room, without interfering noise. Pure-tone audiometry was administered using a Decos (AMC, LUMC) or Interacoustics (UMCN) clinical audiometer and TDH-39 headphones. Calibration of hearing levels was done according to ISO-389.1 (1998).

For the speech-in-noise tests, signals were presented via a standard soundcard (Gina 24/96) on a PC at a sample frequency of 44.1 kHz and were fed through a TDT.
headphone buffer (HB6) and a TDT programmable attenuator (PA4). In the UMCN, signals were fed through an AC-40 clinical audiometer. To ensure standardized and controlled testing conditions, participants received the signals via headphones, at a fixed noise level. The Dutch sentence SRT test is presented through TDH-39 headphones. After finishing this test, the subjects performed the different online speech-in-noise tests, using Sennheiser HDA-200 headphones in an otherwise identical test set-up. The participant was seated in front of a computer touch screen to enter test responses. The noise levels of each test were calibrated with a B&K sound level meter 2260 and a B&K type 4153 artificial ear, with the use of a flat-plate adaptor.

Speech-in-noise test stimuli

Dutch sentence SRT test
The Dutch sentence SRT test (Plomp & Mimpen, 1979a) comprises ten lists of thirteen short meaningful Dutch sentences containing eight or nine syllables each. These sentences are presented either in quiet or in masking noise with a spectrum that matches the long-term average spectrum of the speech. A simple up-down procedure was used to estimate the SRT. Noise level was fixed at a minimum of 65 dBA and the SNR was varied adaptively by changing the speech level. The first sentence was presented at a level below SRT that was gradually increased with 4 dB steps until the sentence was reproduced entirely correct. The level of each consecutive sentence depended on the accuracy of the response to the previous sentence; this was decreased by 2 dB after a correct response and increased by 2 dB after an incorrect response. An errorless reproduction of the entire sentence was required for a correct response. The SRT is calculated as the average SNR over sentence 5 to 13, plus the SNR of a virtual 14th sentence determined from the response to the previous sentence.

Online speech-in-noise tests
The three different online speech-in-noise tests, National Hearing Test, Earcheck and Occupational Earcheck, are based on the intelligibility of speech in stationary masking noise. For each test, the spectrum of the noise is matched to the long-term average spectrum of the speech material used and the root mean square (RMS) level of the noise was scaled to match that of the speech. Speech and noise files are stored in MP3 format and a Macromedia flash player (Macromedia Inc., San Francisco, USA) web application is used to mathematically mix the SNRs of the speech and noise files.

Normally the online tests are performed at a level that is most comfortable and loud enough for the respondent; prior to performing the test, a stimulus is presented without noise and participants are instructed to adjust volume to a level where the stimulus is clearly understood. Next, test instructions are presented on screen. Respondents are recommended to perform the test by headphones, but PC
loudspeakers can also be used for diotic testing. Then the speech stimuli are presented once for familiarization of the stimuli and their corresponding buttons, and then the test starts. Immediately after finishing the test, the test result is shown; the SRT results are classified into categories of hearing ability, e.g. ‘good’, ‘insufficient’ or ‘poor’ hearing, which is presented to the respondent accompanied by appropriate advise. For reasons of time and standardization, all this is eliminated from the experimental set-up and the test is performed at a fixed presentation level, starts immediately after entering the participant’s ID code, and test results are presented as SRT values in dB. All tests are performed according to the up-down procedure with the noise level fixed at a minimum of 65 dBA and the presentation level of speech stimuli varying adaptively with a 2 dB step size, as described above. The signal-to-noise level of the first presentation was fixed at 0 dB. The SNRs presented ranged from –14 to +4 dB. Participants were instructed to listen and enter their response using the buttons on the computer screen. Although these online speech-in-noise tests roughly follow the same test principles, there are some differences between the three versions regarding the procedure and stimuli used, which are described below and summarized in Table 4.2.

**National Hearing Test**

The National Hearing test (NHT) was developed by Smits et al (2004) as an automatic speech-in-noise test to be performed by telephone, or by internet (www.hoortest.nl) (Smits et al, 2006a). Dutch monosyllabic digits are used to construct a set of 80 different digit triplets. A series of 23 triplets is randomly chosen from this set, and SRT is calculated by averaging the SNRs of triplets 5 to 23 and the virtual 24th one, based on the last response. A response is considered correct only when all three digits are identified correctly. The test results are categorized into three classes: ‘good’ hearing (SRT ≤ -5.5 dB), ‘insufficient’ hearing (-5.5 dB < SRT < -2.8 dB) and ‘poor’ hearing (SRT ≥ -2.8 dB). For reasons of comparison to the telephone version of the test, the signals of the internet version simulated the characteristics of the telephone network and were bandwidth limited to 0.3 - 3.4 kHz.

