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: http://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.

M.C.J. Leensen¹
J.A.P.M. de Laat²
A.F.M. Snik³
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
³ Department of Otorhinolaryngology, Radboud University Medical Centre, Nijmegen, The Netherlands

Abstract

Objective: An easily accessible screening test can be valuable in the prevention of noise-induced hearing loss (NIHL). The Dutch National Hearing Foundation developed ‘Earcheck’; an internet-based speech-in-noise test, presenting CVC words in stationary broadband noise. However, its sensitivity to detect NIHL appeared to be low, 51% (Chapter 4). The aim of the current study is to examine ways to improve Earcheck’s sensitivity for (early) NIHL using different forms of noise filtering.

Design: The test’s stationary broadband masking noise is replaced by six alternatives, including noises that have been temporally modulated, spectrally filtered by high-pass or low-pass filters, and combinations of temporal modulation and spectral filtering.

Study Sample: In this multi-centre study, 49 normal-hearing and 49 subjects with different degrees of NIHL participated.

Results: Hearing-impaired subjects deviated more clearly from normal performance when executing the test with alternative masking noises, except for the high-pass filtered conditions. Earcheck with low-pass filtered noise made the best distinction between normal hearing and NIHL, without reducing test reliability. The use of this noise condition improved the sensitivity of Earcheck to 95%.

Conclusion: The use of low-pass filtered masking noise makes speech-in-noise tests more sensitive to detect NIHL in an early stage.
Introduction

Despite the fact that noise-induced hearing loss (NIHL) is preventable, it is still a highly prevalent public-health problem in modern society. NIHL is not only the most reported occupational disease in the Netherlands (Van der Molen, 2010), but it is also a growing concern in the general public, due to the increasing exposure to recreational noise. Young people especially are considered to be at risk of developing NIHL, exposing themselves to potentially damaging loud music when attending discotheques and live concerts or when listening to personal music players. Noise levels during these recreational activities are high and often exceed the occupational limit of 80 dBA set for an 8-hour working day, defined in the European Directive 2003/10/EC (EPC, 2003). Vogel et al. (2010) estimated that more than half of the 1512 adolescents participating in their study exceeded this occupational standard by listening to high volume music. Although research concerning the prevalence of hearing loss caused by leisure noise in youngsters demonstrated inconsistent results (Meyer-Bisch, 1996; Mostafapour et al, 1998; Niskar et al, 2001; Biassioni et al, 2005; Shah et al, 2009; Zhao et al, 2010; Shargorodsky et al, 2010), the reported average sound levels of these activities, ranging from 80 dBA to 115 dBA (SCENIHR, 2008), are high enough to pose a risk to hearing. This is particularly true for individuals being exposed for longer periods, and for young people involved in multiple noisy recreational activities or additionally exposed to occupational noise, resulting in cumulative effects that may lead to an increased prevalence of hearing loss (Torre III, 2008).

Noise-induced hearing loss develops gradually and is often unnoticed until the damage is substantial and severe enough to be measured (Shah et al, 2009). Therefore, the risk of hearing loss is easily underestimated (Vogel et al, 2008). Furthermore, there is a great deal of misconception and unawareness among youngsters about the impact of hearing loss and the effect of overexposure to loud music in general (Chung et al, 2005; Vogel et al, 2008; Vogel et al, 2009; Shah et al, 2009). Adolescents first must become aware that listening to high-volume music may cause hearing damage and that they personally are at risk for hearing loss before the promotion of protective behaviours is useful (Vogel et al, 2009). In addition, self-experienced symptoms after recreational noise exposure might lead to greater awareness (Widen et al, 2009), which can change personal listening behaviour in order to protect hearing. Moreover, if hearing deterioration can be shown at an early stage actions can be taken to prevent further hearing loss (Meyer-Bish, 1996).

An objective hearing screening test that can detect hearing loss in an earlier stage can be of great help in preventing NIHL and raises awareness of possible hearing problems after music exposure (Koopman et al, 2008). Since subjects with NIHL often complain about a reduced ability to understand speech in noisy situations, a speech-in-noise test seems a suitable measure to detect this kind of hearing loss.
In The Netherlands an internet-based speech-in-noise test was implemented, as a screening tool for adolescents exposed to leisure noise. This test, “Earcheck” (Oorcheck in Dutch, www.oorcheck.nl), has been developed by the Dutch National Hearing Foundation and the LUMC Leiden (Albrecht et al, 2005). The test principles are derived from the National Hearing Test (Smits et al, 2004; 2006a). This test is bandwidth limited, whereas Earcheck covers the full bandwidth up to 16 kHz. Earcheck is incorporated in a special educational website aiming at adolescents and young adults in the age range 12 - 24 years, facilitating early NIHL identification and increased awareness about the risks of noise exposure.

The test presents a closed set of nine different CVC words against a background of stationary speech-shaped noise. The test uses an adaptive up-down procedure corresponding to the one described by Plomp and Mimpen (1979a), to assess the speech reception threshold (SRT), i.e. the signal-to-noise ratio (SNR) required to recognize 50% of the speech correctly. A total of 27 stimuli are randomly presented, and the arithmetic average of the SNRs of the last 20 presentations results in the SRT. The Earcheck outcomes are classified into four categories of hearing status, accompanied by an appropriate advice for referral. This self-screening test is easy to administer and takes about three minutes to perform.

Speech-in-noise tests, such as Earcheck measure the speech reception threshold in a stationary noise with the same long-term average spectrum as the speech material used. Because this makes the test independent of absolute presentation level (Plomp, 1986) and of variations in equipment used (Smits et al, 2004; Culling et al, 2005), it is considered to be suitable for online screening purposes. Furthermore, the test is robust against background noise (Jansen et al, 2010), resulting in a test that is reliably administered in an at-home setting (Smits et al, 2004). However, the evaluation study described in Chapter 4 showed that the currently implemented Earcheck was not able to make a clear distinction between normal-hearing listeners and participants with different degrees of NIHL. Although Earcheck demonstrated fairly good test reliability, with a test-retest standard deviation (SD) of 1.2 dB and an intraclass correlation coefficient of 0.75, the test sensitivity for NIHL turned out to be rather low; only 51% compared to the results of the clinical audiogram. This means that half of the NIHL patients were (wrongly) classified as having normal hearing by Earcheck.

Subjects with NIHL exhibit poorer hearing thresholds in the higher frequencies, while thresholds in the lower frequency region remain (nearly) normal. These individuals could be benefitting from their intact low frequency hearing (Quist-Hanssen et al, 1979), by mainly relying on vowel recognition to identify CVC words in noise (Smoorenburg, 1992), especially if a closed set of stimuli is used. Consequently, Earcheck only demonstrated small differences between normal speech reception and the SRTs of listeners with early NIHL, resulting in low sensitivity to discover relatively
mild high-frequency hearing losses (see Chapter 4). Since an adequate screening test can be of major importance in the prevention of NIHL, current study examines possible ways to improve Earcheck’s sensitivity to discover (early) high-frequency hearing loss.

