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.
The applicability of a speech-in-noise screening test in occupational hearing conservation

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

Accepted for publication in the International Journal of Audiology
Abstract

Objective: Noise-induced hearing loss (NIHL) is the most reported occupational health disease in the Netherlands. The internet-based speech-in-noise test Earcheck is designed to detect beginning NIHL and can be a valuable tool in occupational hearing health surveillance. The aim of this study is to investigate the validity of Earcheck compared to regular screening audiometry.

Design: Subjects performed online Earcheck tests at home. The results were compared to a pure-tone screening audiogram obtained during regular occupational health examination. A subgroup performed the measurements twice to assess test-retest reliability.

Study sample: 249 male construction employees who recently had a periodic occupational health examination participated.

Results: An average learning effect of -1.6 dB was found, that reduced with increasing test number. The test-retest variability was 1.6 dB. Sensitivity to detect beginning NIHL was 68%, with a specificity of 71%.

Conclusions: Although sensitivity and specificity values are only moderate, the broad internet application still promises a valuable addition to current practice. The relatively high learning effect indicates that more reliable results can be obtained after a longer test session. When this is put into practice some improvement in sensitivity and specificity may be expected as well.
Introduction

Noise represents one of the most common environmental health hazards. Long term exposure to high daily noise levels may cause noise-induced hearing loss (NIHL), a permanent sensorineural hearing impairment (ISO-1999, 1990). In modern society, large groups of people are frequently exposed to noise, either during recreational activities or in occupational settings. In the past years NIHL was the most reported occupational disease in the Netherlands (Van der Molen, 2010; 2011), especially in sectors with a high concentration of noisy occupations such as the construction industry. In 2012, 12.9% of the workers employed in this sector reported NIHL (Arbouw, 2012). Internationally, a large US analysis showed that the construction industry had the highest number of workers with self-reported hearing impairment attributable to their employment (Tak & Calvert, 2008).

NIHL can mostly be prevented by reducing or eliminating the exposure to noise. Hearing conservation programs in construction mainly focus on employee’s use of hearing protection devices (HPDs) rather than on controlling the noise exposure at its source (Neitzel & Seixas, 2005). Yet few workers use hearing protection consistently enough to prevent hearing damage (Lusk et al, 1998; Hong et al, 2005). NIHL affects the high frequency region (3-6 kHz) and develops gradually (Rösler, 1994). Consequently, it is often unnoticed, especially when hearing loss is still in an early stage (Vogel et al, 2009; Shah et al, 2009). When the damage is substantial and severe enough to be measured, hearing impairment is irreversible. Hence it is of major importance to assess the possibility of hearing damage as early as possible, so precautionary measures can be taken to prevent further development of NIHL (Meyer-Bisch, 1996).

The early detection of hearing loss is thus a crucial aspect of hearing conservation. Periodic testing enables monitoring of the hearing ability of employees at risk, and may help to prevent further development of NIHL. Workers that have been demonstrated to have reduced hearing ability can be motivated better to use HPDs properly (Zohar et al, 1980). The European Directive 2003/10/EC states that employees whose daily noise exposure exceeds the lower exposure action value of 80 dBA have the right to preventive audiometric testing (EPC, 2003). In the Netherlands, pure-tone screening audiometry is incorporated in the periodic occupational health examinations (POHE), to which all construction employees are invited once in at least every 4 years. Participation in this POHE occurs on a voluntary basis.

However, audiometric testing is not always a viable option for construction workers because it is logistically difficult and time-consuming. Despite the attention given to occupational health in the Dutch construction sector, 40-50% of the employees invited for a POHE do not respond to their call. This relatively high non-response reflects the difficulty in attaining the entire population at risk, because
of the widespread and very transient workforce in this sector that is characterized by many small companies, subcontracting, and mobility between workplaces.

Additionally, there are some requirements to assure reliable and valid outcomes in pure-tone audiometry that are not always met in occupational testing, which for practical reasons is often performed at the company or worksite. First, an accurate audiogram requires adequate acoustical isolation, but a sound proof booth is not available in most cases. As a result, possible ambient noise may negatively affect threshold determination. Second, test administration is mostly done during a working day, without a predefined noise-free period prior to testing. Although respondents are advised to wear HPDs prior to examination, hearing threshold assessment might be biased by a temporary threshold shift due to recent noise exposure. Third, this method requires a well-calibrated audiometer and qualified test administrators, making it a costly and time-consuming test method. Therefore alternatives for hearing screening in occupational health need to be found.

Since the effects of NIHL are typically experienced in challenging listening situations such as in background noise, speech-in-noise testing may be a valuable alternative for hearing screening in occupational health. This has the advantage that all stimuli are presented suprathreshold, which is less demanding with regards to the test environment. This makes a quiet room adequate for accurate testing, and there is no need for an isolated testing booth. Speech-in-noise testing can be automated using an adaptive procedure, so it can be used as a self-administered test (Smits et al, 2004; Jansen et al, 2010). Furthermore, calibration can be controlled programmatically and it is relatively insensitive to absolute presentation level or small variations in equipment (Plomp, 1986; Smoorenburg, 1992; Smits et al, 2004; 2006a).

Due to these characteristics, a speech-in-noise test is very suitable for internet-based application. Using the internet to distribute a screening test entails the possibility of evaluating hearing status remotely at home and provides a fast way to reach many employees at risk (Stenfelt et al, 2011). As a result, online hearing screening may lead to higher participation rates in the transient workforce of the construction industry. Moreover, it requires no specialized equipment, can be performed more easily after a period free of occupational noise (e.g. during weekends), and hearing status can be tested more frequently than once every four years, e.g. when complaints arise. So, an easily accessible speech-in-noise test can be a valuable addition to current practice of occupational hearing screening in the construction industry.

In the Netherlands, such a test was developed by the National Hearing Foundation in association with the Leids University Medical Centre; Earcheck (Albrecht et al, 2005). Earcheck is an online speech-in-noise test that measures the ability to understand words in noise by determining the signal-to-noise ratio (SNR) that corresponds to 50% intelligibility. The original version of this test used a stationary broadband noise and
was not sensitive enough to detect mild-to-moderate NIHL (see Chapter 4). By modifying the broadband masking noise this sensitivity increased strongly; by using a low-pass filtered noise that forces the listener to use high-frequency speech information, the test discriminates well between normal and impaired performance due to NIHL. Lab results show that the test has a high sensitivity and specificity of 95% and 98% respectively (see Chapter 5). However, these promising findings were based on results obtained under well-controlled conditions in the laboratory. The question is how this holds when the online test is broadly applied for remote testing in occupational health, when there is lack of control over environmental variables and testing conditions such as settings, presentation level or transducers used.

