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.
Speech-in-noise screening tests by internet, Part 3: Test sensitivity for uncontrolled parameters in domestic usage

M.C.J. Leensen
W.A. Dreschler

Revised manuscript submitted to International Journal of Audiology
Abstract

Objective: The online speech-in-noise test ‘Earcheck’ is sensitive for noise-induced hearing loss (NIHL). This study investigates effects of uncontrollable parameters in domestic self-screening, such as presentation level and transducer type, on speech reception thresholds (SRTs) obtained with Earcheck.

Design: Subjects performed 26 Earchecks that differed regarding presentation level (65, 71, and 77 dBA), presentation mode (monotic or diotic), and masking noise (two different low-pass filtered noises) in the lab. To investigate effects of test environment, participants conducted 8 additional Earchecks at home using different transducer types (headphones or loudspeakers).

Study sample: Thirty noise-exposed workers, either normal-hearing (n=10), or with different degrees of NIHL (n=20), participated.

Results: There was a minor effect of presentation levels exceeding 65 dBA in severely impaired listeners. Diotic presentation mode yielded lower SRTs compared to monotic presentation mode. Normal-hearing test results at home were poorer than in the laboratory, whereas hearing-impaired subjects performed better in domestic testing. Using loudspeakers deteriorated SRTs significantly in comparison to headphones, but only in hearing-impaired subjects.

Conclusions: A monotic presentation mode using headphones is recommended for domestic screening. Since domestic testing affect SRT results, a follow up study using a large study population should assess Earcheck’s validity when performed at home.
Introduction

Noise-induced hearing loss (NIHL) is a permanent loss of hearing caused by sustained exposure to intense noise, present in occupational settings and/or during recreational activities. Because NIHL occurs gradually over several years it comes on insidiously, causing the risk of hearing loss to be easily underestimated (Vogel et al, 2008). Therefore, consciousness about hearing problems may be increased by an accessible and reliable hearing screening test. Internet-based screening tests have attracted widespread interest and their rapid growth and increasing usage illustrate that the internet is a suitable medium to contact the general public (Smits et al, 2006a; Koopman et al, 2008; Swanepoel & Hall, 2010). An easily accessible online self-test offers a new approach to facilitate awareness about the risks of hearing loss due to noise or music exposure. Through early detection of hearing loss, it can help to prevent further development of hearing damage in individuals already affected. Furthermore, internet-based self-testing may offer new methods to monitor hearing health in certain noise-exposed populations, such as in occupational settings.

Since subjects with NIHL often encounter difficulties understanding speech in noisy situations (Chung & Mack, 1979; Smoorenburg et al, 1982; Smoorenburg, 1992; Bosman & Smoorenburg, 1995), a speech-in-noise test can be considered a suitable measure for screening. More importantly, speech presented in stationary noise is a very suitable method for online application, because it has less strict acoustical requirements than traditional pure-tone audiometry due to suprathreshold presentation, requires less strict calibration and enables automated test administration by the use of a simple adaptive procedure. In the past years, several simple and automatic online speech-in-noise screening tests became available, such as digit triplet tests in several countries (Smits et al, 2004; Jansen et al, 2010; Zokoll et al, 2012; Watson et al, 2012), and ‘Earcheck’ in the Netherlands using monosyllables (Albrecht et al, 2005). These tests have proven to be well accepted by the users and offer a reliable self-screening test for hearing loss in general.

Unfortunately, most of these tests lack sensitivity to specifically detect the (mild) high-frequency hearing loss that is typical for beginning NIHL (see Chapter 4), because the words, usually presented in a closed set, can be recognized from low-frequency cues only. However, a recent study showed that when the online speech-in-noise test Earcheck used low-pass filtered masking noise instead of broadband noise, the discriminative power of the test increased substantially (see Chapter 5). Employing a low-pass filtered masking noise facilitates the use of high-frequency speech information, which is advantageous for normal-hearing listeners. Reduced audibility of this high-frequency speech information will limit the potentially positive effect of low-pass filtered noise in subjects with NIHL, since their hearing is affected in this frequency region. Consequently, the test’s sensitivity to detect NIHL improved from
51% to 95%, and this was shown to be possible without a reduction in test reliability (see Chapter 5)\(^3\).

Domestic testing has no or limited influence on test outcome when speech is presented 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, this may not be the case when the masking noise is spectrally filtered. The increased test sensitivity was shown in a well-controlled experiment performed in a lab environment. Self-testing over the internet may be affected by the lack of control over testing conditions, such as the presentation level, the testing environment (such as ambient noise present), and the expected variety of equipment used by the respondents (such as PC settings). In order to investigate the effect of these parameters on Earcheck with either one of the two types of low-pass filtered noise, test results obtained under different test conditions are compared in this study. These test conditions differed with regard to:

- **presentation level**: Presentation level is set individually before starting the test. Level is no critical factor in measuring SRTs in stationary speech-shaped noise since it depends upon the SNR rather than upon the absolute level, as long as speech level clearly exceeds the individual’s threshold (Plomp & Mimpen 1979b; Wagener & Brand 2005; Theunissen et al, 2009) and is within the range of moderate conversation levels where effects of uncomfortable listening levels do not deteriorate speech intelligibility (Studebaker et al, 1999; Dubno et al, 2006; Summers & Cord, 2007). However, in low-pass filtered noises, level-dependency of the SRTs measured may be introduced, as the audibility of unmasked high-frequency speech information might increase at higher levels.

- **presentation mode**: Currently, online speech-in-noise tests for domestic screening presents speech and noise diotically. A monotic presentation mode that allows the testing of both ears separately by headphones may be highly beneficial in cases of mild to moderate asymmetric hearing loss. On the other hand, it is known that speech discrimination under diotic conditions is superior to monotic listening (Plomp & Mimpen 1979a; Kaplan & Pickett 1981; Davis et al, 1990; McArdle et al, 2012). In order to choose the appropriate presentation mode, differences between Earcheck results in monotic and diotic presentation mode will be considered in this study.