The sensitivity of a test is high when it clearly distinguishes between normal-hearing and hearing-impaired listeners (Theunissen et al, 2009). The difference in speech reception between normal-hearing and hearing-impaired listeners varies greatly depending on the nature of the interfering noise. Certain types of maskers may yield more information than a steady-state background. It is well known that listeners with normal hearing sensitivity perform much better when the masking noise is interrupted than when it is stationary (Festen & Plomp, 1990; Phillips et al, 1994; Stuart & Phillips, 1996; Bacon et al, 1998). They take advantage of the relatively high SNR in the silent periods of the interfering noise to extract speech information, in order to achieve higher performance than with a stationary masker. This is called masking release. Conversely, hearing-impaired listeners experience little or no benefit when going from stationary noise to fluctuating noise, even when the hearing loss is mild and more or less restricted to high audiometric frequencies (Phillips et al, 1994; Stuart & Phillips, 1996; Bacon et al, 1998; Versfeld & Dreschler, 2002). The SNR improvement during the gaps in the noise is limited by their elevated thresholds. In addition, they generally show reduced temporal resolution and degraded recovery from forward masking, preventing them from taking full advantage of dips in the masking noise.

For sentence intelligibility, reported differences between normal-hearing and hearing-impaired listeners in interrupted noise are in the range of 7 to 15 dB compared to differences ranging from 2 to 5 dB in stationary noise (Peters et al, 1998). Also previous studies examining word recognition of subjects with NIHL in stationary and interrupted noise only demonstrated significant differences in performance relative to controls in interrupted noise conditions (Phillips et al, 1994; Stuart & Phillips, 1996). Spectral properties of the speech signal and the competing background noise affect the results of a speech-in-noise test as well. Normal-hearing speech reception in noise improved when spectral dips were added to the interfering noise, and this improvement increased as the width of these spectral dips increased (Peters et al, 1998). Hearing-impaired subjects showed a much smaller improvement, indicating reduced audibility for speech in noise with lower intensity. A spectrally filtered noise can also be used to improve discrimination between respondents with NIHL and normal-hearing listeners. Since NIHL affects the higher frequency region, a low-pass filtered masker would facilitate the use of high-frequency speech information, where limitations imposed by reduced audibility will impair speech intelligibility.

Considering the expected larger differences in SRT between hearing-impaired and normal-hearing listeners in time-modulated or spectrally filtered maskers
compared to stationary noise (Festen & Plomp, 1990; Peters et al, 1998), SRT measurements in a modified noise could improve discrimination between hearing-impaired and normal performance, providing a more sensitive measure of hearing impairment. However, this only applies when the reliability of the tests using the modified masking noises remain unchanged or is at least equivalent to that of the original version (Smits & Houtgast, 2007).

The aim of this study is to improve an online speech-in-noise screening test for (early) NIHL. In order to do so, different forms of masking noise modification are investigated, by comparing the speech recognition performance of normal-hearing listeners and hearing-impaired participants with noise-induced hearing loss in these different noise conditions. In addition, the alternative test needs to be reliable and valid, so test-retest results are evaluated and performance on the different speech-in-noise tests is compared to performance on the Dutch sentence SRT test (Plomp & Mimpen, 1979a) and to pure-tone thresholds.

Methods

Participants
The same groups of normal-hearing and hearing-impaired listeners as described in Chapter 4 participated in the current study. Participants were tested at three different audiology departments; LUMC Leiden, UMCN St Radboud Nijmegen and AMC Amsterdam. There were no differences between the subjects tested at the different centres. All subjects were native speakers of the Dutch language.

The normal-hearing (NH) group consisted of 49 listeners (mean age 27.0 years, SD = 8.5 years; 16 male, 33 female), with pure-tone thresholds of 15 dB HL or better across octave frequencies from 0.125 to 8 kHz, including 3 and 6 kHz. The 49 hearing-impaired subjects (mean age 56.3 years, SD = 9.4 years; 47 male, 2 female) were patients of one of the three ENT departments who had recently received audiological evaluations. The inclusion criterion was a combination of one or more pure-tone thresholds greater than 25 dB HL at 2 to 6 kHz and thresholds of 20 dB HL or better at 0.125 to 1 kHz. Also the included subjects had a history of noise exposure, although it is impossible to prove a direct relationship between this exposure to noise and the hearing loss measured. Even though the exact cause of notch-shaped hearing loss remains unknown, the included audiogram configurations are characteristic for NIHL and the results are assumed to be applicable to a NIHL population. In addition, results may be generalized to individuals with high-frequency sensorineural hearing loss due to another cause. Patients with an air-bone gap greater than 15 dB in the tested ear were excluded.
The hearing-impaired participants were divided into subgroups having either a narrow audiometric dip (HI-ND, n = 25, mean age 54.1 years, SD = 7.6 years), corresponding with early NIHL, or a broad dip (HI-BD, n = 24, mean age 58.5 years, SD = 10.2 years), corresponding with more severe hearing loss. Distinction was made based on whether or not their hearing threshold at 2 kHz was affected; when hearing threshold at 2 kHz exceeded the pure-tone average of 0.5 and 1 kHz by more than 15 dB, the patient was classified as having a broad dip. For each group, mean audiometric hearing thresholds of the ears selected for monaural testing are displayed in Figure 5.1.

A power analysis showed that the sample size of 49 in each group will have 84% power to detect a difference in means of 2.0 dB, using a two group t-test with a 0.05 two-sided significance level and assuming a standard deviation of 3.1 dB (Jongmans et al, 2008). Details on power calculation and demographics are reported in Chapter 4. All participants signed informed consent forms before starting the experiment. This study was approved by the ethics committee of the University of Amsterdam.

**Figure 5.1.** Audiometric thresholds of the ear selected for monaural testing, averaged for each of the three subject groups. Error bars represent one SD.
Speech-in-noise tests

Dutch sentence SRT test
Plomp and Mimpen (1979a) developed a speech-in-noise test that consists of 10 lists of 13 short everyday sentences, spoken by a female speaker and presented in a stationary interfering noise with the long term average speech spectrum. This test is considered as the clinical standard in the Netherlands, and the SRT obtained with this test was used as a reference value in this study to which performance on the alternatives of Earcheck is compared. Noise level was fixed and the SRT was measured adaptively according to the standard procedure of Plomp and Mimpen (1979a).

Earcheck
Earcheck (EC) is an online speech-in-noise screening test based on the intelligibility of nine different Dutch CVC words in stationary masking noise. These words were randomly presented three times each. On screen, nine response buttons containing a written representation of the words and a corresponding picture were shown. A tenth button saying ‘not recognized’, was added to prevent respondents from guessing, when the presented stimulus was not understood. Participants were instructed to listen carefully and enter their response using the buttons on the computer screen. The test was performed according to a simple up-down adaptive SRT-procedure with a 2 dB step size and fixed noise level. After an incorrect response, the signal-to-noise ratio (SNR) of the next presentation is increased by 2 dB and after a correct response SNR is decreased by 2 dB. The SRT was calculated as the average SNR over stimuli 7 to 27, and was defined as ‘good’ (SRT ≤ -10 dB), ‘moderate’ (-10 < SRT ≤ -7 dB), ‘insufficient’ (-7 < SRT ≤ -4 dB) and ‘poor’ hearing (SRT > -4 dB) (Albrecht et al, 2005).