A recent evaluation study showed that the influence of these parameters on outcomes of domestic Earcheck testing were small, but significant (Chapter 6). Normal-hearing performance at home was on average 1 dB poorer than in the laboratory, whereas hearing-impaired subjects showed a significant beneficial effect of 1.5 dB from domestic testing. These findings indicate that Earcheck's laboratory-based cut-off points need to be refined for adequate discrimination in domestic screening application. The evaluation study also showed that Earcheck results were significantly affected by the chosen presentation mode and transducer type. Although diotic presentation yielded 1.5 dB better results than monotic listening, monotic presentation is recommended since it has the great advantage of testing each ear separately (see Chapter 6). Intrinsic to monotic presentation is that testing should be done via headphones. This choice is also favorable because test outcomes are more reliable when measured with headphones than using loudspeakers (Smits et al, 2006a; Van Son & Jellema, 2011).

The results of Chapter 5 and 6 indicated that two types of low-pass filtered masking noise showed good applicability; a stationary low-pass filtered noise and a stationary low-pass filtered noise combined with a high-pass interrupted noise that replaces the removed high-frequency part of the noise. These additional fluctuations place demands on temporal resolution of the listener, and might increase the discriminative power between normal and impaired speech recognition performance even more. The current study compares both masking noises when applying the online Earcheck in a noise-exposed population, to come to the best alternative for domestic hearing screening.

The aim of this study is to investigate the value of Earcheck in identifying (early) noise-induced hearing loss in addition to regular pure-tone screening audiometry. To do so, Earcheck results of a large population of construction workers are compared to their screening audiograms. The obtained test reliability, sensitivity, and specificity in this occupationally exposed population will determine the value of Earcheck as a screening tool for occupational NIHL.
Methods

This investigation consisted of a cross-sectional comparison of data from construction workers. Survey data collection was completed through internet-based speech-in-noise testing. Pure-tone audiometric data were collected as part of regular periodic occupational health examinations and were provided by occupational health services. The experimental protocol and all procedures in this study were approved by the ethics committee of the University of Amsterdam (approval number: 2001_187).

Participants

The participants in this study were employees of the construction industry, who were possibly exposed to occupational noise and therefore were at risk of developing noise-induced hearing loss. The eligible study population included only male employees aged 18 years or older, including office workers. Female workers comprised only a small percentage of the total workforce in construction (3.5%), and they were discarded because of their relatively high concentration in non-noise-exposed jobs. Since outcomes of the online Earcheck were to be compared to the regular occupational screening audiogram as performed during a periodic occupational health examination (POHE), only subjects who recently had such an examination were selected for participation.

The selection procedure was carried out by Arbouw, the Dutch national institute on occupational health and safety in the construction industry. This organization collected data from medical records of POHEs of all employees in the construction sector throughout The Netherlands.

First, all medical records from POHEs conducted within 3 months prior to the selection date were selected. Records with invalid or missing audiometric data were excluded from the selection, as were subjects reporting known hearing problems due to other etiologies than noise exposure. Then, three random samples of approximately 1,000 subjects were drawn from all selected records, according to a weighing procedure based on age. Since the mean age in the total population of construction employees is 40 years, the selected population was stratified into three age groups, 18-34 yr, 35-44 yr and 45-64 yr, to get a proportional representation of age groups in the study population. The 2,937 selected employees were invited to participate by sending them an invitation letter.

Pure-tone audiometry

Screening audiograms were assessed during a periodic occupational health examination. POHEs are provided for all employees in the construction industry, irrespective of occupational noise exposure. Pure-tone audiometry was conducted at the workplaces, if possible in a mobile unit equipped with a soundproof booth. Manual audiometers
coupled with TDH-39 headphones were used. These audiometers were annually calibrated according to the ISO-389.1 (1998) standard. Pure-tone air-conduction thresholds were determined for both ears at frequencies 0.5, 1, 2, 3, 4, 6 and 8 kHz, in 5-dB steps. Testing was done during the work shift, but subjects were advised to wear HPDs on the day of testing if they had to work in a noisy environment.

**Earcheck**

In the Netherlands Earcheck was developed as an automatic online speech-in-noise test (Albrecht et al, 2005) that was adapted to be used for NIHL screening (see Chapter 5). In brief, it consists of nine different monosyllables, randomly presented in a low-pass filtered interfering noise. On screen, nine response buttons are shown. A tenth button, saying 'not recognized', is added to prevent respondents from guessing.

After the presentation of a word, the subject’s task was to identify the word he had been presented by clicking on the corresponding button on the computer screen. The level of the noise was fixed and the level of presented words varied according to an up-down procedure, with a step size of 2 dB. This was based on the testing procedure according Plomp & Mimpen (1979a), except for the first stimulus of Earcheck that was presented only once at a fixed SNR of -10 dB. A list of 27 presented words was used to estimate the signal-to-noise ratio at which 50% of the speech material was reproduced correctly. This was defined as the speech reception threshold (SRT), and was calculated by taking the arithmetic average of the SNRs of the last 20 presentations.

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 have been 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/oct). To generate the LP noise, a noise floor of -15 dB was added after filtering. In order to create the LPmod noise, the low-pass filtered noise was combined with a high-pass filtered noise that was modulated by a 16-Hz square wave with 50% duty cycle and modulation depth of 15 dB. Further details of the development of noise stimuli can be found in Chapter 5.

**Procedure**

Subjects were recruited by sending them an information letter about the study and inviting them to participate. Subjects willing to participate were asked to log in to a secured website of Arbouw, to sign an informed consent and give permission to retrieve audiometric data from their medical record for the purpose of this study. After this, they were automatically led through a short online questionnaire that serves as a first screening for including adequate participants. Respondents were asked to complete questions regarding noise exposure at work and during leisure.
time, hearing ability and native language. Reported job title was used to estimate daily noise exposure levels of individual construction workers; time weighted average (TWA) noise exposure levels for standardized job titles were extracted from a database of Arbouw (Arbouw, 1998; for more details see Chapter 2). Registered subjects were assigned a random identification number and were sent an instruction letter and e-mail. The letter was accompanied by a standard low-cost pair of headphones (HQ, type HP 113 LW) to assure that every participant had one in order to conduct the domestic screening test properly. Subjects were asked to perform Earcheck at home, using their own personal computer with corresponding settings. They used their identification number to log in to a special experimental website to perform the online speech-in-noise tests.