**Earcheck**

The Earcheck (EC; ‘Oorcheck’ in Dutch) is a screening test that specifically aims at young persons aged between 12 and 24 years (www.oorcheck.nl). The recorded test files have no bandwidth limitations using both broadband speech and noise, covering the full bandwidth up to 16 kHz. Speech material consists of nine different Dutch CVC syllables, randomly presented three times each. The words are chosen from the Dutch wordlist used for diagnostic speech audiometry (Bosman, 1989), with a phonemic distribution representative for the Dutch language. Consequently, the nine words all contain unique vowels (rat /rαt/, thumb /dœym/, goat /xεit/, chicken /kɪp/, fire /vyr/, lion /lew/, cat /pus/, saw /zax/, and wheel /wil/).
On screen, nine response buttons containing a written representation of the words and a corresponding picture are shown. A tenth button saying ‘not recognized’, is added to prevent respondents from guessing. The SRT is calculated by averaging the SNRs of presentations 8 to 27. The test results are classified in four categories of hearing status. The category ‘good’ hearing corresponds to a SRT of -10 dB or less, and a ‘moderate’ hearing status corresponds to a SRT that is between -10 dB and -7 dB. SRT results lying between -7 dB and -4 dB are categorized as ‘insufficient’ hearing status and a SRT result of -4 dB or higher corresponds to ‘poor’ hearing (Albrecht et al, 2005). The cut-off values defined to classify the SRT results into these different hearing status categories derived from a validation experiment (Martens et al, 2005).

**Occupational Earcheck**

The Occupational Earcheck (OEC; ‘Bedrijfsoocheck’ in Dutch) is a speech-in-noise screening test, developed for use in occupational hearing conservation (www.bedrijfsoocheck.nl). The test is similar to EC in that it has no bandwidth limitations and that
The speech material comprises nine Dutch CVC syllables, represented by nine response buttons and a tenth one labelled ‘not recognized’. However, the speech stimuli are different; although the words are chosen from the Dutch wordlist used for diagnostic speech audiometry as well (Bosman, 1989), the speech set is specifically selected to contain a higher proportion of high-frequency consonants and to include only five different vowels (bed /bɛt/, knife /mɛs/, bag /tæs/, pan /pæn/, cat /pus/, book /buk/, sock /sɔk/, sun /sʌn/, arrow /pɛil/). The words are randomly presented four times each, resulting in a total set of 35 stimuli (Ellis et al, 2006).

The first presentation is at a signal-to-noise ratio of 0 dB, and with every correct response to the subsequent stimulus speech level is attenuated by 2 dB. The actual test starts at the SNR of the first incorrect response, so the starting level is set individually. The SRT is calculated by averaging the SNRs of stimuli 6 to 35 and the test results are classified into five, smaller, categories: ‘good’ hearing (SRT ≤ -10 dB), ‘moderate’ hearing (-10 dB < SRT ≤ -8 dB), ‘insufficient’ hearing (-8 dB < SRT ≤ -6 dB), ‘poor’ hearing (-6 < SRT ≤ -4 dB) and an extra fifth category ‘very poor’ hearing (SRT > -4 dB) (Kuipers, 2007).

The performance of the three subject groups on these online tests will be compared to assess the sensitivity of the tests for NIHL detection. The validity of the online test results will be assessed by comparing the performance on the screening tests to the performance on Dutch sentence SRT test and to pure-tone thresholds. In addition, online test reliability is assessed and compared between the different tests, by analysing the test and retest results.