\(^3\) Similar effects were found for a stationary low-pass filtered noise combined with a high-pass interrupted noise replacing the removed high-frequency part of the noise.
- test environment: Earlier studies on speech-in-noise testing in a living room environment, using either headphones (Ozimek et al, 2009) or loudspeakers (Culling et al, 2005), yielded results that were highly comparable to those obtained in laboratory conditions. However, these living room environments were simulated and hence were similar to all participants. In addition, the variability in computers to be used for domestic testing, with different sound cards and various possible PC settings, is largely unknown. These specifications of domestic equipment might become relevant for test outcomes when part of the masking noise is removed by low-pass filtering, as is the case in the Earcheck test investigated in this study.

- transducer type: Several experiments revealed that variations in the type of headphone had negligible effects on SRT in stationary noise (Culling et al, 2005). However, when SRTs are obtained in a low-pass filtered noise, effects of (differences in) the frequency response of the transducer used may influence speech recognition, especially when there is a peak in high frequency region with less masking energy. Instead of using headphones, testing can also be administered using loudspeakers. The advantage of loudspeakers over headphones is their availability as they are standard PC equipment. In addition, the use of loudspeakers offers the ability to assess hearing capacity with hearing aids. However, when a speech-in-noise test is presented over loudspeakers, rather than headphones, the acoustics of the test environment, such as room reverberation and ambient noise levels, could degrade speech intelligibility (Culling et al, 2005; Soli & Wong 2008; Theunissen et al, 2009).

This study consists of two experiments investigating the performance of Earcheck in different testing conditions. In experiment A participants perform Earcheck in different testing conditions under well-controlled lab conditions, in order to investigate the effect of presentation level and presentation mode on SRT results. In addition, test-retest reliability is assessed. In experiment B the same participants perform Earcheck at home, to examine the influence of test environment and transducer type on the SRTs obtained by Earcheck.

**Methods**

**Participants**
A selection of male construction employees aged 18 years or older who recently had a periodic occupational health examination was invited to participate in this study. In total 30 participants were included, all employed in different occupations in
construction industry. Twenty-four of them reported job related noise exposure. All subjects were native speakers of the Dutch language. The study population was divided into three subgroups. The first subgroup consisted of 10 listeners considered as normal hearing (NH), with pure-tone thresholds below or equal to 20 dB HL for the octave frequencies between 0.25 and 6 kHz, including 3 kHz. Their ages ranged from 32 to 60 years (mean age 46 yrs, SD = 8.7 years), and mean job tenure in construction was 25.6 years (SD = 9.8 years). The remaining 20 subjects were mildly-to-moderately hearing-impaired (HI) participants. They had high-frequency hearing loss defined as one or more pure-tone thresholds greater than 25 dB HL at 2 to 6 kHz. All had normal hearing (≤ 20 dB HL) at frequencies below 2 kHz and none of them used hearing aids or had known middle ear problems. They were employed in construction for 32.2 years (SD = 6.7 years) on average, and all reported some kind of noise exposure, either occupational and/or recreational, in the past.

Analogous to the earlier studies concerning Earcheck described in Chapters 4 and 5, the hearing-impaired participants were divided into two subgroups having either a narrow audiometric dip (HI-ND, n=14, mean age 50 years, SD = 7.0 years), corresponding with early NIHL, or a broad dip (HI-BD, n=6, mean age 56 years, SD = 5.6 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 of the three groups, mean audiometric hearing thresholds are displayed in Figure 6.1.

Test stimuli
The online speech-in-noise test Earcheck (in Dutch: www.oorcheck.nl), was used to assess speech recognition performance in noise (Albrecht et al, 2005). Earcheck consisted of a closed set of nine different monosyllabic words (thumb /doeym/, goat /yeit/, chicken /klip/, rat /rat/, fire /vyr/, lion /lew/, cat /pus/, saw /zay/ wheel /wil/). These words were randomly presented in background noise. On screen, nine response buttons were shown, containing a written representation of the words and a corresponding picture. A tenth button, saying ‘not recognized’, was added to prevent guessing. The test was automated using an up-down procedure with fixed noise level and speech level varied adaptively: after an incorrect response, or ‘not recognized’, the next stimuli was presented at a 2 dB higher level, after a correct response the stimulus level was lowered by 2 dB. The test consisted of 27 trials. The signal-to-noise ratios (SNR) of the last 20 presentations were averaged to calculate the speech reception threshold (SRT), defined as the SNR at which 50% of the speech stimuli was recognized correctly. The SNRs of Earcheck ranged from -30 to -6 dB, and the starting level was fixed at -10 dB.
The masking noise used in this test was a low-pass filtered noise, either without (LP) or with temporal modulations in the high-frequency part (LPmod). Both masking noises were derived by digitally filtering a stationary broadband noise with a long term average spectrum similar to that of the speech stimuli, using a low-pass filter with a cut-off frequency of 1.4 kHz and with a steep roll-off slope (100 dB/octave). To generate the LP noise, a noise floor was added after filtering, that consisted of the speech-shaped noise attenuated by 15 dB (see Figure 6.2 for schematic representation). In order to create the LPmod noise, the stationary speech-shaped noise was modulated by a 0.016-kHz square wave with 50% duty cycle and modulation depth of 15 dB, to generate fluctuating noise. This fluctuating noise was digitally filtered by a high-pass filter with a cut-off frequency of 1.4 kHz and steep roll-off slopes (100 dB/octave) and was added to the low-pass filtered stationary noise (Figure 6.2). See for more details Chapter 5.