Standard Earcheck administration procedure was utilized in this study. At the start of the test session, a word without noise was presented repeatedly, and participants used their PC volume control or a slider on screen to adjust the volume to a level at which the presented word was clearly understandable. This presentation level was used for the presentation of all consecutive test stimuli. To allow subjects to become acquainted with the stimuli, a preliminary sequence of the nine different test words was presented in noise preceding the actual tests.

All testing was done monaurally. Both ears were tested consecutively, and the ear to be tested first was chosen randomly. In addition, both ears were tested using LP and LPmod masking noise, resulting in four different test conditions per test session. The masking noise condition was counterbalanced across subjects. To assure this balanced design of the tests, the four test conditions were programmed to appear in the correct order and participants were automatically led through the complete testing procedure. Participants were instructed to perform the first Earcheck during the weekend (preferably on Sunday) to prevent occupational noise exposure from biasing the SRT results. Reminders were sent to participants that did not perform their test on the given date, until the test was completed. To assess Earcheck’s reliability when used for domestic screening, a subpopulation of 32 employees performed the test in test and retest. The repeated test session was performed on the same day as the first test session.

**Statistical analysis**
The data were analysed using SPSS (version 19.0) and R software (R Foundation 2008, from http://www.R-project.org).

To account for effects of repeated measurements within both ears of each individual, linear mixed effect models were used to estimate the effects of different

---

5 Subjects that logged in with an odd identification number performed the test with the right ear in LP noise first and even numbered subjects started with their left ear in LPmod noise.
masking noise conditions, tested ear, and test repetition on SRTs. This method accounted for the nested data structure of repeatedly testing two ears within one participant, and also handled missing data that were introduced in the data due to some incomplete measurements. The variations between both 'subjects' and 'ear' were treated as random effects. Fixed factors of primary interest were 'masking noise' and 'ear'. Since the systematic difference between both masking noise conditions was known, the primary interest of this study was whether shown effects are similar in both masking conditions. In order to investigate this, two-way interaction terms between each of the predicting variables and 'masking noise' were incorporated in the model as well. Only the factors and interaction terms that showed a significance contribution to the fitted model, tested with conditional F-tests at the 0,05 level, were investigated for significant coefficients of each level. When coefficients proved to be significant, the term was retained in the model.

To assess test reliability, a similar linear mixed effects model was fitted to the data of the subpopulation, also incorporating the fixed factor 'repetition' in the model, to assess systematic differences in SRT outcomes between test and retest sessions. Test-retest reliability can be expressed by two different parameters; the intraclass correlation coefficient (ICC), a relative index of reliability representing the ratio of the between-subject variability to the total variability in the data, and the standard error of measurement (SEM), a measure of absolute consistency reflecting the precision of individual outcomes (Weir, 2005). Two-way random, absolute agreement, single measures ICCs were calculated over the test and retest results within each masking noise condition. This type of ICC takes the systematic difference between test sessions into account and yields results that can be extrapolated to other situations (Weir, 2005). SEM can be derived from ICC, according to the following equation:

\[
SEM = SD \sqrt{1-ICC}
\]  \hspace{1cm} \text{(Equation 7.1)}

where SD is the standard deviation of the test and retest scores from all subjects.

However, when there are two levels of trials, there is an alternative way of calculating the SEM, by dividing the SD of the differences (SDdiff) resulting from paired t-testing by \(\sqrt{2}\).

Finally, the obtained SEM was used to define the minimum detectable difference (MDD), which can be considered as a real change in a subject’s score, above measurement error. MDD was calculated from SEM, according to

\[
MDD = SEM \times 1.96 \times \sqrt{2}
\]  \hspace{1cm} \text{(Equation 7.2)}

creating a so-called 95% confidence interval of change.
Pearsons product correlation coefficients were calculated to determine if there was an association between Earcheck results and hearing thresholds of the screening audiogram.

To further explore the relationship between Earcheck and pure-tone thresholds, linear regression was performed. Since this analysis compared two test methods, Deming’s regression was used. Whereas the ordinary linear regression method assumes that only the independent measurements are associated with measurement errors, the Deming method (Deming, 1964) takes measurement errors for both methods into account. To use this technique it was necessary that the ratio of the variances of measurement error of the two test methods was known. The measurement errors for the Earcheck using either LP or LPmod noise were derived from the test-retest analysis. For the PTA-value obtained by a screening audiogram, a value of 5 dB could be taken (Helleman & Dreschler, 2012).

Receiver operating characteristics (ROC) curves were calculated to establish the correct cut-off values of Earcheck and assess corresponding test sensitivity and specificity in detecting NIHL. Sensitivity is the ability of Earcheck to detect participants that actually show audiometric hearing loss, and specificity is the ability of Earcheck to detect absence of hearing loss in those showing no elevated hearing threshold levels.

**Results**

**Participants**

In total, 256 male construction employees signed up to participate in this study. After application, seven participants did not complete the online hearing tests. Of the 249 participants that did complete the Earcheck session, nine were excluded from analysis based on their answers in the online questionnaire; three of them reported otological problems other than NIHL or presbyacusis, two used hearing aids and four were not native speakers of the Dutch language.

In addition, 30 participants performed their first hearing test at the end of a working day, instead of during the weekend. Since possible temporary effects of noise exposure on their speech intelligibility cannot be ruled out entirely, these subjects were omitted as well. This leaves 210 participants included in the analyses. The distribution of audiometric hearing threshold levels of this population of 210 participants is displayed in Figure 7.1.

Participants were predominantly middle-aged (mean age 45.7 years, SD = 10.0 years). The majority of them, 83.3%, were exposed to job related daily noise levels exceeding 80 dBA. Their mean job tenure in construction was 25.8 years (range 1 – 46 years,
SD = 11.4 years). Two-thirds of the study population reported to often wear their HPDs and only 10.5% never used HPDs. About 74.2% of the study participants indicated that their hearing ability wasn’t good, and 19.1% reported to have tinnitus. In addition, 70.3% of the participants experienced some problems understanding speech in noise, and 27.7% had also problems in quiet.

Of the total population, 32 subjects conducted a test and retest session on the same day to assess test-retest reliability: the test-retest subgroup. T-tests and chi-square analyses indicated that there were no significant differences in characteristics between the subgroup and the total population.