**Results**

In total, 98 subjects completed the experiment. The number of participants in each centre, demographic characteristics and outcomes of the Dutch sentence SRT test were shown in Table 4.1 for both the normal-hearing and the hearing-impaired listeners. A one-way ANOVA showed small significant differences between the NH participants of the three centres for sentence SRT ($F_{2,95} = 100.60, p < 0.001$). This is probably associated with significant differences in hearing thresholds; normal-hearing subjects tested in the AMC had better threshold levels at the lower frequencies (0.125 – 1 kHz) than subjects tested in the other centres (repeated measures ANOVA for all frequencies $F_{1,2} = 8.16, p = 0.001$). There are no significant centre differences in the normal-hearing Dutch sentence SRT results in noise ($F_{2,95} = 2.47, p = 0.096$). The SRTs obtained are in excellent agreement with the normal-hearing average of –5.5 dB reported by Plomp and Mimpfen (1979a). Since there is no difference in sentence SRT in noise between the
normal-hearing measurements in the three centres, the participants tested in the
different centres may be considered as belonging to the same population.
Consequently, no further distinction between centres is made and all participants are
analysed together within each of the three subject groups.

**Mean SRTs and reliability of the online screening tests**
The three internet-based speech-in-noise tests were evaluated by performing them
twice. The mean test and retest results of the three subject groups are shown in Table
4.3. Test and retest differences and correlations are analysed in terms of test reliability,
which is compared across the three online tests. These analyses are conducted using
data of all 98 participants together.

<table>
<thead>
<tr>
<th>Test</th>
<th>Subject group</th>
<th>Test 1 mean (SD)</th>
<th>Test 2 mean (SD)</th>
<th>Mean Δ test-retest</th>
<th>SEM</th>
<th>ICC test-retest</th>
<th>Slope (%/dB)</th>
<th>Se for NIHL (%)</th>
<th>Sp for NIHL (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>NH</td>
<td>-12.0 (1.7)</td>
<td>-12.6 (1.5)</td>
<td>0.82 (p&lt;0.001)</td>
<td>1.24</td>
<td>0.75</td>
<td>13.0</td>
<td>51</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-10.7 (3.0)</td>
<td>-11.4 (2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-8.4 (2.4)</td>
<td>-9.6 (2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OEC</td>
<td>NH</td>
<td>-9.6 (1.4)</td>
<td>-9.7 (1.7)</td>
<td>0.45 (p=0.015)</td>
<td>1.26</td>
<td>0.68</td>
<td>11.0</td>
<td>92</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-7.6 (2.0)</td>
<td>-7.9 (2.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-5.9 (2.7)</td>
<td>-7.2 (1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHT</td>
<td>NH</td>
<td>-7.5 (1.2)</td>
<td>-7.9 (0.8)</td>
<td>0.29 (p=0.097)</td>
<td>1.20</td>
<td>0.79</td>
<td>13.4</td>
<td>53</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-6.3 (1.7)</td>
<td>-6.7 (2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-2.9 (2.5)</td>
<td>-2.8 (2.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Displayed ICC’s are Two way Mixed model, Type Consistency, Single measures ICC’s. Also, sensitivity (Se) and specificity (Sp) for the detection of NIHL for each of the three online speech-in-noise tests are included.

**Learning effect**
First the difference in SRT over the two test sessions is calculated. This difference
estimates a systematic change in performance due to learning or fatigue. Averaged
across all participants, the performance on the second test was better than on the
first, for all online tests. Nevertheless these differences between the SRTs for the test
pairs were small, ranging from 0.29 dB (NHT) to 0.82 dB (EC). A paired-sample t-test revealed that the test-retest difference was significant for EC (p < 0.001) and OEC (p = 0.015), suggesting a small learning effect for these tests. These effects however, are of a magnitude that is smaller than inter-individual differences and thus are not particularly relevant for clinical purposes. NHT did not show a significant learning effect (p = 0.097).

**Test-retest reliability**

Reliability can be defined as the consistency of a test’s results across series of observations. A measure to express this is the standard error of measurement (SEM), that is calculated dividing the within subject standard deviation of the differences by \( \sqrt{2} \). A speech-in-noise test can only differentiate between subjects with different degrees of hearing loss if the SEM is small. The test-retest reliability of all three tests turns out to be comparable; SEM values ranged from 1.20 dB for NHT to 1.26 dB for OEC. In addition, test reliability can be expressed by the intraclass correlation coefficient (ICC), which is also comparable across the three tests (0.68 – 0.79). In correspondence with the slightly lower SEM, the ICC is highest for the NHT (Table 4.3).