**Figure 6.1.** Mean audiometric thresholds for each group for left and right ear. Error bars represent one SD.

**Procedure and set-up**

This study consisted of two experiments. Experiment A was completely conducted at the laboratory of the department of Clinical and Experimental Audiology at the AMC Amsterdam. In experiment B, participants performed the internet-based Earcheck at home in 8 different conditions. The experimental protocol and all procedures in this study were approved by the ethics committee of the University of Amsterdam (approval number: 2001_187).
Experiment A
During their visit to the AMC Amsterdam, the subjects were tested individually in a double-walled sound-proof booth (size: 3x2x2.2 m). At the start of each test session, pure-tone audiometry was conducted, using a Decos Audionigma PRO audiometer connected to Telephonic TDH-39 headphones. Air conduction thresholds of both ears were obtained at octave frequencies from 0.125 Hz to 8 kHz, including 3 and 6 kHz. The audiometer was calibrated according to ISO-389-1 (1998). Following audiometric testing, participants performed Earcheck in several different testing conditions. Speech and noise signals were played via an Echo soundcard (Gina 24/96) on a PC at a sample frequency of 44.1 kHz. Then they were routed via a TDT Programmable Attenuator (PA4), ensuring separate attenuation as required for a

Figure 6.2. Schematic presentation of the creation of the two different masking noise conditions. A: spectral representation of the broadband masking noises indicated as the stationary noise (dark grey in upper section) and 16 Hz modulated 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 spectra of the stationary and modulated noise. D: schematic representation of the modified masking noises. The upper section shows the stationary LP conditions after combing the filtered results of C with a -15 dB noisefloor. The lower section shows the LPmod condition after combining the filtered results of the stationary filtered noise with a complementary modulated filtered noise as represented in C.
given presentation level, and through a TDT Headphone Buffer (HB7). Subjects received the signals via HDA 200 Sennheiser headphones. The noise levels of the test were calibrated with a B&K type 2260 sound level meter and a B&K type 4153 artificial ear with flat-plate adaptor.

Standard Earcheck administration procedure was utilized in this study (Albrecht et al, 2005). Respondents were instructed to listen carefully and enter their responses through the corresponding button on the computer screen. To allow subjects to become acquainted with test stimuli and procedure, a preliminary sequence of the test words was presented preceding testing. This sequence was presented at an individually chosen comfortable listening level that was set prior to the administration of the actual speech-in-noise tests. Mean individually set preferred listening level was 66.2 dB (range 56.2 – 73.0 dBA, SD = 5.6 dBA).

In order to assess the effect of presentation level, Earcheck was performed at three fixed intensities. Noise levels were set at either normal conversation level or levels that were 6 dB or 12 dB higher; 65, 71 and 77 dBA. Also, testing was done in either monotic (testing each ear consecutively) or diotic signal presentation. The diotic presentation mode was not tested at 71 dB to reduce testing time in order to limit experiment A to one single visit to the AMC. This led to eight different conditions that were all tested in test and retest (referred to as ‘repetition’). Monotic retest measurements were limited to a single ear to reduce testing time. The ear to be retested was randomly chosen and remained the same throughout the experiment in each participant. See Table 6.1 for an overview of the different testing conditions.

The various test conditions were conducted in two blocks. Each block contained the 13 different tests using either one of the two masking noises; in 8 tests and 5 retests (Table 6.1). To avoid possible order effects, the tests in each experimental block were counterbalanced across subjects; the eight tests varying in presentation level and presentation mode were counterbalanced using an 8x8 Latin square design, and the five retests were presented in opposite order. For each subject, masking noise condition for the first block of trials was randomly selected. In total, the laboratory experiments comprised 26 Earcheck tests in different conditions.

Experiment B
In order to evaluate differences in SRT results when Earcheck was performed at home as opposed to in a well-controlled lab situation, all participants repeated a selection of eight Earchecks at home (Table 6.1), using their own personal computer. Participants were instructed to complete the test in a quiet environment, at a volume level that is comfortable. Six of the test conditions (3 tests differing in presentation mode (monotic, testing both ears, and diotic) in 2 masking noises (LP/LPmod)) had to be performed with headphones (Table 6.1). To reduce variability in testing equipment in
this study, all participants received relatively simple headphones (HQ, type HP 113 LW) for domestic testing. Finally, Earcheck with both types of masking noise was done using participants’ own PC loudspeakers.

### Table 6.1. Summary of different conditions tested in experiment A (26 tests) under laboratory conditions and experiment B (8 tests) obtained at home.

<table>
<thead>
<tr>
<th>Masking noise</th>
<th>Presentation mode</th>
<th>Experiment A: lab</th>
<th>Experiment B: home</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Presentation level (dBA)</td>
<td>Presentation level (dBA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>LP</td>
<td>monotic</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>diotic</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>LPmod</td>
<td>monotic</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>diotic</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

Experiment A was performed in two experimental blocks, each containing the test and retest measurements with either one of the two masking noises. When monotic testing was performed twice in a certain condition, both left and right ear are measured.

### Statistical analyses

In total, Earcheck was performed in 34 different conditions by each of the participants. To account for effects of repeated measurements within each individual, linear mixed effect models were used to estimate the difference in SRT over the various testing conditions. This method could also handle the missing variables that were incorporated in the acquired data, due to the experimental design. To perform correct analyses including all variables of interest, the total dataset of SRT results was split up into three subsets, before linear mixed effects models were estimated:

- Dataset A included the 26 test results of experiment A, and was used to investigate the effect of presentation level, presentation mode and repetition.
- Dataset B was used to investigate the differences between test environments and contains all 32 results of tests conducted with headphones in experiments A and B.
- Dataset C was used to study the influence of transducer type and consisted of the results of the four diotic tests either through headphones or through loudspeakers, performed at home in experiment B.