**SRT outcomes**

To assess possible systematic differences in SRT results between masking noise conditions and ears linear mixed effect models were run with random effects for ‘subject’ and ‘ear’, and fixed effects for ‘masking noise’ (LP, Lpmod) and ‘ear’ (left, right). A two-way interaction between these fixed effects was also included in the model, but this did not significantly contribute to it (F[1,403] = 3.31, p = 0.0696). ‘Masking noise’ significantly affected SRT outcomes (F[1,403] = 1109.64, p < 0.001), as was expected beforehand. On average, SRTs obtained in low-pass noise were 4.6 dB better than SRTs obtained under low pass filtered modulated noise, reflecting the differences in SRT between masking noise found in earlier studies described in Chapters 5 and 6. ‘Ear’ did not show any significant contribution to the model (F[1,207] = 1.69, p = 0.194). Since significant right-left differences were not found, data were pooled into a group of 420 ears for further analyses.

---

**Figure 7.1.** Audiometric thresholds of the study population.
Test-retest reliability
A subpopulation of 32 participants, and thus 64 ears, performed test and retest measurements on the same day. The mean test and retest results of this population in both noise conditions are shown in Table 7.1.

Learning effect
The difference in SRT over the two test sessions was calculated. This difference estimated a systematic change in performance due to effects of either learning or fatigue. The results in Table 7.1 show that the performance on the second test was better than on the first, in both masking noises. A linear mixed effect model was fitted to the test-retest data with random effects for ‘individual ear’, and fixed effects for ‘masking noise’, ‘ear’ and ‘repetition’ (test, retest). Two-way interactions between the fixed effects were also included in the model, but these showed no significant contribution to the model. The factors ‘masking noise’ ($F[1,187] = 360.17, p < 0.001$) and ‘ear’ ($F[1,62] = 0.07, p = 0.786$) showed effects similar to those found in the total study group.

The model showed that ‘repetition’ ($F[1,187] = 51.62, p < 0.001$) significantly affected SRT results; test and retest outcomes differed by 1.6 dB averaged over all testing conditions. The insignificant interaction terms of ‘repetition’ with either ‘masking noise’ or ‘ear’ indicated that the learning effect was similar for both ears and in both masking noises.

The results of the test and retest sessions were compared against each other in Figure 7.2; the number of datum points laying below the diagonal line representing absolute agreement indicated the better SRT results of the second test. The intra-session Pearson correlation coefficients of 0.74 for LP noise and 0.65 for LPmod noise showed reasonably good agreement between test and retest results.

Table 7.1. Mean SRTs and test-retest characteristics of Earcheck in both masking noise conditions (n=64 ears).

<table>
<thead>
<tr>
<th></th>
<th>SRT 1 mean (SD)</th>
<th>SRT 2 mean (SD)</th>
<th>SRT diff mean (SD)</th>
<th>ICC</th>
<th>SEM</th>
<th>MDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>-17.3 (3.4)</td>
<td>-18.7 (2.7)</td>
<td>-1.4 (2.3)</td>
<td>0.65</td>
<td>1.63</td>
<td>4.49</td>
</tr>
<tr>
<td>LPmod</td>
<td>-13.0 (2.7)</td>
<td>-14.8 (2.9)</td>
<td>-1.8 (2.4)</td>
<td>0.54</td>
<td>1.68</td>
<td>4.61</td>
</tr>
</tbody>
</table>

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

The model showed that ‘repetition’ ($F[1,187] = 51.62, p < 0.001$) significantly affected SRT results; test and retest outcomes differed by 1.6 dB averaged over all testing conditions. The insignificant interaction terms of ‘repetition’ with either ‘masking noise’ or ‘ear’ indicated that the learning effect was similar for both ears and in both masking noises.

The results of the test and retest sessions were compared against each other in Figure 7.2; the number of datum points laying below the diagonal line representing absolute agreement indicated the better SRT results of the second test. The intra-session Pearson correlation coefficients of 0.74 for LP noise and 0.65 for LPmod noise showed reasonably good agreement between test and retest results.
To explore the observed learning effect of -1.6 dB in more detail, the effect of practice over the subsequent trials was investigated. To do so, the improvements in listeners’ performance over the eight tests during their test and retest session were assessed. Figure 7.3 shows the mean SRT at the particular position during the test, as a function of the trial number (1-8) during the test sessions, determined for both noise conditions separately.

In order to characterize this learning effect, an exponential curve, described by Rhebergen et al. (2008), was fitted to the data according to the following equation:

\[
SRT(n) = SRT_{\text{final}} + a \times 2^{-(n-1)/N}
\]  

(Equation 7.3)

where \( n \) denotes the trial number (\( n = 1, \ldots, 8 \)), \( SRT_{\text{final}} \) denotes the SRT value that is reached after the learning has leveled off, \( a \) is the size of the total learning effect (expressed in dB), and \( N \) is the average amount of tests that is required to halve the size of the learning effect. \( SRT_{\text{final}} \), \( a \), and \( N \) are free parameters that have been estimated by a least squares fit to the individual data.

The exponential curve was fitted for each of the two masking noise conditions, and the estimated parameters are displayed in Table 7.2. The amount of learning \( a \) was estimated to be 2.6 dB in LP noise and 3.5 dB in LPmod noise.

**Figure 7.2.** Domestic Earcheck results of the retest are plotted against results from the first test, separated for masking noise. Solid straight lines represent perfect agreement, dashed lines indicate the learning effect in each masking noise condition.
For LP noise, the parameter $N$ was 1.3, which means that with an observed SEM of 1.6 dB, about 1.9 trials are required to bring the SRT within one SEM of its final value. In addition, about 2.8 trials are needed to bring the SRT within one decibel of the final SRT. Since the learning effect $a$ was larger in LPmod noise, more trials would be required to reach final SRT; the amount of trials required to bring the SRT either within one SEM or one dB of the final SRT value were 2.6 and 3.5 respectively.

### Figure 7.3.
Mean SRT as function of trial number in both test sessions for the test-retest subgroup. Mean SRT is calculated across combinations of listeners at that trial, and for both noise conditions separately.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Mean SRT (dB SNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-20</td>
</tr>
<tr>
<td>2</td>
<td>-18</td>
</tr>
<tr>
<td>3</td>
<td>-16</td>
</tr>
<tr>
<td>4</td>
<td>-14</td>
</tr>
<tr>
<td>5</td>
<td>-12</td>
</tr>
<tr>
<td>6</td>
<td>-10</td>
</tr>
<tr>
<td>7</td>
<td>-18</td>
</tr>
<tr>
<td>8</td>
<td>-16</td>
</tr>
</tbody>
</table>

**Table 7.2.** Parameter estimates obtained by the least squares fit to the individual data, for both noise conditions.

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>LPmod</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SRT_{\text{final}}$ (dB)</td>
<td>-19.47</td>
<td>-15.86</td>
</tr>
<tr>
<td>$a$ (dB)</td>
<td>2.57</td>
<td>3.48</td>
</tr>
<tr>
<td>$N$</td>
<td>1.32</td>
<td>1.40</td>
</tr>
</tbody>
</table>

**Test-retest reliability**

An important measure to evaluate a specific measurement procedure is the test-retest reliability; the consistency of a test’s results across series of observations. Two-way random, absolute agreement, single measures ICCs were calculated over the test and
retest results within each masking noise condition (Table 7.1). For the stationary low pass filtered noise an ICC of 0.65 was found. For Earcheck using low pass filtered modulated noise, the ICC was somewhat lower, 0.54.