**Intelligibility functions**

Another important test characteristic is the slope of its psychometric function. Steeper intelligibility functions result in more precise speech-in-noise tests, with greater discriminative power. The steepness of the performance intensity (PI) function indicates the rate at which speech information becomes intelligible with increasing signal-to-noise ratio. Although the adaptive procedure only yields SRTs, PI functions can be estimated based on a fit of the proportions correct at different presentation levels. The SNR of every presentation was corrected by the individual SRT for that test, in order to correct for inter-individual differences. Then for each SNR the proportion correct was calculated. For the different speech-in-noise tests, the intelligibility scores per corrected SNR were obtained from pooled data of both test sessions. A test-specific psychometric function was fitted to the data, using the following logistic regression function:

\[
P(SNR) = \gamma + (1 - \gamma) \frac{1}{1 + e^{\left[-(SNR - SRT) \cdot 4s\right]}}
\]

(Equation 4.1)

where \( p \) is the proportion correct at a given signal-to-noise ratio, \( \gamma \) is guess level, and \( s \) represents the slope of the psychometric function at SRT. In the case of EC and OEC that both use nine different stimuli guess level \( \gamma \) is set at 0.11. The NHT is based on correct triplet recognition and thus its guess level is around zero. The mean intelligibility functions obtained for the EC, OEC and NHT are shown in Figure 4.2, in separate
plots for NH and HI listeners. There was no further distinction of the hearing impaired subjects in ND or BD groups; to assure a sufficient amount of measurements in each group, intelligibility functions are based on pooled data of all 49 HI participants.

The three online tests show similar slopes, the exact slopes are reported in Table 4.3. For all tests, slopes are slightly steeper for normal-hearing subjects than for hearing-impaired listeners. The slope of the OEC curve is somewhat shallower than the slope of the other two online tests, despite the specific test properties chosen to improve the test’s precision (e.g. selection of speech material, higher number of stimuli). The PI functions of the Dutch sentence SRT test are displayed for comparison. These functions were much steeper, with slopes of 30%/dB in the normal-hearing group and 28.8%/dB in the hearing-impaired listeners.

![Figure 4.2. PI functions of the three different online tests and the Dutch sentence SRT test, showing proportion correct as function of the signal-to-noise ratio of the presentations. The presented SNR is relative to the individual SRT and then corrected to the average test SRT. To assure a sufficient amount of measurements in the HI group, no further distinction of the hearing impaired subjects in ND or BD groups was made, and intelligibility functions are based on pooled data of all 49 HI participants.](image-url)
Relationship between SRTs results and the pure-tone audiogram

In the following analyses only results of the first test session were taken into account, as this is representative for real-life test performance. Retest results showed a slightly lower (thus better) SRT due to training effect, and thus are not representative for respondents performing the test only once.

First, mutual Pearson correlation coefficients are calculated, using results of the total study population, in order to establish the amount of association between the different tests. All correlations are significant (p < 0.001), and are shown in Table 4.4.

Test validity relates to the correlation between the test’s results and other, accurate, measures of the same behaviour. In order to assess the validity of the online speech-in-noise tests, their results are compared to the results of the Dutch sentence SRT test. All online tests are significantly correlated with this test (p < 0.001). The correlation coefficients show moderate association (Table 4.4), and are quite similar for the three tests, ranging from 0.65 (EC) to 0.77 (NHT). In Figure 4.3, bivariate scatterplots are given, presenting the 50% points of the three online screening tests versus results of the Dutch sentence SRT test. In each panel, datum points of the listeners with normal-hearing are in closer proximity to one another, whereas the datum points from the listeners with hearing loss show a wider distribution, reflecting a larger inter-subject variability. Excluding all data of the normal-hearing participants yielded correlation coefficients of the same magnitude as those presented in Table 4.4.

Correlations of the online test results and pure-tone thresholds are analysed in the same way. For this purpose, three pure-tone averages (PTA) are calculated, concerning the average over the frequencies 0.5, 1, 2 and 4 kHz that are important for speech intelligibility (PTA_{0.5,1,2,4}); the average of the thresholds at 2 and 4 kHz, which yielded the strongest correlation with results of the Dutch sentence SRT test in studies of Smoorenburg (1990; 1992) (PTA_{2,4}); and the PTA of the higher, noise-sensitive,
frequencies 3, 4 and 6 kHz (PTA$_{3,4,6}$). Correlation coefficients between these pure-tone averages and online test results are statistically significant (p < 0.001) yet lower than the correlation coefficients found comparing PTA-values to the Dutch sentence SRT (Table 4.4). The correlation coefficients for the three tests do not differ much, but $r$ is highest for NHT and lowest for EC. Correlation is highest when SRTs are compared to PTA$_{0.5,1,2,4}$ and becomes slightly lower when online test results are compared to PTA$_{3,4,6}$.