---

Here we deviate from “real-life” domestic testing, but the application of low-cost headphones is a feasible option in a screening program. This approach also solves the problem that headphones are not always available for each PC.
For all analyses of SRTs over different conditions, the variation between ‘subjects’ and ‘ears’ within subject were treated as random effects. In addition, Earcheck was performed in two experimental blocks, each with a different interfering noise. Because of this experimental design, variation between ‘masking noise’ within subjects was also treated as a random effect. For the same reason ‘repetition’ was considered as a random effect, but this was not shown to be a significant random factor. Fixed factors of primary interest were ‘presentation level’, ‘presentation mode’, ‘repetition’ (in dataset A), ‘test environment’ (in dataset B), and ‘transducer type’ (in dataset C). In each model, ‘masking noise’ and ‘subject group’ were included as fixed factors, to account for their known systematic differences in SRT results (see Chapter 5). Two-way interaction terms between each of the fixed variables and ‘masking noise’ and ‘subject group’ were incorporated in the model as well, since the primary interest of this study is whether shown effects of the tested parameters are similar in both masking conditions and in the three subject groups. When there was a reasonable a priori expectation of interaction between two other factors, these were also included. Only the factors and interaction terms that showed a significant contribution to the fitted model, tested with conditional F-tests at the 0.10 level, were investigated for significant coefficients of each level. When coefficients proved to be significant, the term was retained in the model. Results of the models are displayed as the estimated effects for the fixed effect levels and interaction terms retained in the model and their 99% confidence intervals, relative to the reference condition of monotic test results of the normal-hearing group, presented at 65 dBA in LP noise. In case of an interaction between two variables, the difference between a certain condition and this reference is obtained by summing up the coefficients obtained for each separate factor contributing to that interaction.

The repeated measurements of SRT in both test and retest, and over the different testing conditions, were used to assess the test-retest reliability of Earcheck. The intraclass correlation coefficient (ICC) is a relative index of test precision. It represents the ratio of the between-subjects variability to the total variability in the data, and is used to determine the consistency of the position of individual scores relative to others (Weir, 2005). In addition, the standard error of measurement (SEM), a measure of absolute consistency, was calculated. This measure quantifies the precision of individual outcomes on a test, and assesses the reliability within individual subjects (Weir, 2005). It combines an overall standard deviation (SD) of all measurements and the ICC as follows: SEM = SD*√(1-ICC). When there are two levels of trials, which is the case in test and retest, SEM calculation can be simplified by dividing the SD of the differences by √2. The data were analysed using SPSS (version 19.0) and R software (R Foundation 2008, from http://www.R-project.org).
Results

Results of experiment A
The mean SRT results of the three subject groups in the different testing conditions are shown in Figure 6.3. Linear mixed effect models were run as previously described and all fixed variables showed significant effects, explained in more detail below. The coefficients of the full model are presented in Table 6.2.

Effect of masking noise
In this model, ‘masking noise’ showed a significant main effect (F[1,119] = 69.13, p < 0.001), as was expected based on previous results (see Chapter 5). On average, SRTs obtained in low-pass noise were 4.0 dB better (p < 0.001) than SRTs obtained under low pass fluctuating noise. There were no significant interactions found between ‘masking noise’ and the other fixed factors, except for ‘subject group’ (described below), indicating that the effects of the various testing conditions were similar in both interfering noises.

Effect of subject group
The main effect of ‘subject group’ (F[2,27] = 25.97, p < 0.001) showed that SRTs were different between subject groups. However, only SRTs of the hearing-impaired participants with a broad dip were significantly higher (thus poorer) than found for the participants in both other subject groups. HI-BD subjects showed SRT results that were on average 7.4 dB worse than the NH results (p < 0.001). Although SRTs of the hearing-impaired group with a narrow dip were 1.4 dB poorer than those of the normal-hearing group, these differences turned out to be not statistically significant (p = 0.116) (Table 6.2). ‘Subject group’ showed significant interactions with both ‘masking noise’ and ‘presentation level’. The significant interaction between ‘subject group’ and ‘masking noise’ (F[2,27] = 5.29, p = 0.012) indicated that the mean SRT difference of 4.0 dB between both noise conditions differed between the three subject groups. Indeed, the difference between the two noise conditions was significantly smaller (2.3 dB) in the HI-BD group than in the two other subject groups (p = 0.039).

Effect of presentation level
Increasing presentation levels over the range from 65 to 77 dBA showed only minor effects on SRTs measured. The main effect of ‘presentation level’ was not significant (F[2,593] = 0.58, p = 0.558), but there was a significant interaction between ‘subject group’ and ‘presentation level’ (F[4,593] = 2.22, p = 0.065): the model indicated an improvement in SRT at higher presentation levels only in the severe hearing-impaired participants. In the HI-BD group, SRTs were on average 1.0 dB better at both higher
levels compared to testing at 65 dBA (p = 0.024 for 71 dBA and p = 0.026 for 77 dBA). In the normal-hearing and the mildly hearing-impaired listeners groups SRTs at different presentation levels were similar (Table 6.2 and Figure 6.3).

**Effect of presentation mode**

SRTs were obtained in two different presentation modes: monotic and diotic. In monotic presentation mode both left ear and right ear were tested separately. Although we were not primarily interested in inter-ear differences, our analysis showed that average SRTs obtained from testing the left ear were 0.55 dB higher, thus worse, than SRTs resulting from right ear testing. However, since the focus of this analysis was on presentation mode rather than on differences between ears, the full model considered results obtained in either monotic or diotic presentation mode, regardless of the ear tested.

**Figure 6.3.** Mean SRT results for the three subject groups over the 26 different experimental conditions of experiment A. Error bars represent 95% CI. Results in the stationary LP noise are displayed in the left panel, results in the fluctuating LP noise in the right. The upper panel shows test results, retest results are presented in the lower panel.
‘Presentation mode’ was a significant main effect in the model ($F[1,119] = 26.18$, $p < 0.001$). Results from diotic speech-in-noise testing were slightly better than when testing was done monotonically. Averaged over all testing conditions, listeners showed a benefit of 0.85 dB performing the test with both ears over monotic testing. There was no statistical interaction between ‘presentation mode’ and ‘subject group’, demonstrating all groups benefitted equally from diotic presentation.

**Effect of repetition**

Most of the test conditions were performed in test and retest, in order to assess a possible test-retest difference. A change in SRT between similar test conditions over time might indicated a learning effect (if the SRT improved) or signs of fatigue (if the SRT deteriorated). Indeed, the full model showed a main effect of ‘repetition’ ($F[1,119] = 20.71$, $p < 0.0001$); retest results were slightly better than test results, indicating an average learning effect of 0.54 dB. This effect was similar in all subject groups and in the two masking noise conditions.