The obtained standard errors of measurement for both masking noise conditions were comparable (Table 7.1). The minimum difference to be real, derived from this SEM, indicated that a difference between two Earcheck measures should be at least 4.5 dB to be considered as a real difference in speech intelligibility (Table 7.1).

**Validity**

Test validity relates to the correlation between the test’s results and other, accurate, measures of the same behavior. To assess Earcheck’s validity, the SRT results obtained during the first test session were compared to single frequency hearing threshold levels of the corresponding ear, as obtained by routine screening audiometry. In addition, the pure-tone average (PTA) over the noise-sensitive frequencies 3, 4 and 6 kHz was calculated (PTA\(_{3,4,6}\)) and compared to SRT results. In Table 7.3 the Pearson’s correlation coefficients for the SRTs obtained with Earcheck and the different audiometric parameters are given. All coefficients showed significance at 0.001 level. Significant positive linear associations between speech-in-noise intelligibility measured with Earcheck and audiometric hearing thresholds were observed. The association became stronger with increasing frequency, since the higher frequencies are more prone to noise-induced damage. The correlation coefficients indicated only moderate relations, which was best for Earcheck with LP noise compared to PTA\(_{3,4,6}\) (r = 0.58).

<table>
<thead>
<tr>
<th>Table 7.3. Pearson correlation coefficients of Earcheck outcomes and hearing thresholds (HTLs) at different frequencies from 1 to 6 kHz and PTA(_{3,4,6}). All correlation coefficients are significant at p-value &lt;0.001.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HTL 1kHz</strong></td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>LP</td>
</tr>
<tr>
<td>LPmod</td>
</tr>
</tbody>
</table>

The sensitivity and specificity of Earcheck to detect NIHL relative to the screening audiogram depend on the cut-off values that were used to distinguish between normal and impaired hearing. First, PTA\(_{3,4,6}\) was used as the audiometric reference measure, and hearing loss was defined as having PTA\(_{3,4,6}\) > 40 dB HL. Then the relationships between the sensitivity and specificity of Earcheck were explored in more detail by calculating the receiver operating characteristic (ROC) curve. The ROC curve exhibits the sensitivity versus 1-specificity) of the model for each possible
threshold in the range of estimated probabilities. The area under the curve (AUC) reflects the overall discriminative value of the model (Table 7.4). This was slightly higher for the Earcheck in LP noise (0.82) than when LPmod noise was used (0.79). Then the SRT values of Earcheck above which a subject is classified as having impaired hearing were calculated using the Deming’s regression equation and corresponding cut-off values were -15.1 dB for Earcheck with LP noise and -11.2 dB for Earcheck with LPmod noise. Details of the regression are presented in Table 7.4.

### Table 7.4. Results from Deming’s regression analysis, fitting the relationship between SRT and PTA$_{3,4,6}$, and from fitted ROC curves.

<table>
<thead>
<tr>
<th></th>
<th>Intercept (dB) (95% CI)</th>
<th>Slope (95% CI)</th>
<th>Cut-off PTA$_{3,4,6}$</th>
<th>Cut-off SRT</th>
<th>AUC</th>
<th>Se</th>
<th>Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LP</strong></td>
<td>-20.71 (-21.22 to -20.20)</td>
<td>0.14 (0.12 to 0.16)</td>
<td>40</td>
<td>-15.1</td>
<td>0.82</td>
<td>62.4</td>
<td>86.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>-15.8</td>
<td>0.79</td>
<td>59.3</td>
<td>83.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>-16.5</td>
<td>0.80</td>
<td>67.7</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>-17.2</td>
<td>0.78</td>
<td>67.9</td>
<td>71.0</td>
</tr>
<tr>
<td><strong>LPmod</strong></td>
<td>-14.83 (-15.33 to -14.34)</td>
<td>0.09 (0.08 to 0.11)</td>
<td>40</td>
<td>-11.2</td>
<td>0.79</td>
<td>67.1</td>
<td>71.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>-11.7</td>
<td>0.76</td>
<td>67.3</td>
<td>68.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>-12.1</td>
<td>0.76</td>
<td>70.7</td>
<td>65.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>-12.6</td>
<td>0.74</td>
<td>74.4</td>
<td>58.8</td>
</tr>
</tbody>
</table>

Different criteria are used to define NIHL, and corresponding cut-off values of the SRT outcomes are calculated. Based on these cut-off values, and analysis of the ROC curve, the area under the curve (AUC), sensitivity (Se) and specificity (Sp) are obtained.

Applying these calculated cut-off values showed that Earcheck performed at home had a moderate sensitivity and specificity in detecting NIHL compared to the screening audiogram; respectively 62% and 87% in LP noise, and 67% and 71% in LPmod noise. The observed sensitivity and specificity also depend on the value of PTA$_{3,4,6}$, defined to classify NIHL, and the procedure described above was repeated for different cut-off values. The corresponding parameters are presented in Table 7.4, showing an increase in sensitivity to distinguish between normal and impaired hearing with decreasing PTA-value, although the area under the curve slightly decreased.

Since this study aimed to assess the applicability of Earcheck in NIHL screening purposes, the sensitivity to detect mild NIHL was of main interest. Beginning NIHL
was defined as having $PTA_{3,4,6}$ exceeding 25 dB HL. Using the regression equation, cut-off values of -17.2 dB in LP noise and -12.5 dB in LPmod noise were established (see Table 7.4). For this criterion, Earcheck with LP noise showed a sensitivity of 68% and specificity of 71% in detecting beginning NIHL. For Earcheck with LPmod noise the observed sensitivity was higher, 74%, but the corresponding specificity was lower (59%).

The scatterplots of the outcomes of both Earcheck conditions versus the $PTA_{3,4,6}$ reference measure and the Deming’s regression line are shown in Figure 7.4.