In Figure 4.3, also bivariate scatterplots of the SRT results of the three online tests versus PTA$_{0.5,1,2,4}$ are displayed.

**Group differences between tests**

Finally, the results for NH listeners and both groups of HI participants with different degrees of NIHL are compared. This is done in order to establish the amount of separation in recognition performance measured with each test and to obtain the tests’ sensitivity to identify subjects with NIHL. Again, only the speech reception thresholds of the first test were used for this evaluation.

A one-way ANOVA of the individual SRT results of each test is performed to investigate differences between the subject groups. This shows that the main effect of listener group is significantly different (EC: F[2,95] = 20.56, p < 0.001, OEC: F[2,95] = 32.04, p < 0.001, NHT: F[2,95] = 60.01, p < 0.001). Post-hoc t-tests with Bonferroni corrections for multiple comparisons show small but significant differences between all three subject groups, except for the EC; normal-hearing EC results do not differ from EC results of the hearing-impaired subjects with a narrow dip (p = 0.079).

Boxplots representing the obtained SRT results for each subject group are displayed in Figure 4.4. The boxes represent the inter-quartile range, the median is demonstrated by the vertical line inside the box and error bars represent the 5th and 95th percentile of the results in each group. The dashed lines in Figure 4.4 correspond to hearing status categories as defined for the different tests. The overlap between normal-hearing listeners and subjects with narrow audiometric dips in the EC results is evident, since a high proportion of the mildly impaired participants are classified as having ‘good’ hearing. Although the group differences are statistically significant, NHT results are distributed similarly. OEC displays a somewhat better separation between the normal-hearing and hearing-impaired listeners. This test correctly classified the majority of the hearing-impaired listeners, while only half of the normal-hearing subjects fall into the ‘good’ hearing category.

**Sensitivity and specificity for NIHL**

In order to investigate how well the different speech-in-noise tests discriminate between NH and HI respondents, SRT results and the cut-off values defined for each test were used to calculate their sensitivity and specificity. Test sensitivity refers to the percentage of HI participants classified correctly as having hearing ability worse than
Figure 4.3. Bivariate scatterplots of SRT measured with each of the online speech-in-noise tests against sentence SRT in noise and PTA, separated for the three subject groups; NH listeners (black), HI subjects with narrow noise dips (grey) and HI subjects with broad noise dips (white).
‘good’; test specificity refers to the number of normal-hearing subjects correctly classified as having ‘good’ hearing. The results for each of the three screening tests are shown in Table 4.3. The highest test sensitivity for NIHL was found for OEC, almost all HI participants were classified as ‘moderate’, ‘insufficient’ or ‘poor’. Yet, this accounted for the majority of the normal-hearing group as well (Figure 4.4), resulting in a very low specificity of 49%, meaning that half of the NH subjects is also classified as being hearing impaired. The two other tests show, in spite of a high specificity of 94% for NHT and 90 % for EC, low sensitivity values of 55% and 51% respectively. This means that a large number of the subjects with NIHL, especially those with narrow dips, will be classified as being normal-hearing when performing NHT or EC.

**Figure 4.4.** Boxplots of SRTs of the different online speech-in-noise tests, for NH listeners (dark grey), HI listeners with a narrow dip (gray) and HI listeners with a broad dip (light grey). Error bars show 5th and 95th percentile of SRT values. Vertical dashed lines indicate the outcome categories of each test, ranging from good, on the left, through moderate (mod) and insufficient (insuf), to (very) poor, on the right of each plot.
Discussion

Noise-induced hearing loss is a highly prevalent public-health problem that is irreversible yet preventable. Increased public awareness about NIHL and early detection of hearing loss in groups at risk can help to prevent the development of NIHL. Since monitoring by pure-tone audiometry, as applied in an occupational setting, is not easily accessible for a broad population, alternative ways of monitoring, e.g. speech-in-noise testing, appear attractive, especially when such tests can be conducted through the internet. A speech-in-noise test has the advantage that it is independent of presentation level and less sensitive to background noise, making it suitable for use as a self-test at a remote test site (e.g. at home), in less well-controlled conditions.