All test results will be described by the term ‘speech reception threshold’ (SRT). For the purpose of this study, SRT is defined as the signal-to-noise ratio (in dB) that yields 50% intelligibility, rather than as absolute threshold level.

Stimuli
Speech
The speech material used was the closed set of nine different monosyllables comprising the speech stimuli of Earcheck. These CVC words were chosen from the Dutch wordlist used for diagnostic speech audiometry (Bosman, 1989), with a phonemic distribution representative for the Dutch language (Albrecht et al, 2005). Consequently, the nine words all contained unique vowels (thumb /døeym/, goat /γeit/, chicken /klp/, rat /rat/, fire /vyr/, lion /lew/, cat /pus/, saw /zaχ/ wheel /w¡l/).
Homogenizing the speech material

When using an adaptive procedure to assess speech intelligibility, it is important that the speech stimuli are of equal difficulty when heard in noise, to yield consistent and accurate results. One way to achieve this was to adjust the words in level with respect to an optimized perceptual homogeneity. These level corrections were derived from word-specific intelligibility functions, determined based on online test results of previously performed tests.

Earcheck was implemented online in April 2004. Test results were centrally stored, and data collection until December 2007 was available to determine word-specific intelligibility functions. Tests that resulted in a within-subject standard deviation of more than 2.5 dB were considered unreliable and were excluded (Martens et al, 2005). This resulted in a dataset of approximately 100,000 test results that were available for these analyses. Since the SRT measured with Earcheck is calculated by averaging the SNR of presentation 8 to 27, only these presentations were selected. Each word was presented at various signal-to-noise ratios during the adaptive procedure. In order to compensate for inter-individual differences in overall performance, relative SNRs were constructed by correcting the presentation level for the individual SRT. Since it was known whether the response at that SNR was correct or incorrect, the proportion correct could be calculated for each word at each relative SNR. Based on a fit of these proportions correct word-specific psychometric functions were estimated, using the following logistic regression function:

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

(Equation 5.1)

Where \( P \) is the proportion correct at a given relative signal-to-noise ratio, \( \gamma \) is guess level, and \( s \) represents the slope of the psychometric function at SRT. When using a closed-set of speech stimuli the guess rate is related to the number of alternatives (1/n), thus in this case \( \gamma \) is 0.11.

The relative SNRs at the 50% points for each intelligibility function resulting from this fitting procedure were used to adjust the RMS level of the particular word in order to achieve equal intelligibility. These level corrections were applied to the individual CVC words, meaning that the resulting dB-level of each word differs. To define the SNR in the measurements, the average speech level, i.e. the average level of all word-specific dB-levels, was used.

Masking noise

First a broadband stationary masking noise was constructed with a spectral shape similar to the long-term average spectrum of the homogenized word material. Then an experimental set of interfering noises was created by modulating and/or filtering this speech-shaped noise.
**Speech-shaped noise**

A stationary speech-shaped noise was generated by filtering a white noise, using a FIR filter. This filter was based on the long-term average speech spectrum of the concatenation of all test words, according to the methods described by Versfeld et al. (2000). The filtered noise was scaled to match the level of the speech material. This provides a reference condition against which SRTs in other types of noise can be compared.

**Alternative noises**

An experimental set of alternative Earcheck versions was created. In these tests, the homogenized words were presented in different background conditions, all of which were derived from the speech-shaped noise matching the long-term spectrum of the speech stimuli. Six different masking noises were created, either by spectrally filtering or temporally modulating this speech-shaped noise, or by a combination of this filtering and modulating. The appropriate parameters for the spectrally filtered noise conditions were determined using the speech intelligibility index (SII) according to ANSI S3.5 (1997). The SII model can predict the audibility of speech by calculating the proportion of total speech information that is available to the listener, as function of the SNR of the presentation and listeners' hearing threshold level.

The SII was calculated for several audiograms ranging from normal-hearing to severe noise-induced hearing loss and for various versions of Earcheck. These model-based predictions provided insight into the effects of different kinds of filtered masking noise on the SNR required for correct speech reception. Relevant parameters of filtered noise conditions, such as cut-off frequency, noise floor and filter shape (HP/LP/notch), are varied in order to predict their effects on the SRT. The noises that generated the largest differences between normal and impaired hearing ability and that resulted in a SRT that can be reliably measured at a remote test site, were chosen for the experiment.

We realize that this analysis was partly a first order approximation. The SII-model is validated for speech in stationary noise. Although an extended version is available for SII predictions in fluctuating noise (Rhebergen et al, 2006), this model cannot predict hearing-impaired speech reception and is not used in this study. However, modulation frequencies between 10-20 Hz are known to generate the lowest SRTs when using monosyllabic speech material, and several studies report 16 Hz as an optimum modulation rate (Festen & Plomp, 1990; Smits & Houtgast, 2007).

The following masking noises have been selected for the experiments:
1. Earcheck: a broadband stationary speech-shaped noise, as described above.
2. 16-Hz: a broadband interrupted noise, with a modulation depth of 15 dB.
3. LP: a low-pass filtered stationary noise, with a -15 dB noise floor
4. **LPmod**: a low-pass filtered stationary noise, combined with high-pass filtered 16-Hz modulated noise, with a modulation depth of 15 dB.
5. **HP**: a high-pass filtered stationary noise, with a -15 dB noise floor
6. **HPmod**: a high-pass filtered stationary noise, combined with low-pass filtered 16-Hz modulated noise, with a modulation depth of 15 dB.
7. **NF**: a broadband stationary noise, consisting of only the noise floor of -15 dB.

The characteristics of the filtered noise are specified in Table 5.1 and schematically illustrated in Figure 5.2. The spectrally filtered noises are digitally filtered with either a low-pass or a high-pass filter, employing a cut-off frequency of 1.4 kHz and a steep roll-off slope of more than 100 dB per octave. For all interrupted noises, the speech-shaped noise was modulated by a 16-Hz square wave, with 50% duty cycle. The final condition was a low-level broadband noise, referred to as the ‘noise floor’, created by attenuating the speech-shaped noise by 15 dB. In each alternative noise condition this noise was additionally present, to ensure that the noise floor was sufficiently high to mask potential ambient noise levels. In addition, the noise floor produced more or less equivalent masked thresholds for all subjects, minimizing differences in speech audibility among subjects.