**Figure 7.4.** Scatterplots of SRT values against $PTA_{3,4,6}$ for both masking noise conditions. The black symbols represent normal-hearing ears, grey symbols represent hearing-impaired ears ($PTA_{3,4,6} > 25$ dB HL). Dashed lines represent the cut-off value for SRT that corresponds to hearing impairment. Solid lines represent the Deming’s regression line.

**Discussion**

The internet-based Earcheck has been validated in this study against the current practice of pure-tone screening audiometry. The results of 210 construction workers at risk for noise-induced hearing loss demonstrate that domestic hearing screening by Earcheck has a moderate sensitivity to detect beginning NIHL; the observed sensitivity and specificity are 68% and 71% when stationary low-pass filtered masking noise is used, and 74% and 59% in low-pass filtered noise with modulations.

These values are lower than the test characteristics obtained in a previous evaluation study in which all measurements were conducted under well-controlled
laboratory conditions (Chapter 5); those findings showed very high sensitivity and specificity of 94% and 92% in LPmod noise, and even higher values in stationary LP noise (95% and 98% respectively).

It is well known that speech intelligibility in noise is not perfectly related to pure-tone thresholds (Smoorenburg, 1992), because both methods assess different functionalities of hearing. Since 70.3% of the study population indicated to experience difficulties with understanding speech in noisy situations, testing speech intelligibility in noise has a higher relevance for hearing in daily functioning. Although there are large inter-individual differences, there is a weak relationship between the measured SRT values and self-reported speech intelligibility in noise. In the subgroup reporting no difficulties, the average SRT is -18.7 dB, in the participants who sometimes experience difficulties the average SRT is -17.2 dB and in the subgroup often having difficulties this is -15.0 dB (F[2,403] = 33.20, p < 0.001) in the LP noise condition. Nevertheless, pure-tone audiometry is generally considered the gold standard to which new hearing tests should be compared.

Several issues may be considered to have played a role in the moderate association between speech reception in noise and audiometric results observed in this study. First, field testing may incorporate effects of uncontrollable parameters that increase variability in both screening audiometry and domestic speech-in-noise testing. Second, some methodological considerations should be taken into account. When Earcheck is administered at home, some uncontrollable parameters could affect test outcomes, such as presence of ambient noise, individually set presentation levels, and variety in the equipment used regarding its quality and frequency response. A previous study showed that the differences between normal and impaired performance were smaller at home than under well-controlled lab conditions (Chapter 6). Although an effect of presentation level could not be convincingly established, performing the test at a higher presentation level, might have been responsible for the average improvement of 1.2 dB in hearing-impaired domestic SRTs. On the other hand, unfavorable effects of remote testing were shown in the normal-hearing subjects, who had 1 dB poorer test results at home than in the lab (Chapter 6). In addition, the test-retest reliability in the current study shows that the variability of Earcheck testing at home of 1.63 dB is slightly greater than the SEM of 1.25 dB observed in LP noise in a clinical setting (Chapter 5). Also, industrial screening audiometry is less reliable than diagnostic audiometry obtained in a clinical setting, due to less controlled test conditions. Dobie (1983) observed that workers referred for otologic evaluation have hearing levels that were, on average, 5 dB better than indicated by screening audiometry. Additionally, test-retest variability in industry, expressed in SDdiff was approximately 3 dB higher (6.7-8.3 dB up to 4 kHz) than reported for clinical audiometry (4-5 dB up to 4 kHz) (Dobie, 1983). In conclusion, both
measurement methods compared here may show greater variability in the field than opposed to lab testing. Nevertheless, the quality of the test results in this validation study does reflect the quality of these measurements in daily practice (Grobbee & Hoes, 2009).

Another complicating factor is that for screening on NIHL the insensitivity of Earcheck for conductive hearing losses is to be preferred above the sensitivity of the gold standard, pure-tone audiometry. Listeners with conductive hearing loss (sometimes due to a recent flue or cold) will demonstrate elevated air conduction thresholds. Bone conduction thresholds, allowing for identification of conductive hearing losses, cannot be measured reliably in screening audiometry. As a consequence these patients will be referred for further diagnostic testing in order to assess potential NIHL. Earcheck results, however, will most likely indicate that this is not necessary. So, the discrepancy between Earcheck and screening audiometry for conductive losses is one of the reasons for poorer sensitivity and specificity values, while for this group the Earcheck outcomes may be expected to be more reliable than the gold standard of screening pure-tone audiometry.

Furthermore, some methodological issues that might have affected the observed test parameters should be considered. First of all, the observed sensitivity and specificity depend upon the study population used. Although sensitivity and specificity are considered to be more or less constant and not directly influenced by the prevalence of a disease, it has been shown that they vary according to differences in disease severity, and thus an indirect effect of the prevalence of the disease cannot be ruled out (Grobbee & Hoes, 2009).

The current study population consists of construction workers selected randomly without any criterion regarding their hearing status, as opposed to the study population participating in our previous evaluation study described in Chapter 5. Those participants were selected to have either normal hearing or NIHL, which resulted in a NIHL prevalence of 50%. As a result, listeners with beginning NIHL that are of particular interest for Earcheck validation were not involved in the evaluation study. Current study participants have a continuum of hearing threshold levels, of which 40% is considered hearing-impaired. Now that subjects with small losses are also included, they may blur the previously observed clear distinction between NH and NIHL.

Second, Earcheck outcomes could be influenced by the fact that starting level is fixed at -10 dB for each participant, hence the intelligibility of the first stimulus depends on amount of hearing loss. Figure 7.5 shows the mean signal-to-noise ratio for the different positions in the adaptive procedure, stratified for 1-dB SRT groups and for both noise conditions. Only data points representing means from at least 30 SNRs are shown. This shows that for the majority of participants the SNR for the
different positions in the adaptive procedure decreases as function of presentation number. Although the first seven presentations are omitted in calculating the SRT, Figure 7.5 shows that the mean signal-to-noise ratios still improve after these 7 presentations, at least for the better performing listeners with highly negative SRTs. The final SRT obtained for these listeners might thus be an underestimation of the true speech reception threshold. Consequently, test reliability could be improved by choosing a fixed starting point at a lower SNR than -10 dB to exclude effects of the fixed starting level in NH listeners.

![Figure 7.5](image-url)  
*Figure 7.5. Mean signal-to-noise ratio for the different presentation numbers in the adaptive procedure, separated for both masking noise conditions. Results are shown for different SRT groups, representing results of at least 30 participants. The vertical line divides the procedure in presentations used to approach threshold, and presentations used to calculate the SRT. For the LP noise, the lines represent SRTs of -9 dB until -22 dB, for the LPmod noise, the lines represent SRTs of -7 dB until -17 dB.*

To assure intelligibility of the first stimulus for each respondent, another possibility is to set the starting point individually, for instance by starting the actual adaptive test procedure of 27 presentations after the first two or three incorrect responses are given. Besides changing the starting point, a reduction of the number of stimuli used to calculate the SRT could also lead to more accurate SRT derivations, hence improve test reliability.