An existing Dutch online screening test, the National Hearing Test (Smits et al, 2004; 2006a), has been adapted to serve as a screening instrument in either a population of occupationally exposed workers (Occupational Earcheck) or a young population exposed to leisure noise (Earcheck). In both populations noise-induced hearing loss is expected to be prevalent, and the screening tests could be of great significance in the identification of this NIHL. However, this study was conducted to investigate the value of these tests in discovering early noise-induced hearing loss and shows that their sensitivity to detect NIHL is only low.

Age effects

The normal-hearing and hearing-impaired subjects tested in this study are not well matched with respect to their age; the HI listeners were on average older than the NH students. Since speech-in-noise recognition becomes poorer with increasing age, especially when age is above 50 years (Plomp & Mimpen, 1979b), the two subject groups differ in more ways than just hearing status. The HI subjects measured in this study had a mean age of 56 years that, according to the data presented by Plomp & Mimpen (1979b), corresponds to a median SRT of only 1 dB poorer than the median SRT for the young NH subjects. Moreover, Van Rooij & Plomp (1992) showed that speech recognition in the elderly was influenced mainly by the progressive high-frequency hearing loss that develops with age, rather than by a decrement in cognitive performance. This made us assume that the difference in age did not strongly affect our results, and this assumption is supported by the fact that the differences in SRT between the subject groups are actually smaller than expected despite the older HI subjects.

Test reliability

The online speech-in-noise screening tests did show reliable results in this population of normal-hearing listeners and patients with NIHL. The test-retest reliability is mainly determined by the slope of the psychometric function; steeper slopes mean that a
small change in SNR would result in a large change in performance. The slopes of intelligibility functions of both EC (13.0%/dB) and NHT (13.4%/dB) turned out to be comparable and slightly higher than the slope obtained for OEC (11.0%/dB). These slopes are considerably steeper than the slopes of the CVC intelligibility functions reported by Bosman & Smoorenburg (1995), who found a slope of 8.1%/dB for normal-hearing listeners and a similar slope of 7.8%/dB for a group of 20 subjects with NIHL, probably because their test used an open response format. However, slopes reported for similar speech-in-noise screening tests are somewhat steeper; Smits et al. (2004) obtained psychometric functions of NHT results of 10 normal-hearing subjects with slopes of 16%/dB using headphones and 20%/dB using telephone. The French digit triplet test by telephone also reached a higher slope, of 17.1%/dB (Jansen et al, 2010).

Comparing the test and retest results yielded high intraclass correlation coefficients of 0.68 – 0.79 and standard errors of measurement that are relatively small: around 1.2 dB. The reliability of the National Hearing Test is examined earlier by a series of studies (Smits et al, 2004; Smits & Houtgast, 2005; Smits & Houtgast, 2007). These investigations reported measurement errors ranging from 0.9 to 1.1 dB obtained in a normal-hearing population, which are slightly better than the SEM found in this study.

In addition, a significant test-retest difference was obtained for OEC and EC, suggesting a small learning effect. However, these effects were only 0.5 dB and 0.8 dB respectively and since this is smaller than the measurement error they can be considered as not clinically relevant. These results indicate that the three tests are reliable in measuring speech reception in noise. Reliability measures showed comparable results over the three online tests, but all results turned out to be slightly poorer for OEC. This is unexpected, since this test was specifically designed to have a higher precision.

**Test validity**

Previous studies reported moderately high correlations coefficients of 0.72 – 0.84 between results of the Dutch sentence SRT test and pure-tone averages for normal-hearing and hearing-impaired listeners (Smoorenburg, 1992; Bosman & Smoorenburg, 1995; Smits et al, 2004). The findings of this study show similar correlations; r is 0.82 comparing the results of the Dutch sentence SRT test with both PTA_{0.5,1,2,4} and PTA_{2,4}.