<table>
<thead>
<tr>
<th>Name</th>
<th>SNR start</th>
<th>Filtering</th>
<th>Modulation</th>
<th>Noise floor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earcheck</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16 Hz</td>
<td>-10</td>
<td>-</td>
<td>16 Hz squarewave</td>
<td>-15 dB</td>
</tr>
<tr>
<td>LP</td>
<td>-10</td>
<td>LP (1.4 kHz)</td>
<td>-</td>
<td>-15 dB</td>
</tr>
<tr>
<td>LPmod</td>
<td>-10</td>
<td>LP (1.4 kHz)</td>
<td>16 Hz squarewave</td>
<td>-15 dB</td>
</tr>
<tr>
<td>HP</td>
<td>-10</td>
<td>HP (1.4 kHz)</td>
<td>-</td>
<td>-15 dB</td>
</tr>
<tr>
<td>HPmod</td>
<td>-10</td>
<td>HP (1.4 kHz)</td>
<td>16 Hz squarewave</td>
<td>-15 dB</td>
</tr>
<tr>
<td>NF</td>
<td>-15</td>
<td>-</td>
<td>-</td>
<td>-15 dB</td>
</tr>
</tbody>
</table>

The noises were generated such that the spectral part of the filtered noise that was included or the temporal part that was “on” was identical to the steady-state noise. Accordingly, the overall level of the modified noises was slightly reduced. No adjustments in level were made to compensate for this difference. This way, the benefits of removing parts of the background spectrum can be examined without the confounding effect of increases in the level of the remaining part of the spectrum.
**Figure 5.2.** Schematic presentation of the creation of the different masking noise conditions. A: Spectral representation of the broadband masking noise indicated as the stationary EC noise (dark grey in upper section) and modulated 16-Hz noise (light grey shadowed in lower section). B: Schematic presentation of the filters; LP shows the low-pass filter, HP shows the high-pass filter. C: Representation of the filtered noise spectra of the stationary and modulated noise. D: Schematic representation of the modified masking noises. The upper section shows the stationary LP and HP conditions after combining the filtered results of C with the noisefloor of -15 dB. The lower section shows the LPmod and the HPmod conditions after combining the filtered results of the stationary filtered noise with a complementary modulated filtered noise as represented in C.
Constructing the tests
All processing was done using a 16-kHz sampling rate and the processed signals were converted to a 44.1 kHz rate. Speech and noise files were stored in MP3 format and a Macromedia Flash player (Macromedia Inc., San Francisco, USA) web application was used to mathematically mix the SNRs of the speech and noise files, according to an adaptive procedure with fixed noise level and variable speech level.

The constructed noises all had durations of 10 seconds, and they were recorded preceding each test for calibration purposes. Of each noise, a fragment of 2 seconds was randomly chosen to be used as test stimulus. Rise and fall times of 0.5 seconds were applied.

The SNRs of these modified tests ranged from -30 to -6 dB. In case subjects gave a correct response on -30 dB or an incorrect response on –6 dB the next stimulus was presented at the same SNR, due to ceiling effects. The starting level was fixed at 0 dB for the original Earcheck, and at lower SNRs for the modified masking noises (Table 5.1).

Procedure and set-up
Subjects were tested individually in a sound-insulated booth. At the beginning of the experiment, a pure-tone audiogram was recorded at the octave frequencies of 0.125-8 kHz and additionally at 3 and 6 kHz, using a Decos (AMC, LUMC) or Interacoustics (UMCN) clinical audiometer and TDH-39 headphones. In addition, bone conduction was measured at 0.25, 0.5, 1, 2 and 4 kHz. All consecutive speech-in-noise testing is done in case of monaural signal presentation. For the normal hearing listeners the tested ear was either the subject’s best ear, or, in case of symmetric hearing, the right ear. For the NIHL subjects, the ear showing the most pronounced audiometric dip was selected, but in all cases it was checked that the asymmetry did not lead to cross hearing to the contralateral ear. Following audiometric threshold testing, participants performed the different speech-in-noise tests. Signals were played out 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) via TDH-39 headphones. In the UMCN, signals were fed through the AC-40 audiometer.

First, the Dutch sentence SRT test was assessed in quiet (SRT\textsubscript{q}), using list 1 and 2 as developed by Plomp and Mimpen (1979a). These measurements were used to set the masking noise level of all consecutive speech-in-noise tests. This noise level was fixed at 65 dBA, or at SRT\textsubscript{q} + 20 dBA to ensure audibility in cases of highly elevated SRT in quiet where 65 dBA is not high enough above threshold. Next, two sentence lists in stationary noise were performed. The order of the lists in noise was counterbalanced. The participants received the sentences monaurally to the test ear and were instructed to repeat them as accurately as possible. A sentence was scored correct if all words in that sentence were repeated correctly.
After finishing the sentence tests, the subjects performed the various Earcheck tests with different masking noises. Again, signals were presented monaurally, using Sennheiser HDA-200 headphones in an otherwise identical test set-up. However, at the LUMC no internet access was available in the audiologic booth and the online Earcheck measurements had to be carried out in a quiet room. Ambient noise levels were monitored during all test sessions and are considered to have no effect of performing the supra-threshold speech-in-noise tests. See for more details Chapter 4. The participant was seated in front of a computer touch screen to enter the responses of the different versions of the Earcheck tests. The experiment was divided into two blocks, a test and a retest. Between these blocks subjects paused for approximately fifteen minutes resulting in an intermediate period of 45 minutes between each test and retest pair.

The sequence of test conditions was counterbalanced according to a Latin square method, to avoid learning effects and confounding effects of measurement condition order. The noise levels of each test were calibrated with a B&K type 2260 sound level meter and a B&K type 4153 artificial ear, with the use of a flat-plate adaptor.

Results

First, the SRT results obtained with Earcheck in the various masking conditions are analysed with respect to differences in masking noise and hearing ability. Second, the different Earcheck test and retest results and intelligibility functions are analysed to assess test-retest reliability. Third, correlations between word recognition in different masking noises and both performance on the Dutch sentence SRT test and pure-tone thresholds are analysed, to assess test validity. Finally, for the most discriminating test the sensitivity and specificity are calculated, as this will be the most appropriate candidate for a future NIHL screening test.

Effect of masker types on test results

The effects of the various masking noises employed in Earcheck are examined by analysing the performance of both normal-hearing subjects and participants with different degrees of NIHL. The average SRT results of these groups are displayed in Figure 5.3, for each masking noise condition. Only the speech reception thresholds of the first test are considered for this evaluation, because this is the most representative for people who will do the test only once, as this will be the case in normal practice.

The highest SRTs in each group are generated by the unmodulated Earcheck and lowest SRT values are found when only the noise floor is present. All other modified noise conditions yield SRT values that lie between these extremes. A repeated
Speech-in-noise screening tests for NIHL; improving sensitivity

Measures analysis of variance shows a significant main effect of both ‘test condition’ (F[6,564] = 799.92, p < 0.001) and ‘subject group’ (F[2,94] = 122.78, p < 0.001). Also the interaction between ‘test condition’ and ‘subject group’ is significant (F[12,564] = 34.59, p < 0.001), indicating that the differences between the subject groups vary between the different test conditions.

Differences between subject groups
To further investigate these differences between subject groups for each test condition, test results are analysed using one-way ANOVA’s and post-hoc t-tests with Bonferroni correction for multiple comparisons. As is presented in Table 5.2, the main effect of ‘subject group’ is significant in each test condition. Post-hoc t-tests show significant differences between nearly all subject groups for all tests, except for HP results; SRT results of the hearing-impaired subjects with a narrow dip do not differ from SRTs obtained by the two other subject groups. In addition, HPmod makes no significant distinction between the two hearing-impaired subjects groups.