---

**Table 6.2.** Model coefficients of dataset A: all test results of the 26 laboratory conditions.

<table>
<thead>
<tr>
<th>Model terms</th>
<th>Coefficient</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-19.29</td>
<td>-20.58 – -18.01</td>
</tr>
<tr>
<td>Masking noise: LPmod</td>
<td>4.02</td>
<td>3.03 – 5.05</td>
</tr>
<tr>
<td>Subject group: HI – ND</td>
<td>1.38</td>
<td>-0.36 – 3.13</td>
</tr>
<tr>
<td>Subject group: HI – BD</td>
<td>7.41</td>
<td>5.22 – 9.59</td>
</tr>
<tr>
<td>LPmod * HI – ND</td>
<td>0.71</td>
<td>-0.58 – 2.01</td>
</tr>
<tr>
<td>LPmod * HI – BD</td>
<td>-1.71</td>
<td>-3.33 – -0.09</td>
</tr>
<tr>
<td>Level: 71</td>
<td>-0.07</td>
<td>-0.58 – 0.44</td>
</tr>
<tr>
<td>Level: 77</td>
<td>-0.23</td>
<td>-0.66 – 0.19</td>
</tr>
<tr>
<td>Level 71 * HI – ND</td>
<td>-0.29</td>
<td>-0.95 – 0.37</td>
</tr>
<tr>
<td>Level 71 * HI – BD</td>
<td>-0.96</td>
<td>-1.78 – -0.13</td>
</tr>
<tr>
<td>Level 77 * HI – ND</td>
<td>0.02</td>
<td>-0.54 – 0.59</td>
</tr>
<tr>
<td>Level 77 * HI – BD</td>
<td>-0.80</td>
<td>-1.50 – -0.10</td>
</tr>
<tr>
<td>Mode: diotic</td>
<td>-0.85</td>
<td>-1.19 – -0.52</td>
</tr>
<tr>
<td>Repetition: retest</td>
<td>-0.54</td>
<td>-0.77 – -0.31</td>
</tr>
</tbody>
</table>

All coefficients are expressed in dB. In case of an interaction between two variables, the particular difference between a certain condition and the reference situation is obtained by summing up the different individual coefficients contributing to that condition.
An important feature of a measurement procedure is the test-retest reliability. Reliability could be described as the consistency of a test’s results across series of observations. Intraclass correlation coefficients and standard error of the measurements were calculated for each of the five conditions tested in test and retest and over the total of 13 obtained SRTs in both masking noises, and are displayed in Table 6.3. For the stationary low-pass filtered noise ICC values between 0.78 and 0.86 were found, resulting in an overall ICC value of 0.81. For Earcheck using low pass filtered modulated noise, overall ICC was somewhat lower, 0.68. The coefficients for the five different conditions ranged from 0.65 to 0.81. The obtained SEMs ranged from 1.33 to 1.76 (Table 6.3) and were similar in both masking noise conditions.

Table 6.3. The intraclass correlation coefficient (ICC) and standard error of measurements (SEM) calculated for the different test conditions (based on test and retest results) and over the total set of 13 repeated measurements.

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th></th>
<th>LPmod</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>SEM</td>
<td>ICC</td>
<td>SEM</td>
</tr>
<tr>
<td>Diotic 65 dBA</td>
<td>0.84</td>
<td>1.51</td>
<td>0.76</td>
<td>1.63</td>
</tr>
<tr>
<td>Diotic 77 dBA</td>
<td>0.80</td>
<td>1.70</td>
<td>0.79</td>
<td>1.43</td>
</tr>
<tr>
<td>Monotic 65 dBA</td>
<td>0.83</td>
<td>1.53</td>
<td>0.81</td>
<td>1.33</td>
</tr>
<tr>
<td>Monotic 71 dBA</td>
<td>0.78</td>
<td>1.60</td>
<td>0.73</td>
<td>1.53</td>
</tr>
<tr>
<td>Monotic 77 dBA</td>
<td>0.85</td>
<td>1.39</td>
<td>0.65</td>
<td>1.76</td>
</tr>
<tr>
<td>Total</td>
<td>0.81</td>
<td>1.62</td>
<td>0.68</td>
<td>1.70</td>
</tr>
</tbody>
</table>

*ICC are two-way random model, absolute agreement, single measure.*

Results of experiment B

In experiment B, participants repeated Earcheck in 8 conditions at home, using their own personal computer. Only 26 participants completed these tests; domestic test results were unavailable for three normal-hearing listeners and one hearing-impaired listener with a broad dip. The laboratory results of these participants were excluded before comparing test results acquired in the lab environment to the outcomes obtained at home, using database B. Because the volume setting in domestic testing is unknown the linear mixed model did not account for the factor ‘presentation level’. In addition, the factor ‘repetition’ could not be taken into account, since measurements at home were only performed once. As a result, the linear mixed modeling averaged SRTs obtained in the laboratory over the three presentation levels and over test and
retest. Mean SRT results for the different subject groups and test conditions are presented in Figure 6.4.

All fixed test parameters showed significant main effects. Since the main effects of ‘masking noise’ (F[1,75] = 233.23, p < 0.001) and ‘subject group’ (F[2,23] = 21.91, p < 0.001) concerned the reference condition of lab testing, these effects were similar to those found in experiment A, even though the study population was slightly smaller. There were significant interactions between ‘subject group’ and both ‘masking noise’ and ‘test environment’, and between ‘test environment’ and ‘presentation mode’. The coefficients of the full model are presented in Table 6.4.