Third, the mean learning effect of -1.6 dB observed in the test results is rather high, compared to previous findings that report learning effects for Earcheck ranging from -0.5 to -0.7 dB (see Chapter 5 and 6). However, these results are derived from
well-controlled lab testing as opposed to the domestic test administration in current study. Besides the uncontrollable parameters that may influence domestic test precision, an important difference between both test situations is the amount of instruction offered prior to testing. In the present study, participants received only written instructions; they were displayed on screen, and sent by e-mail and in the letter accompanying the provided headphones. After logging in into the experimental website, the test words are presented on screen and are played once, before test administration begins. Although similar steps were taken in the lab study, the test administration was preceded by an oral explanation of Earcheck when showing the response screen. Before starting the test, participants were able to consult the test leader if there were any questions. The small amount of instruction at home might have led to an underestimation of SRT results of the tests conducted first. This is confirmed by assessing SRTs as a function of trial number. The results of this analysis show that test outcomes improve over the eight trails conducted (Figure 7.3). An exponential curve could be fitted to this data, which estimated that after two trials the result is within one SEM of the final SRT in LP noise, and after three trials in LPmod noise.

The systematic error in repeated measures, which is expressed as the learning effect, is taken into account in the reliability calculations and causes lower ICC values and accordingly higher SEMs. To adequately assess reliability parameters, the significant systematic error should be diminished. In the case of a learning effect, additional trials should be conducted until a plateau in performance occurs, and the ICC should be calculated on the trials in the plateau region only (Weir, 2005). When the 2-3 first measurements from the test-retest subgroup are excluded, the reliability parameters could be calculated over the last 5-6 tests that roughly form a plateau. Indeed, the ICC improved, to 0.69 for LP noise and 0.67 for LPmod noise, and the SEM decreased slightly, to 1.5 dB for both masking noise conditions, indicating less variability in the test. Of course, this is only an optimized estimation of Earcheck’s reliability when performed at home. In order to truly eliminate the systematic error, several consecutive Earchecks should be performed under identical test conditions to reach a plateau in performance, and accurately assess test reliability over the trials in the plateau region.

Practically this implies that Earcheck, when used for NIHL screening purposes, should also incorporate at least two practice trials prior to actual testing. The exact number of trials needed, and whether or not this should be repeated every time the same individual performs the test, should be addressed by future research.

Another alternative method to reduce the learning effect due to insufficient amount of instruction might be to apply Earcheck for testing small groups of construction employees at one central location, for instance at the company site. Although the
The big advantage of easy accessible domestic testing is lost this way, the test is still easy to implement and quick to perform. The test set-up could be adequately installed in a quiet but not necessarily sound-isolated room, and accurate instruction and education could be given.

Moreover, group testing might motivate more workers to participate in hearing screening, and after a controlled administration these workers could continue testing their hearing at home. Earcheck might even be considered as a supplementary low-cost test performed during POHE, to assess additional information about the functional hearing status of the workforce in the construction industry.

The big advantage of an internet-based screening test is the easy accessibility and low requirements of the test. This study showed that it is easy to develop and administer an online experiment to remotely test hearing ability. One possible application of a reliable internet test would be to monitor subtle changes in speech recognition in noise over time. To illustrate this, a pilot experiment was conducted simultaneously with the experiment described here, to investigate a possible effect on hearing of occupational noise exposure during a representative working week. For details, see the appendix.

In short, 140 participants repeated a second Earcheck after ending their last working day of the week following their first test. Their difference in SRT outcomes is 0.5 dB in LP noise and 1.0 dB in LPmod noise. Although this was significantly lower than the obtained learning effect participants still perform better when conducting Earcheck for the second time. Since the difference in SRT was not negatively associated with the intensity of noise exposure, an effect of exposure to noise during the intermediate period could not be proven. Apparently, the effects of noise exposure during a working week on SRT are, if any, too small to be detected by Earcheck in its current form. If an adapted procedure will allow a better precision, it seems worthwhile to repeat such a study. With the current procedure, monitoring of speech recognition only seems to be applicable over longer time intervals, as larger differences are expected to occur over longer time periods.

Test-retest results showed that the minimum detectable difference of current domestic Earcheck testing equals 4.5 dB. This is quite large. Reduction of the learning effect and SEM, by performing more tests, should result in lower MDD values. A considerable reduction is necessary to arrive at SEM and MDD values that are comparable to SEM and MDD values reported for screening audiometry (Helleman & Dreschler, 2012).

The present study was conducted to validate Earcheck using two different masking noises. Again, results of this study show only small differences between test performance using stationary low-pass filtered noise or modulated low-pass filtered
noise. Better discrimination between normal and impaired performance due to additional fluctuations is not proven. Although the test sensitivity to detect beginning hearing loss was slightly higher in LPmod noise, the specificity is lower. Both parameters of Earcheck with LP noise are more balanced, and accordingly the area under this ROC curve was slightly higher than for LPmod noise. In addition, results obtained in LP noise showed stronger correlation with pure-tone audiometry than SRTs in LPmod, and also reliability parameters were slightly better for this condition, as was the case in our previous studies described in Chapter 5 and 6. Furthermore, in LP noise fewer practice trials are needed to reduce the learning effect than in LPmod noise. So, although the differences are small, stationary LP noise is the recommended masking noise condition in Earcheck for domestic screening purposes.

**Conclusion**

Present study has validated Earcheck against a pure-tone screening audiogram. Earcheck provides the benefit of an easily accessible, self-administered alternative method for hearing screening. The Earcheck version with low-pass masking noise has a sensitivity of 68% and a specificity of 71% to detect beginning NIHL. Test-retest reliability was relatively high, 1.6 dB, as was the mean learning effect of -1.6 dB. Since this systematic difference was mainly observed in the first tests, improved instruction and the use of practice trials is recommended in order to reduce this effect and increase test reliability.

The broad internet application still promises an attractive and valuable tool for hearing screening in large populations. The easy accessibility of this test facilitates addressing large segments of the workforce, which is a major advantage over pure-tone screening audiometry in a widespread and very transient population as in the construction industry. Moreover, Earcheck allows for more frequent and on demand testing, and enables testing during leisure time, which facilitates the presence of an adequate noise-free period before testing.