All other tests result in significant differences between the three subject groups. While the results for both stationary noises show small differences across the groups, the SRT measured in the interrupted noise, either broadband or combined with
low-pass filtering, increases when the respondent has more severe NIHL. However, the greatest differences between the subject groups are found when using the low-pass filtered masking noise.

Table 5.2. Results of a one-way ANOVA investigating the main effect of ‘subject group’ performed for each configuration of Earcheck and the mean differences (in dB) between the three subject groups for each test.

<table>
<thead>
<tr>
<th>Test</th>
<th>F-value ANOVA</th>
<th>Δ NH/HI-ND mean</th>
<th>Δ NH/HI-BD mean</th>
<th>Δ HI-ND/HI-BD mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earcheck</td>
<td>32.7</td>
<td>-2.6</td>
<td>-4.2</td>
<td>-1.6*</td>
</tr>
<tr>
<td>16 Hz</td>
<td>85.4</td>
<td>-3.9</td>
<td>-7.2</td>
<td>-3.2</td>
</tr>
<tr>
<td>LP</td>
<td>162.0</td>
<td>-5.9</td>
<td>-11.7</td>
<td>-5.8</td>
</tr>
<tr>
<td>LPmod</td>
<td>93.5</td>
<td>-5.1</td>
<td>-8.9</td>
<td>-3.8</td>
</tr>
<tr>
<td>HP</td>
<td>5.6</td>
<td>-0.9#</td>
<td>-1.7*</td>
<td>-0.9#</td>
</tr>
<tr>
<td>HPmod</td>
<td>20.5</td>
<td>-2.3</td>
<td>-3.5</td>
<td>-1.2#</td>
</tr>
<tr>
<td>NF</td>
<td>61.8</td>
<td>-2.6</td>
<td>-5.6</td>
<td>-3.0</td>
</tr>
</tbody>
</table>

# not significant, *significant at <0.05. Other differences were significant at p < 0.001. All p-values are corrected using Bonferroni correction for multiple comparisons.

Reliability of the online screening tests
It is the goal of this study to find an Earcheck condition that discriminates better between NH and HI listeners. A great spread in test results and large differences in SRT values of these groups are important test parameters in providing a sensitive measure of disability. However, this only applies when the reliability of the improved test is comparable to that of the original version. To evaluate the reliability of each test condition, test and retest results are investigated. The mean test and retest results of the three subject groups are displayed in Table 5.3, with relevant test characteristics of the various Earcheck configurations with respect to test-retest reliability also presented. These characteristics are obtained using results of the total population of 98 participants.

In the two LP noise conditions, some of the hearing-impaired participants have several consecutive incorrect responses at -6 dB SNR, resulting in an invalid SRT. Although this results in a test outcome that still clearly deviated from normal-hearing performance indicating poor performance (see Figure 5.3), these data are omitted when determining the test-retest reliability characteristics.
Learning effect

One measure of reliability is the difference between test and retest results, estimating a possible systematic change in performance due to learning or fatigue. Table 5.3 shows that the performance in the second test is better than in the first for all conditions, although the differences are small (< 1 dB). Paired-sample t-tests reveal

---

**Table 5.3.** Mean SRT values and test-retest characteristics for the various configurations of Earcheck.

<table>
<thead>
<tr>
<th>Test</th>
<th>Subject group</th>
<th>Test 1 mean (SD)</th>
<th>Test 2 mean (SD)</th>
<th>Mean test-retest difference$^1$</th>
<th>SEM</th>
<th>ICC test-retest</th>
<th>SD inter</th>
<th>Slope (%/dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earcheck</td>
<td>NH</td>
<td>-9.4 (1.5)</td>
<td>-11.2 (1.3)</td>
<td>1.62 (p&lt;0.001)</td>
<td>1.17</td>
<td>0.84</td>
<td>2.79</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-6.7 (2.9)</td>
<td>-8.6 (2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-5.1 (2.4)</td>
<td>-6.1 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Hz</td>
<td>NH</td>
<td>-22.6 (1.3)</td>
<td>-23.2 (1.3)</td>
<td>0.70 (p=0.001)</td>
<td>1.45</td>
<td>0.86</td>
<td>3.74</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-18.6 (3.2)</td>
<td>-19.6 (4.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-15.4 (2.6)</td>
<td>-16.1 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>NH</td>
<td>-20.8 (1.3)</td>
<td>-21.3 (1.7)</td>
<td>0.71 (p&lt;0.001)</td>
<td>1.25</td>
<td>0.93</td>
<td>5.38</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-14.9 (3.4)</td>
<td>-16.3 (3.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-9.1 (2.6)</td>
<td>-9.2 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPmod</td>
<td>NH</td>
<td>-16.9 (2.1)</td>
<td>-17.8 (1.8)</td>
<td>0.48 (p=0.028)</td>
<td>1.39</td>
<td>0.87</td>
<td>4.31</td>
<td>10.6</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-11.8 (2.4)</td>
<td>-11.8 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-8.0 (2.0)</td>
<td>-8.1 (1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>NH</td>
<td>-19.6 (1.8)</td>
<td>-21.0 (2.2)</td>
<td>0.73 (p=0.009)</td>
<td>1.92</td>
<td>0.41</td>
<td>2.22</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-18.7 (2.4)</td>
<td>-18.9 (2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-17.9 (2.4)</td>
<td>-18.1 (2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPmod</td>
<td>NH</td>
<td>-16.0 (2.0)</td>
<td>-16.5 (2.2)</td>
<td>0.46 (p=0.073)</td>
<td>1.77</td>
<td>0.60</td>
<td>2.79</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-13.6 (3.0)</td>
<td>-13.9 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-12.4 (2.2)</td>
<td>-12.9 (2.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>NH</td>
<td>-25.0 (1.5)</td>
<td>-25.8 (1.5)</td>
<td>0.73 (p&lt;0.001)</td>
<td>1.28</td>
<td>0.83</td>
<td>3.07</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>HI-ND</td>
<td>-22.4 (2.8)</td>
<td>-22.7 (2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HI-BD</td>
<td>-19.4 (2.1)</td>
<td>-20.3 (2.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^1$ Test-retest parameters are calculated excluding the subjects with an invalid SRT for LP (n = 9) or LPmod (n = 13) conditions. Displayed ICCs are Two-way Mixed model, Type Consistency, Single measures ICCs.
that test-retest differences are significant for all tests, except for HPmod (p = 0.073). Although this suggests a learning effect for these tests, the effects are of a magnitude that is smaller than inter-individual differences and thus are not considered relevant for clinical applicability.

**Test-retest reliability**

Reliability can also be defined as the consistency of a test’s results across series of observations. This is expressed as the standard error of measurement (SEM), calculated by dividing the standard deviation of the differences by $\sqrt{2}$ (Weir, 1995). This SEM is smallest for the original Earcheck, 1.17 dB (Table 5.3). The SEM values of the modified test conditions are all larger, although this deviation is only small for the LP and NF conditions (1.25 dB and 1.28 dB respectively). Both tests using high-pass filtered masking noise show the highest SEMs.