**Effect of test environment**

![Graph showing SRT results for different conditions](image)

**Figure 6.4.** Mean SRT results for the three subject groups over the 32 different experimental conditions of experiment B, displayed for lab and domestic testing. The upper panel shows test results obtained in monotonic presentation mode, results obtained in dichotic presentation mode are displayed in the lower panel. Results in the stationary LP noise are displayed in the left panel, results in the fluctuating LP noise in the right. Lab results are averaged over the three presentation levels used. Error bars represent 95% CI.
The significant main effect of ‘test environment’ ($F[1,672] = 12.41, p < 0.001$) indicated that results of at home testing were $1.0 \text{ dB}$ poorer than SRTs obtained in the laboratory environment ($p < 0.001$). However, this only held for the normal-hearing listeners. There was a significant interaction between ‘test site’ and ‘subject group’ ($F[2,672] = 27.12, p < 0.001$) showing that the domestic SRT results for the hearing-impaired subjects were significantly better than the outcomes in lab conditions (Table 6.4). These differences, of $0.78 \text{ dB} (p < 0.001)$ and $1.90 \text{ dB} (p < 0.001)$ for HI-ND and HI-BD respectively, seemed to increase with hearing loss (Figure 6.4). In addition, the interaction term between ‘test environment’ and ‘presentation mode’ was shown to be significant ($F[1,672] = 5.79, p = 0.016$). The positive main effect of diotic listening ($F[1,51] = 18.75, p < 0.001$) in the lab environment of $0.86 \text{ dB} (p = 0.001)$ was even larger when Earcheck was performed at home; the diotic SRT results at home were on average $1.60 \text{ dB}$ better than monotic test outcomes ($p = 0.016$).

**Table 6.4.** Model coefficients of dataset B: all test results of the 32 tests with headphones.

<table>
<thead>
<tr>
<th>Model terms</th>
<th>Coefficient</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masking noise: LPmod</td>
<td>4.68</td>
<td>4.07 – 5.30</td>
</tr>
<tr>
<td>Subject group: HI – ND</td>
<td>1.63</td>
<td>-0.34 – 3.60</td>
</tr>
<tr>
<td>Subject group: HI – BD</td>
<td>7.11</td>
<td>4.77 – 9.44</td>
</tr>
<tr>
<td>LPmod * HI – ND</td>
<td>0.04</td>
<td>-0.72 – 0.80</td>
</tr>
<tr>
<td>LPmod * HI – BD</td>
<td>-2.24</td>
<td>-3.14 – -1.34</td>
</tr>
<tr>
<td>Test environment: at home</td>
<td>1.04</td>
<td>0.46 – 1.62</td>
</tr>
<tr>
<td>At home * HI – ND</td>
<td>-1.82</td>
<td>-2.49 – -1.14</td>
</tr>
<tr>
<td>At home * HI – BD</td>
<td>-2.94</td>
<td>-3.74 – -2.14</td>
</tr>
<tr>
<td>Mode: diotic</td>
<td>-0.86</td>
<td>-1.26 – -0.46</td>
</tr>
<tr>
<td>At home * diotic</td>
<td>-0.74</td>
<td>-1.34 – -0.14</td>
</tr>
</tbody>
</table>

*All coefficients are expressed in dB. In case of an interaction between two variables, the particular difference between a certain condition and the reference situation is obtained by summing up the different individual coefficients contributing to that condition.*
**Effect of transducer type**

The 26 participants that conducted Earcheck at home performed the diotic tests with both headphones provided for this study and loudspeakers of their own personal computer, using each of the two interfering noises (dataset C). Mean SRTs of the three subject groups are displayed in Figure 6.5. Since all tests were performed in diotic presentation mode, ‘tested ear’ was not included as a random factor in the linear mixed model. The coefficients of the full model are presented in Table 6.5.

![Figure 6.5](image)

**Figure 6.5.** Mean SRT results for the three subject groups displayed for the two types of transducer used in experiment C. Results in the stationary LP noise are displayed in the left panel, results in the fluctuating LP noise in the right. Error bars represent 95% CI.

This model also showed significant main effects of ‘masking noise’ (F[1,25] = 158.41, p < 0.001) and ‘subject group’ (F[2,23] = 4.14, p = 0.029). Whereas the effect of masking noise was similar as found in previous analyses, the differences between subject groups were smaller than obtained in the previous models. Because the hearing impaired groups performed better at home than under lab conditions, especially when signals were presented diotically, the differences between subject groups in this dataset were reduced. HI-ND still showed SRTs that did not differ significantly from normal performance (p = 0.099), whereas HI-BD subjects showed SRT results that were on average only 2.8 dB higher than the NH results (p < 0.001). This probably also explained why the interaction term between ‘subject group’ and ‘masking noise’ did not contribute significantly to this model.

‘Transducer type’ did not show a main effect on test outcomes (F[1,49] = 1.37,
p = 0.247), but there was a significant interaction between ‘transducer type’ and ‘subject group’ (F[2,49] = 3.36, p = 0.043). Detailed analyses showed that SRTs with headphones did not differ significantly from SRTs obtained via speakers in the normal-hearing group (p = 0.247). However, there was a significant effect of transducer type in both hearing-impaired groups (see Figure 6.5); when the speech and noise signals were played through speakers HI-ND and HI-BD listeners have SRTs that were on average 2.68 (p = 0.015) and 2.43 dB (p = 0.068) higher, respectively.

**Discussion**

This study aims to investigate the effects of several testing parameters that may vary when the online speech-in-noise test Earcheck is used for domestic screening. Subjects with and without NIHL performed Earcheck in several conditions, both in a well-controlled lab environment and at home using their own personal computer.