The online test results showed lower correlation coefficients ranging from 0.66 for EC results to 0.72 for NHT results when compared to PTA_{0.5,1,2,4}. This value for NHT is comparable to the correlation of 0.77 reported for both the French and the Dutch digit triplet in noise test and PTA_{0.5,1,2,4} (Jansen et al, 2010; Smits et al, 2004). The latter one also found a correlation for NHT with the Dutch sentence SRT test of 0.87, that was higher than found in this study (r = 0.77). All correlations found between the SRTs and PTA_{3,4,6} are slightly lower than found for the other PTA-values.
Results of the online tests
The online tests thus prove to yield reliable results, but are not strongly related to pure-tone thresholds and the Dutch sentence SRT results in this population. More importantly, the results of this evaluation study show that the speech-in-noise tests differentiated only to a limited degree between speech intelligibility of participants with noise-induced hearing loss and normal hearing. Normal-hearing participants do reach – on average - lower SRTs than hearing-impaired listeners but the differences between normal speech reception and SRTs of listeners with a narrow audiometric dip are only small or, in case of EC, even statistically insignificant. Consequently, the sensitivity of the tests to discover relatively mild high-frequency hearing losses appears to be rather low. This is reflected in the sensitivity and specificity results; NHT and EC yield a sensitivity of 55% and 51% respectively in this study population, meaning that both tests classify almost half of the listeners with NIHL incorrectly as normal-hearing listeners. On the other hand, the specificity of these two tests is high; 94% and 90% respectively. This is in contrast with the OEC that yield a high test sensitivity of 92% in this population but shows low specificity of 49%, incorrectly classifying the majority of the normal-hearing listeners. Apparently the additional fifth category had led to cut-off values that are too high to classify normal-hearing listeners as having a ‘good’ hearing.

The sensitivity and specificity of a test depend critically on the cut-off value that is used to distinguish between normal and impaired (hearing) performance. A change in the cut-off values used in EC and NHT will lead to a higher number of correctly classified hearing-impaired subjects, but this will go along with poorer specificity.

Current findings thus show that subjects with NIHL, especially with a narrow audiometric dip, will perform quite similar to normal-hearing participants on internet-based speech-in-noise tests, probably because they can benefit from their preserved hearing in the low and mid frequencies. This implies that the current tests for auditory screening and monitoring via internet, although proven to be a valuable tool for screening purposes in a general population, have a limited applicability for populations in which noise-induced hearing loss is prevalent, e.g. in occupational health care and in prevention programs for young persons. In fact, these results do not only concern respondents with NIHL, but may also be generalized to individuals with a high-frequency sensorineural hearing loss, regardless of the etiology.

Future research
Although low sensitivity for NIHL was expected, this is a disadvantageous finding, since early self-identification of hearing loss may result in increased awareness and appropriate audiological follow-up for those affected, thereby preventing NIHL. It would be very useful to investigate ways to improve sensitivity of these tests to
discover high-frequency hearing loss. In order to do so, possible adaptations involving either the speech material or the masking noise used, might be considered. For the speech stimuli, words seems to be the best choice, since a self-test should be quick, easy and automated. We hypothesized that listeners with NIHL mainly rely on vowel recognition for their identification of CVC words in noise, especially in the small closed sets of stimuli that are presented. If this is true, one would expect a greater deviation between EC and OEC results, as the latter presents words containing high frequency consonants and only five different vowels, increasing the auditory similarity of the stimuli and forcing the listener to use high-frequency information to discriminate between the words. Although the difference in SRT results of OEC between NH listeners and HI subjects with a narrow dip is slightly greater than found for the EC, there is no overall difference between these tests.

In addition, the use of different types of noise, such as filtered or interrupted noise, may discriminate better between normal and impaired performance, increasing the validity of the tests for the detection of NIHL. This will be considered in a follow-up study, in which we aim to make the tests more appropriate for screening on (early) NIHL, described in Chapter 5.

**Conclusion**

In this study three online Dutch speech-in-noise screening tests were evaluated in a population of normal-hearing listeners and participants with different degrees of noise-induced hearing loss, concerning their sensitivity for detecting NIHL. The tests showed reliable results, although correlations with Dutch sentence SRTs and pure-tone thresholds were moderate. SRT results of subjects with mild NIHL deviated only slightly from normal performance. Consequently, the sensitivity of the tests to discover high-frequency hearing loss is low, and the tests in their current form are not appropriate to be used for screening of NIHL in an early stage.

Since these online screening tests can play an important role in the prevention of NIHL possible adaptations leading to an improvement in test sensitivity for NIHL are to be investigated, in order to obtain a valid screening tool for high-frequency hearing loss detection. This will be described in Chapter 5.