Speech intelligibility measurements should differentiate between subjects with different degrees of hearing loss. Therefore results should show small SEM and large inter-individual standard deviations ($SD_{\text{inter}}$) (Wagener & Brand, 2005). Table 5.3 shows that the spread in SRT values over subjects is higher than the SEM for each test condition. $SD_{\text{inter}}$ was higher for the low-pass filtered and interrupted noises than for the stationary noise.

In addition, test reliability can be expressed by the intraclass correlation coefficient (ICC) of the test and retest results (Table 5.3). These ICCs are comparable for both stationary broadband Earchecks (EC: 0.84 and NF: 0.83). The ICC’s for the modified test conditions are larger, except for both high-pass filtered test conditions that yield rather low ICCs. The highest value for ICC was found for LP noise (0.93).

**Intelligibility functions**

Another important test characteristic is the slope of the psychometric function, which is a good indicator of the precision of the test. Test-specific intelligibility functions are determined based on a fit of the proportions correct at different presentation levels, for the pooled data of both test sessions. The SNR of every presentation is corrected by the individual SRT for that test and for each relative SNR the proportion correct is calculated. This data is fitted with the logistic regression formula (Equation 5.1), with guess rate 0.11. These intelligibility functions are shown in Figure 5.4. The results for the two groups of HI subjects were merged, in order to obtain a sufficient number of data points per condition.

The horizontal arrangement of the 50% points of the psychometric functions represents the differences in SRTs obtained in the various testing conditions. The online tests showed relatively similar slopes (Table 5.3). The stationary broadband Earcheck has a slope of 11.5%/dB, and the alternative test conditions yield comparable slopes, ranging from 10.0 to 12.2%/dB; at 8.6%/dB only the slope of HP is slightly
shallower than that of the other tests. For all tests, slopes are slightly steeper for normal-hearing than for hearing-impaired listeners, but these differences were very small.

**Figure 5.4.** Performance-intensity functions of Earcheck employing the different masking noises for both the normal-hearing group (left) and the hearing-impaired listeners (right), showing proportion correct as function of the signal-to-noise ratio of the presentations. The presented SNR is relative to the individual SRTs and then corrected to the average test SRT. To assure a sufficient amount of measurements in the HI group, no distinction between hearing-impaired subjects in ND or BD groups was made, and intelligibility functions are based on the pooled data of all 49 HI participants.

**Relationship between SRT results and the pure-tone audiogram**

Test validity relates to the correlation between the test’s results and other parameters of auditory functioning. Therefore performance on the alternative tests is compared to performance on the Dutch sentence SRT test (Plomp & Mimpen, 1979), the clinical standard in the Netherlands. In addition, the relationships of the different SRT results and pure-tone thresholds are assessed.

In these analyses only results of the first test session are taken into account. This is representative for real-life test performance, since retest results show slightly better SRTs due to training effect. First, mutual Pearson correlation coefficients between the different Earcheck variants are calculated, in order to establish the amount of association between the different test conditions (Table 5.4). The correlations between
online test results and the Dutch sentences SRT test are all statistically significant (p < 0.001) and show different degrees of association (Table 5.4). The SRTs for both HP conditions show only weak correlation with the sentence test outcomes. Both test versions using broadband stationary noise show moderate correlation; for the unmodulated Earcheck r is 0.71. The SRTs obtained in interrupted noise and low-pass filtered background show reasonably high correlations (around 0.80).

### Table 5.4. Bivariate correlation coefficients (Pearson’s r) for SRT in noise of different tests with SRT results and PTA-values. All correlation coefficients are significant at p < 0.001.

<table>
<thead>
<tr>
<th></th>
<th>EC</th>
<th>16 Hz</th>
<th>LP</th>
<th>LPmod</th>
<th>HP</th>
<th>HPmod</th>
<th>NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Hz</td>
<td>0.77</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>0.76</td>
<td>0.88</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPmod</td>
<td>0.76</td>
<td>0.86</td>
<td>0.91</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>0.39</td>
<td>0.42</td>
<td>0.40</td>
<td>0.40</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPmod</td>
<td>0.62</td>
<td>0.63</td>
<td>0.56</td>
<td>0.64</td>
<td>0.56</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>0.77</td>
<td>0.80</td>
<td>0.83</td>
<td>0.81</td>
<td>0.44</td>
<td>0.62</td>
<td>-</td>
</tr>
<tr>
<td>Sentence SRT</td>
<td>0.71</td>
<td>0.80</td>
<td>0.80</td>
<td>0.79</td>
<td>0.40</td>
<td>0.60</td>
<td>0.79</td>
</tr>
<tr>
<td>PTA$_{0.5,1,2,4}$</td>
<td>0.75</td>
<td>0.84</td>
<td>0.91</td>
<td>0.88</td>
<td>0.35</td>
<td>0.57</td>
<td>0.82</td>
</tr>
<tr>
<td>PTA$_{3,4,6}$</td>
<td>0.76</td>
<td>0.86</td>
<td>0.92</td>
<td>0.88</td>
<td>0.34</td>
<td>0.57</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Earcheck results obtained in the different masking noise conditions are compared with pure-tone thresholds. Table 5.4 displays correlation coefficients of SRTs and the pure-tone averages PTA$_{0.5,1,2,4}$ (important for speech intelligibility) and PTA$_{3,4,6}$ (noisesensitive frequencies). All correlation coefficients between SRT results and PTA$_{0.5,1,2,4}$ are statistically significant (p < 0.001), although correlations vary from weak (0.35 for HP) to high (0.91 for LP). The coefficients for PTA$_{3,4,6}$ are about the same as those found for PTA$_{0.5,1,2,4}$. Again, the highest correlation is obtained comparing hearing thresholds with SRT in low-pass filtered noise.

In addition, the correlation of SRT results and pure-tone thresholds at each single frequency is calculated, which is displayed in Figure 5.5. The correlation coefficients of Earcheck outcomes in stationary noise increase as frequency increases, reflecting that the greatest variation in this study population is in the higher frequencies. Except for the curves for SRTs in both high-pass filtered masking noise, which display weak correlation and are rather flat, all modified test conditions show strong correlations.
with audiometric thresholds, especially in the high frequency range. Results of the low-pass filtered condition show the highest correlation coefficients with hearing thresholds of 2 kHz and higher.

**Figure 5.5.** Correlation coefficients of SRTs (Pearson’s r) with pure-tone thresholds of the total study population, displayed as functions of frequency, for each of the test configurations.

**Sensitivity and specificity for NIHL**

It is shown that the correlation coefficients between pure-tone thresholds and SRTs in the two low-pass filtered noise conditions are fairly high, and that also the largest group differences occur for these tests. This suggests that a speech-in-noise test in low-pass noise can be used to make better distinction between NH listeners and respondents with NIHL and thus to screen for noise-induced hearing loss.