**Test-retest reliability**

The lab results of experiment A prove that Earcheck is a reliable test, yielding relatively high ICCs and good within-subject variability. Yet, the ICCs found in this study are slightly lower than the values found in the evaluation study described in Chapter 5. Also, the SEMs are slightly higher than those reported in previous studies, which range from 0.9 - 1.4 dB (Plomp & Mimpen, 1979a; Smits et al, 2004; Vaillancourt et al, 2005; Van Wieringen & Wouters, 2008). The reliability of Earcheck using stationary LP

<table>
<thead>
<tr>
<th>Model terms</th>
<th>Coefficient</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-20.30</td>
<td>-22.05–-18.56</td>
</tr>
<tr>
<td>Masking noise: LPmod</td>
<td>4.63</td>
<td>3.87–5.39</td>
</tr>
<tr>
<td>Subject group: HI – ND</td>
<td>-0.28</td>
<td>-2.46–1.90</td>
</tr>
<tr>
<td>Subject group: HI – BD</td>
<td>2.81</td>
<td>0.23–5.41</td>
</tr>
<tr>
<td>Transducer type: speakers</td>
<td>0.73</td>
<td>-0.52–1.98</td>
</tr>
<tr>
<td>Speakers * HI – ND</td>
<td>1.95</td>
<td>0.40–3.50</td>
</tr>
<tr>
<td>Speakers * HI – BD</td>
<td>1.70</td>
<td>-0.13–3.54</td>
</tr>
</tbody>
</table>

*All coefficients are expressed in dB. In case of an interaction between two variables, the particular difference between a certain condition and the reference situation is obtained by summing up the different individual coefficients contributing to that condition.*
noise is slightly better than that with modulated LP noise, in terms of ICC and SEM. Test-retest reliability for testing at home could not be assessed, but the higher variability in domestic results suggests that this is somewhat poorer than in the lab (see variability in Figure 6.4).

The test-retest measurements in low-pass filtered masking noise show a small but significant learning effect of 0.5 dB that is comparable to the test-retest differences obtained in the evaluation study described in Chapter 5, and in other studies developing a speech-in-noise screening test (Smits et al, 2004; Vailancourt et al, 2005, Wagener & Brand, 2005; Jansen et al, 2010). Although small, this effect should be accounted for when performing consecutive tests for domestic hearing screening.

**Effect of masking noise condition**

For all subject groups, the SRTs in LP noise are somewhat lower (e.g. better) than in LPmod noise, consistent with the shape of the noise level distributions of both noises. But the average difference is significantly higher in the NH group than in HI-BD group (4.0 versus 2.3 dB, respectively). These findings are similar to the differences between LP and LPmod noises shown in the previous evaluation study (see Chapter 5). Apparently, the ability to benefit from the unmasking of the high frequencies is smaller for listeners with severe hearing impairment, since additional fluctuations in the noise have a less disturbing effect in this group. As a result, SRTs obtained in LP noise differentiate better between subjects groups than the SRTs in LPmod. There are no differences regarding the effects of any of the tested measurement parameters between both masking noise conditions.

**Presentation mode**

Earcheck was performed in diotic as well as monotic presentation mode, testing both ears consecutively. When testing monotonically, a small difference of 0.6 dB between left and right ears is observed. This right ear benefit in speech recognition is often seen in speech-in-noise testing, and can be explained by left hemispheric dominance for speech and language processing (Kimura, 1961). Our results agree with differences of 0.6-0.7 dB reported in literature (Plomp & Mimpren, 1979a, McArdle et al, 2012). Also, it is known that speech discrimination under diotic conditions is superior to monotic listening (Kaplan & Pickett, 1981; Davis et al, 1990; Van Hoesel & Litovsky, 2011; McArdle et al, 2012). The diotic benefit in this study is 0.9 dB in the lab and 1.6 dB at home (see Table 6.4). Previous studies showed an improvement in SNR due to diotic listening ranging from 1.3 dB to 1.5 dB (Davis et al, 1990), for either sentences in noise (Plomp & Mimpren, 1979a; Bronkhorst & Plomp, 1998) or digit triplets (Smits et al, 2006a). Hence, the diotic advantage found in the lab experiment is smaller than expected.
Effect of test environment
The most important finding in this study is that the SRT results in domestic testing differ from the SRT results obtained under lab conditions. In agreement with the findings of Culling et al. (2005), a slight overall deterioration in domestic test results could be expected because of poorer testing conditions at home, attributable to possible room reverberation or ambient noise, and quality of used sound cards and headphones that is likely to be lower than in the lab. Accordingly, the SRTs of the normal-hearing group are on average 1 dB worse when the test was performed at home rather than in the lab. Conversely, a positive effect of domestic testing is seen in both hearing-impaired groups. Although this overall effect is small in HI-ND subjects (0.8 dB), HI-BD subjects show an average benefit of almost 2 dB.

In order to gain more insight into the origin of these dissimilarities, we calculated the individual differences between the SRTs obtained in the lab (results were calculated separately for monotic and diotic conditions tested at 65 dB and 77 dB) and at home. These differences were plotted against the pure-tone average of the hearing thresholds at 3, 4 and 6 kHz (\( \text{PTA}_{3,4,6} \)) for the corresponding ear of the individual participants in Figure 6.6. This shows that the difference in SRT, in favor of domestic testing, becomes larger with increasing hearing loss. Apparently, the more severely hearing-impaired listeners show larger benefit from domestic testing, suggesting an increase in audibility of the high-frequency speech information when testing is conducted at home.

Though these effects need further analyses, we will discuss the possible role of differences between lab testing and domestic testing regarding spectral differences between the stimuli, the level of presentation, and the presence of background noise.

Spectral differences
The first possible factor of influence might be differences in the spectra of the test signal, possibly induced by different sound cards, PC settings, or the transfer characteristics of the transducers used. With well-matched speech and noise passing through the same playback device, the variation in devices will leave the signal-to-noise ratio in each frequency band unaffected (Culling et al, 2005), and intelligibility is mainly determined by SNR (Plomp, 1986; Wagener & Brand, 2005). However, part of the masking noise used in the Earcheck test investigated in the present study is removed by low-pass filtering, hence the specifications of domestic equipment may become relevant for the test outcomes and its discriminative power. Specific amplification of the less masked high-frequency region can improve audibility of the speech stimuli and hence test performance.