Earcheck cannot replace a pure-tone screening audiogram, but can be a valuable addition to current occupational health practice by better reaching the target population and raising their awareness about noise exposure and the risk of hearing loss. Future modifications in the procedure, including a better instruction and the use of practice trials may be expected to increase the test-retest reliability, and thereby the applicability of Earcheck as a screening test in occupational health care.
Appendix

Applicability of Earcheck in monitoring small changes in speech intelligibility over time.

The major advantage of Earcheck is the easy accessibility due to its internet application. This makes the test not only suitable for hearing screening, but also for monitoring hearing status over time. The potential of Earcheck to detect small longitudinal changes in speech intelligibility is investigated in a small pilot study investigating the effects of noise exposure during one working week on hearing status; the so-called week-effect.

**Methods**

140 of the participants in current study performed the Earcheck test session twice; before and after their working week. This was done in order to investigate the influence of occupational noise exposure on SRTs measured using Earcheck. If a ‘week-effect’ could be established, Earcheck might be used to detect small changes in SRT outcomes, at least averaged over group results. The first Earcheck was always conducted during the weekend and the second test was performed on Friday after finishing work. After completing the second test, participants were linked to a short closing questionnaire, that consisted of 6 questions regarding noise exposure prior to the test and during the intermediate working days. In order to remind participants to conduct the second test on Friday, all received an e-mail after completing the first test.

The noise exposure during the intermediate workweek was estimated using the time-weighted average (TWA) set for the reported job title (Arbouw, 1998). Based on the equal energy principle this TWA was adjusted according to the answers in the closing questionnaire. When subjects stated they were exposed to noise for less than 50% of the time, the estimated TWA was reduced with 3 dB. When subjects reported to ‘often’ have worn HPDs the TWA was also decreased by 3 dB. When a subject was involved in one or more noisy recreational activities during the intermediate period, his noise exposure level was increased by 3 dB. The noise exposure levels derived from these calculations were divided into three categories; no noise exposure (TWA < 80 dBA), moderate noise exposure (TWA 80-89 dBA) and severe noise exposure (TWA ≥ 90 dBA).

**Results**

To assess the week-effect, a linear mixed effects model was fitted to the SRTs of the 140 participants that performed two test sessions. The variations between ‘individual ear’ was treated as a random effect. Fixed factors of primary interest were ‘masking noise’ (LP, LPmod), ‘ear’ (left, right) and ‘repetition’ (test session 1, test session 2). The
main effects of ‘masking noise’ and ‘ear’ were the same as those observed in the total population (Table 7.A1). ‘Repetition’ showed a significant effect on SRTs measured ($F[1,789] = 8.06$, $p = 0.005$), and also significantly interacted with ‘masking noise’ ($F[1,789] = 4.05$, $p = 0.045$). SRTs obtained during the second test session were on average better than results of the first test session; this difference was 0.48 dB in LP noise and 0.96 dB in LPmod noise.

Since these differences in test session were smaller than those obtained in the test-retest subgroup, a second mixed effect model was fitted to the data of all participants that conducted two Earcheck sessions. This model incorporated a fixed factor for the subgroup that performed the tests (‘test-retest’, ‘week-effect’) and an interaction term of ‘repetition’ and ‘subgroup’ (Table 7.1A). The significant interaction term showed that the differences between SRTs obtained during the first and second test session were indeed 0.98 dB smaller in the week-effect subgroup than in the test-retest subgroup ($F[1,948] = 12.81$, $p = 0.004$).

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed factors</th>
<th>Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Masking noise: LPmod</td>
<td>4.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Ear: right</td>
<td>-0.49</td>
<td>0.1059</td>
</tr>
<tr>
<td></td>
<td>Repetition: retest</td>
<td>-0.49</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td>Retest * LPmod</td>
<td>-0.48</td>
<td>0.0455</td>
</tr>
<tr>
<td>II</td>
<td>Masking noise: LPmod</td>
<td>4.81</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Repetition: retest</td>
<td>-1.46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Subgroup: week-effect</td>
<td>0.23</td>
<td>0.6571</td>
</tr>
<tr>
<td></td>
<td>Retest * LPmod</td>
<td>-0.45</td>
<td>0.0330</td>
</tr>
<tr>
<td></td>
<td>Retest * week-effect</td>
<td>0.98</td>
<td>0.0004</td>
</tr>
<tr>
<td>III</td>
<td>Masking noise: LPmod</td>
<td>0.49</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>Noise exposure: none</td>
<td>0.08</td>
<td>0.8302</td>
</tr>
<tr>
<td></td>
<td>Noise exposure: moderate</td>
<td>0.39</td>
<td>0.0402</td>
</tr>
<tr>
<td></td>
<td>Noise exposure: severe</td>
<td>1.07</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Table 7.1A. Coefficients and p-values of the fixed factors in the three different mixed models.

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.
This smaller difference between test sessions in the week-effect subgroup may have been caused by a reduced speech recognition performance due to occupational noise exposure during the intermediate test period. In order to investigate this hypothesis, the relationship between the week-effect, calculated as the difference in SRT outcomes between the first and the second test session, and noise exposure categories was examined.

A third linear mixed effect model, with week-effect as the dependent variable, ‘individual ear’ as random effect and fixed effects of ‘masking noise’ and ‘noise exposure’ (none, moderate, severe) was fitted. Both ‘masking noise’ (F[1,251] = 5.91, p = 0.016) and ‘noise exposure’ (F[1,251] = 3.32, p = 0.038) significantly affected the observed week-effect. There was no significant interaction between these terms. The model showed that the week-effect for LP noise was 1.07 dB for the severely exposed workers, and this was closer to the learning effect of the test-retest group than for the moderately exposed and non-exposed workers who showed almost no week-effect (Table 7.1A). In LPmod noise, the week-effects were on average 0.49 dB larger.

Discussion

The results of this pilot experiment show a smaller difference in SRT outcomes between the first and second Earcheck session measured over a working week than in the test-retest group. Apparently, these smaller difference in the week-effect subgroup cannot be explained by negative effects of intermediate noise exposure. Either, the exposure to noise did not significantly affect SRT results, or the time between the end of workday and test performance at retest was long enough to establish recovery of temporary hearing damage. A third possibility is that Earcheck is not sensitive enough to detect these small differences, even over group results. The relatively high MDD of 4.5 dB observed in this study (Table 7.1) indicates that indeed Earcheck can only detect differences between consecutive SRT results when they are quite large.

Although monitoring small changes in SRT over time is a potentially valuable application of Earcheck, the results of this study cannot prove that Earcheck in its current form can be used for this purpose.