When a test is deployed for screening purposes it should have both a high sensitivity and a high specificity. Test sensitivity refers to the percentage of hearing impaired participants classified correctly as having a hearing ability worse than ‘good’; test specificity refers to the number of normal-hearing subjects that is correctly classified as normal-hearing listener. The sensitivity and specificity of a test depend critically on the cut-off criterion chosen to distinguish between normal hearing and noise-induced hearing impairment. Figure 5.6 shows receiver operating characteristics (ROC) curves that are calculated, based on the results of the first test session, to explore this relationship in more detail. A second curve is added to these ROC curves,
representing the relationship between the sensitivity, specificity and the cut-off value for each test, when discriminating between hearing-impaired participants with either a narrow or a broad audiometric dip (Figure 5.6).

Figure 5.6. ROC-curve, showing sensitivity and specificity of Earcheck using LP noise (left) and using LPmod (right), depending on cut-off values discriminating between normal-hearing and hearing-impaired subjects (solid line) and values separating mild hearing loss (HI-ND) from severe hearing loss (HI-BD, dashed line). The symbols correspond to the chosen cut-off values, as represented in Table 5.5.

The ROC curves provide insight into the optimal combination of cut-off value and the resulting sensitivity and specificity. The SRT value that yielded a high sensitivity without a sufficient decrease in test specificity is chosen as the appropriate cut-off value. The chosen values of the two Earchecks with LP noise are shown in Table 5.5, and are compared to the values of the original test (see Chapter 4). After choosing the cut-off values in order to differentiate between normal-hearing and noise-induced hearing impairment, the Earcheck with low-pass filtered interfering noise showed the best combination of a high sensitivity and a high specificity of the two test alternatives. It turns out that the use of low-pass masking noise increased test sensitivity from 51% obtained for the stationary broadband test to 95%, while the specificity increased from 90% to 98%.
Discussion

Prevention of noise-induced hearing loss may be improved by an easily accessible self-test, as early detection of NIHL may lead to increased awareness about hearing loss and initiate proper audiological follow-up for affected individuals. The on-line speech-in-noise test Earcheck is considered to be useful for NIHL screening purposes, but an evaluation study showed that the sensitivity of this test to identify NIHL is too low (see Chapter 4). This study was conducted in order to improve the sensitivity of Earcheck for detecting noise-induced hearing loss, by investigating the effects of spectrally and temporally modified interfering noises on speech intelligibility, relative to the original test.

Results of the original test

The speech material of Earcheck was homogenized in order to increase test accuracy. After this process, the noise was adapted in order to maintain the match to the long-term average spectrum of the speech. Mean SRT for normal-hearing subjects on this test was -9.4 dB, which is somewhat poorer than the SRT found for the original version without homogenization (see Chapter 4). Both hearing-impaired groups showed slightly poorer performance as well. This resulted in greater SRT differences between the three subject groups when performing the homogenized EC. It is important to note that this improvement in discrimination was achieved while the overall test accuracy was unchanged. However, a larger learning effect was observed.

Modified test results

Earcheck using stationary speech-shaped noise generated the highest SRTs of all test conditions, since a spectrally matched broadband noise is the most effective masker.
The modified noises resulted in lower SRT values for all subjects. This improvement in SRT when going from stationary noise to modified noise can be defined as masking release. Interestingly, hearing-impaired participants showed less masking release than normal-hearing listeners for the majority of test conditions, resulting in greater differences relative to normal-hearing performance.

**Interrupted masking noise**

SRT values obtained in the 16-Hz interrupted noise improved due to listening in the dips of the masking noise. Although the greatest effect is seen in the normal-hearing group, some masking release is also present in hearing-impaired listeners. This decreases with increasing degree of hearing loss, resulting in only little masking release in the most severely affected HI listeners. The participants with NIHL showed reduced ability to take advantage of the dips in the masking noise relative to normal-hearing listeners. Consequently, the differences between normal-hearing and hearing-impaired listeners increase for interrupted masking noises, as was also reported in previous studies (Festen & Plomp 1990; Phillips et al, 1994; Stuart & Phillips, 1996; Smits & Houtgast, 2007; Peters et al, 1998; Bacon et al, 1998).

**Low-pass filtered masking noise**

Hearing-impaired listeners show the smallest masking release for the two low-pass filtered masking noises compared to all other noise conditions. Their masking release is considerably less than found for normal-hearing listeners; not only the subjects with broad audiometric dips - and poorer thresholds in the high frequencies - show very little benefit from the low-pass filtering of the masking noise, also participants with a narrow dip perform significantly poorer than normal. The high masking release found for the normal-hearing listeners shows that they are able to use the additional high-frequency speech information presented at a higher SNR due to the low-pass filtering. Hearing-impaired subjects have less advantage of this additional spectral information. Their masking release turns out to be strongly related to the degree of hearing loss, indicating that their lack of benefit probably originates from poorer audibility for speech due to higher hearing threshold levels in the high-frequency region.

The NH subjects and HI-ND listeners show higher mean SRTs in the low-pass filtered and interrupted noise condition (LPmod) than in the steady state low-pass filtered masking noise (LP). As relatively less silence occurs in the LPmod condition than in LP, this is consistent with the shape of the noise level distributions. On the contrary, the hearing-impaired participants with a severe loss do not show any difference between the SRTs obtained in the two low-pass filtered conditions. Apparently their ability to take advantage of the less masked high-frequency speech information is very limited since a small increase the degree of high-frequency masking does not affect their performance.
High-pass filtered masking noise
Finally two high-pass filtered background conditions were tested. We expected little or no effect of these conditions because the low-frequency hearing thresholds responsible for possible intelligibility differences in these noises, are comparable among all participants. These HP conditions were mainly chosen to complete our experimental design, in order to control for potentially unexpected results of the low-pass filtered conditions. However, these test conditions also provide useful additional information.

The most striking result of the HP conditions is that they do not discriminate between normal-hearing and hearing-impaired listeners at all. Unlike the higher SRTs in LP noise, hearing-impaired participants do not perform different from normal in HP noise. This indicates that their audibility for speech in noise at higher SNR in the lower frequency region is unaffected by their hearing loss. The addition of fluctuations to this noise type in HPmod results in significantly higher SRTs than in stationary HP noise for all subject groups. However, subjects with NIHL show slightly smaller masking release for the HPmod condition than normal-hearing listeners. Although their hearing thresholds at the lower frequencies are (nearly) normal, they already seem to show some signs of reduced temporal resolution. Because the high-pass filtered noise conditions did not discriminate well enough between the subject groups they are considered unsuitable to be used in a screening test for NIHL, as was expected.

Test reliability
Although all subjects benefit from both temporal dips in the noise and low-pass spectral filtering of the background, the benefits are consistently less for HI listeners. These alternative masking noises increase the contrasts between normal-hearing and hearing-impaired participants. However, these tests can only be applied as screening tests if the noise modifications do not adversely affect their reliability. The results of this study show that this is not the case.