Headphone presentation, instead of loudspeaker use, offers the advantage of monotic testing, eliminates reverberation, and reduces ambient noise. The use of the
same headphones in this study reduces one of many sources of spectral variability in speech reception tests at home. But the headphones used at home and in the lab differ both in quality and in frequency response. As a consequence, a systematic difference in SRT results cannot be excluded. Although the type of standard headphones was selected for its relatively flat frequency response, a slight difference in frequency response could have been more beneficial for the hearing-impaired groups if it amplifies the high-frequency region. By using the same headphones in the current study, the possible effect of differences in frequency responses is not established, but should be kept in mind when interpreting domestic Earcheck results.

**Level effects**

Another factor may be that the stimuli at home were presented at a higher, self-chosen, presentation level than used in the lab. It is likely that particularly hearing-impaired respondents increase the presentation level to maximize their performance. This is confirmed when comparing the mean individually set presentation levels of the different groups recorded preceding the lab tests: 64.6 dB for NH versus 70.6 dB for

**Figure 6.6.** Individual differences in SRT obtained in the lab and at home plotted against their PTA\textsubscript{3,4,6} for results obtained in LP noise. A positive difference indicates a superior SRT home relative to SRT lab. Left panel shows the differences relative to lab results at 65 dBA, right panel shows differences relative to lab results at 77 dBA. Black symbols represent the results of the NH subjects, grey symbols denote the HI-ND group and white symbols represent the NI-BD group, for results of the left ear (\(\triangleleft\)), the right ear (\(\triangle\)) and the diotic condition (\(\Diamond\)).
HI-BD. This difference in presentation level may be even more pronounced when the test is performed in the presence of ambient noise in a non-isolated environment.

Previous studies reported that speech intelligibility in interrupted noise improves with increasing presentation level (Stuart & Phillips, 1997), due to increased audibility and steeper forward masking slopes at higher noise levels (Rhebergen et al, 2010b). Whereas the effect of the latter factor only applies to intelligibility in LPmod noise, audibility is an important factor in speech reception in both noises. Because of the high masking release in the low-pass filtered noises, relatively low SNRs are reached, and as a result speech at threshold is presented at low levels. Consequently, high-frequency speech information may fall below the elevated hearing thresholds of the HI subjects, who may be expected to experience more benefit as presentation level increases (Summers & Molis, 2004; Rhebergen et al, 2010b). Indeed, the lab results of experiment A show that presentation level has no effect on SRTs in NH and HI-ND subjects, but HI-BD subjects have a 1 dB benefit for presentation levels exceeding 65 dBA. This finding demonstrates that audibility is improved at higher presentation levels for subjects with more severe high-frequency losses.

**Presence of background noise**

In order to assess the effect of transducer type used on the SRT outcome of Earacheck, domestic diotic test results obtained with loudspeakers and headphones were compared. Previous findings of studies investigating domestic screening methods indicated an advantage of using headphones that ranged from 1.1 dB to 1.4 dB (Culling et al, 2005; Smits et al, 2006a; Van Son & Jellema, 2011), due to detrimental effects of poor listening conditions when loudspeakers are used. In this study, the expected poorer outcomes for loudspeaker testing are only shown in both hearing-impaired subject groups. Loudspeaker presentation produces a significant elevation of SRT of 2.6 dB averaged over all hearing-impaired participants. Normal-hearing listeners, however, do not show any significant effects of the transducer type used.

**Differences between subject groups**

Current study aimed to investigate the effects of uncontrollable parameters in domestic speech-in-noise testing. To assess whether effects were different for NH listeners and subjects with NIHL, three subject groups were examined. All the fitted models (presented in Tables 6.2, 6.4 and 6.5) show that SRTs increase with increasing hearing loss, but only the mean SRT of the HI listeners with severe hearing loss is significantly poorer than the mean SRTs of the other groups. Yet, SRTs of the mildly hearing impaired group do not differ significantly from normal-hearing performance, indicating a limited ability to differentiate between NIHL and normal performance. Deviant from the first evaluation study (see Chapter 5), current subject groups were not selected on their hearing, but composed by classifying the partaking construction
workers, with a continuum of hearing threshold levels, after their inclusion in the study. Subjects were considered normal-hearing when all pure-tone thresholds were below or equal to 20 dB HL. As a result, several near normal-hearing subjects, with only slightly elevated audiometric thresholds at 1 or 2 frequencies, are included in the HI-ND group. Although these are the subjects particularly of interest in case of screening purposes, this might explain the insignificant differences found in SRTs between these groups. Our current results show that Earcheck is not sensitive enough to detect these very mildly hearing impairments.

**General applicability**

Overall, LP and LPmod noises used in this study are equally affected by the different test parameters investigated in the experiments, meaning that both masking conditions can be employed in Earcheck for domestic screening purposes. Considering the greater inter-individual spread in SRT results in LP noise, and the slightly better reliability results for this noise condition, the stationary LP noise is considered the best alternative for online NIHL screening.

Although better SRTs are obtained in diotic testing, monotic testing using headphones has the great advantage of testing each ear separately. This produces a more comprehensive assessment of the respondent’s hearing ability, particularly in the case of a mildly or moderately asymmetric hearing losses. Also, headphones are preferred to eliminate the acoustic effects of test environment. If one does use Earcheck for diotic screening in domestic testing, the diotic benefit of 1.6 dB should be taken into account.

Our results do not show any differences between the NH and the HI-ND groups, and significant differences between the NH and HI-BD groups found in the lab experiment decrease when testing was done at home; the normal-hearing listeners show reduced SRTs for domestic testing, whereas the both HI groups improve their performance. Now that domestic testing turns out to affect SRT results, actual results of Earcheck performed at home by a larger study population than currently tested should be compared to pure-tone audiometry, in order to assess Earcheck’s validity and reliability when performed at home.

Nevertheless, an online self-test such as Earcheck provides a fast and easy way to reach many people, which makes the test highly applicable for adult hearing screening. When the discrimination between normal and impaired hearing could be improved by taking into account the variability arising from domestic test administration, Earcheck can be considered as a valuable test for NIHL screening at home.