Test precision is largely determined by the steepness of the slope of the psychometric function. All alternative test conditions have intelligibility functions with comparable slopes, which are similar to that of the original Earcheck (11.5%/dB). Shallower intelligibility functions for speech recognition in interrupted noise, as reported in earlier studies (Festen & Plomp 1990; Phillips et al, 1994; Stuart & Phillips, 1996) are not found in this study.

The standard error of measurement, expressing the consistency of a test, was 1.17 dB for the unmodified Earcheck. All alternative test conditions showed SEM values that are slightly higher but still comparable, except for the HP conditions that yielded much higher SEMs. The best value is found for the low-pass filtered noise condition, 1.25 dB. SEMs are much smaller than the inter-individual differences between the
groups, so the precision of the tests should be high enough to differentiate between individual listeners. Another measure for test consistency, the intraclass correlation coefficient, also showed good reliability for the interrupted and LP noises, ranging from 0.83 to 0.93. Again, the best value is found for the low-pass filtered background. Finally, the test-retest differences are calculated for each condition. These are all comparable and much smaller than the average difference obtained for the original Earcheck. Although most of these test-retest differences are statistically significant, suggesting a small learning effect, the magnitude of this learning effect is smaller than the measurement error and can be considered as not clinically relevant.

Test validity

When a test is used for hearing screening, it is important that high correlations exist between its outcome and the presence of hearing loss. Generally, hearing ability is tested by pure-tone audiometry, and hearing loss for speech is determined by the Dutch sentences SRT test, which is considered the gold standard. SRT results of the unmodulated Earcheck show a moderate correlation with the Dutch sentence SRT test ($r = 0.71$). This level of association can be explained by the fact that speech material is quite different (sentences vs. words) as well as the response format (open set vs. closed set). The correlation coefficients found for most of the Earcheck variants are somewhat higher (0.79 to 0.80), except for the high-pass filtered noise conditions that both showed weak correlations with the Dutch sentence SRT test results (0.40 and 0.60).

Considering hearing tested by pure-tone audiometry, it is known that no perfect correlation exists between speech reception thresholds in stationary noise and pure-tone thresholds, because speech reception in noise requires more than just audibility. Previous studies found correlation coefficients ranging from 0.72 to 0.77 between sentence SRT in noise and PTA for NH listeners and subjects with NIHL (Smoorenburg, 1992; Bosman & Smoorenburg, 1995). Current findings show a comparable correlation coefficient of Earcheck results in stationary noise and PTA (0.75). Word recognition in the modified test conditions showed a stronger association with PTA, with high correlation coefficients ranging from 0.82 for 16Hz to 0.91 for LP results. However, the HP noises again show much lower correlation coefficients with pure-tone thresholds. Figure 5.5 presents the results of an analysis of correlation coefficients of all SRT results compared with single-frequency pure-tone thresholds. The curves for the HP conditions are almost entirely flat with an r-value around 0.5 or less, indicating that SRTs measured in HP noise are not strongly associated with hearing ability. It is clear that the correlation coefficients between SRT and pure-tone thresholds attain a maximum value at the higher frequencies and obviously this variation is closely related to the SRT results for the modified noises. Again, SRT in low-pass filtered background yielded the highest correlation with hearing loss.
In summary, Earcheck employing low-pass filtered masking noise reveals the most pronounced differences between hearing-impaired and normal performance and shows the highest correlations with high-frequency pure-tone thresholds, without significantly reducing test reliability. Therefore this modification can be considered as the best alternative to screen for NIHL. Indeed analysis revealed a high sensitivity of 95%, as only two mildly hearing-impaired listeners had recognition performances within the normal range. The test specificity of 98% was also very high.

The approach used here may be assumed to be applicable to other types of speech-in-noise tests as well. If this hypothesis is valid, manipulation of the masking noise can increase the sensitivity and specificity with respect to NIHL for different types of speech-in-noise screening tests. In addition, the results of this study may be generalized to any individual with a high-frequency sensorineural hearing loss regardless of the etiology, which is important for the broad application of screening tests such as Earcheck in the Netherlands.

Study limitations
The results presented here were obtained in lab situations, under well-controlled test conditions and with standardized parameters. A potential limitation of a broad application of this screening test over the internet might be the lack of control over environmental variables at the remote test site, such as ambient noise levels, and over testing conditions such as PC settings, presentation level and transducers used. These parameters have no or little influence on test outcome when presenting speech in a broadband stationary noise with a spectrum matched to the long-term average speech spectrum (Plomp, 1986; Smits et al, 2004; Culling et al, 2005; Ozimek et al, 2009). However, their influence on test results using these modified masking noises is unknown.

Online the presentation level will be individually adjusted before starting the screening test. In this study noise level was fixed at 65 dBA, representing general conversation level. It is reasonable to expect that this could be below the preferred volume of hearing-impaired participants, and that actual testing will be done at higher presentation levels. There is no level effect when measuring SRTs in stationary noise (Plomp, 1986; Wagener & Brand, 2005), but SRT results in interrupted noise become better with increasing presentation level (Rhebergen et al, 2010b). As audibility plays an important role in the intelligibility differences in the low-pass filtered noise conditions, higher presentation levels could result in better SRTs in this noise condition as well. Nevertheless, this effect will be limited by the noise floor, resulting in a more or less equal audibility for all subjects. Also, the presence of background noise at the remote test site may affect test results. However, the added noise floor may limit this effect as well; it is set at -15 dB to ensure that the noise floor was above uncontrollable ambient noise levels.
When performing the test at home, participants are strongly advised to perform the test in a quiet room with headphones, although the use of speakers is allowed as well. Culling et al. (2005) showed that variations in the type of headphone had negligible effects on SRT in stationary noise. In addition, SRTs found for speech-in-noise tests in a living room environment were similar to those obtained under headphones in laboratory conditions (Ozimek et al, 2009), even when loudspeakers were used (Culling et al, 2005). However, other studies indicate that a different set of reference values is needed when speech-in-noise tests are performed using loudspeakers instead of headphones (Smits et al, 2006a; Jongmans et al, 2008).

So there are some remaining uncertainties that need to be investigated before the test can be implemented online for use at home. However, this study shows that the lab results of Earcheck with low-pass filtered masking noise are promising for the purpose of screening for NIHL.

**Conclusion**

The aim of this study was to examine whether the online speech-in-noise test Earcheck would perform better as a screening test for noise-induced hearing loss when using a modified masking noise, as compared to the original version of the test. Earcheck with a low-pass filtered masking noise showed the best discriminative power between subjects, and was strongly correlated with results of the Dutch sentence SRT test and with pure-tone thresholds, especially in the high-frequencies. This speech-in-noise test can be considered a very useful test for (early) NIHL, due to the small measurement error and the large spread in SRT values in this population of listeners with different degrees of NIHL. The test had a sufficiently high sensitivity of 95% and specificity of 98%, and is thus considered applicable to use as a valid screening test for NIHL. This manipulation of the masking noise may also be applicable to increase the sensitivity and specificity with respect to NIHL or any type of high-frequency hearing loss for other types of speech-in-noise screening tests.