Title: Statistical tool to detect small hearing threshold shifts

Author: Tronstad, Tron Vedul
Academic degree: PHD candidate
Affiliation:
1) SINTEF ICT, Acoustics group, Trondheim, Norway.
2) Norwegian University of Science and Technology, NTNU, Faculty of Information Technology, Mathematics and Electrical Engineering, Department of Electronics and Telecommunications, Trondheim, Norway.

Address for correspondence:
SINTEF ICT, Acoustics group
c/o Tron Vedul Tronstad
P.O.Box 4760 Sluppen
N-7465 Trondheim
Norway

E-mail: tronvedul.tronstad@sintef.no

Keywords: Hearing Conservation/Hearing Loss Prevention; Statistical Process Control; Noise-Induced Hearing Loss; Early Detection;

News summary: Detection of small hearing threshold shifts can improve the safety of people working in noisy environments. Such early warnings can be used to implement individual counteractions at an early stage. Statistical process control has not been used on the hearing before and opens up new possibilities when it comes to early detection of a damage.

Abstract

Objectives: The aim of this study was to present a new tool that can be used in the prevention of noise induced hearing loss. Even if noise is well regulated in most countries, still many workers are exposed to high noise and suffer from permanent threshold shifts. A new strategy to prevent such damage seems to be necessary.

Methods: A statistical process control scheme was presented, both able to detect large and small hearing threshold shifts. Monte Carlo simulations were used to assess the performance of this hearing monitoring scheme. Different hearing threshold shifts were simulated to evaluate the performance of different hearing development scenarios. Additionally a strategy to handle outliers were presented.

Results: It is possible to detect hearing threshold shifts smaller than the standard deviation of the hearing tests performed. This means that a permanent damage below 5 dB can be detected and acted upon. Outliers can also be automatically detected and treated, increasing the robustness of the monitoring.

Conclusion: The suggested statistical framework can be used as an early warning indicator of noise induced hearing loss to improve the workers safety. Individual counteractions can be implemented, reducing the risk of further damage.

1. Introduction

Noise-induced hearing loss (NIHL) is one of the largest occupational health issues in the world. Approximately 22 million workers (17 %) are exposed to hazardous sound in the US (Tak et al. (2009)) and in Germany the number is 4-5 million (12-15 %) (Concha-Barrientos et al. (2004)). In Norway approx. 10 % of the workers report being exposed to hazardous sound more than four hours per day (Lie et al. (2013)). WHO further states that the sound level exposure is higher in developing countries than those just mentioned, indicating even higher prevalence of occupational NIHL (Concha-Barrientos et al. (2004)).

Even if most countries have legislations regulating occupational noise, still there is a high number of noise related health issues. In Norway, for instance, 60 % of the reported work-related health diseases were attributed to noise in 2009 (Lie et al. (2013)), while The National Institute for Occupational Safety and Health (NIOSH) reported that around 30 % of the reported cases from manufacturing industries were hearing loss in the US (NIOSH (2010)). These are alarming numbers, especially when taking into account the individual and social cost associated with hearing impairments.

Hearing conservations programs, including sound exposure monitoring, education, personal safety equipment, and hearing measurements, are effective ways to reduce the incidence of NIHL (Crandell et al. (2004), Keppler et al. (2015)). However, the individual susceptibility to sound/noise makes it very difficult to prevent all damage (Sliwinska-Kowalska and Pawelczyk (2013), Spankovich and Le Prell (2013)). Preventing NIHL is beneficial, both for the society and, not the least, for the affected person, including the nearest family.

In occupational settings the most common way of detecting NIHL is by performing air-conducted pure-tone audiometry. As Stuart et al. point out, due to the large test-retest variation in such tests, an audiologist must see a 10-15 dB change in hearing threshold to be 95 % confident that the difference did not occur due to measurement error (Stuart et al. (1991)). This leads to a very reactive control regime where a large hearing threshold shift must be present before a person is detected as "damaged". New strategies can therefore be necessary to prevent hearing impairments, especially for the most susceptible.

Since there still does not exist any good test that can be used to find the noise sensitive individuals (Vos (2005)), early detection of small permanent threshold shifts could be used as indicators. If a negative development of the hearing can be stopped or reduced at an early stage, the number of people with profound hearing loss can be reduced. The hypothesis is that detection of an incipient hearing damage, can indeed be used to prevent further negative development of the hearing. This yields for all individuals, not only the susceptible ones. The "stone ears", i.e. the robust individuals, is, however, well protected by the current noise limits used in most occupational legislations.

If the hearing is viewed as a "process", where a permanent threshold shift indicates an out-of-control situation, it is possible to construct control charts for the detection of when a NIHL has occurred. A requirement for such statistical process control (SPC) to work, is that measurements of the process are done frequently. This means that reliable tests of an individual's hearing level (HL) must be performed regularly to realize such control charts. In this paper an automatic hearing test, described by Vinay et al., is used as the underlying test for the monitoring of the hearing (Vinay et al. (2015)). The automatic hearing test is currently embedded in a hearing protection device used in offshore installations, and takes approx. 2 minutes to complete. Since the hearing test tones are played through the earpiece of the hearing protection device, the background noise is less of a problem and the need of an audiometric booth is removed. Since the test is quick and the users can initiate the test whenever and wherever they want, this has led to a large increase in hearing test data. The process control regime is, however, not limited to this hearing test, but is possible to apply to any hearing test that is carried out frequently enough.

If a control chart should be useful it must comply with the limits of the current regime. This means that the hearing test itself should give values of the absolute HL and be at least as precise as normal audiometry. In addition, the control chart must be able to detect changes smaller than 15 dB between two consecutive measurements (The Norwegian Labour Inspection Authority (2013)).

This paper will introduce a statistical framework for detection of small threshold shifts using frequent measurements of the hearing threshold. Such an approach can result in earlier detection of a negative development of the hearing and can become a new barrier against NIHL.

The following sections of this paper will first describe the concept of statistical process control (SPC), including different control charts for detecting both large and small changes in a process. Second, a process control strategy for the hearing will be presented, followed by a section with Monte Carlo simulations of the selected control strategy. These simulations will give examples of how the SPC might perform for different types of hearing threshold shifts. Third, real data from an offshore installation will be presented showing how the process control functions in practice. Finally a discussion of the findings is presented.

2. Statistical Process Control

To monitor a process's average value the most common techniques are the $\overline{X}$ chart, for the detection of large shifts, and cumulative sum of deviations (CUSUM), or exponentially weighted moving average (EWMA), for detection of small shifts. $\overline{X}$ is the average value of the subgroup, and is the input to the control chart. If individual observations are given as input the $\overline{X}$ chart becomes an X chart (since the observed value is not a mean value, but unique observations). Additionally there exist control charts for detecting a shift in the process variability, e.g. moving range chart (mR chart), range chart (R chart) and an s chart for monitoring the standard deviation. Below, only X, mR and CUSUM charts are presented, since these are used in the consecutive discussion. The reader is referred to e.g. Montgomery for a more thorough description of all the different control charts (Montgomery (2013)).

2.1. Average Run Length

One measure for evaluating the performance of different control strategies is the average run length (ARL). This is a number describing how many measurements, on average, are needed to detect a certain shift in a parameter value. Even if some authors disagree about using ARL as a performance measure, it is quite a common way to compare different control charts (Montgomery (2013)).

A sensitive control chart has a small ARL value when the process has gone out of control. This means that the error is detected quickly when a change in the process has occurred. The control charts should also have a large ARL value when the process is in control. This means that the false alarm rate is low. These two situations can be compared to the usual type 1 and type 2 errors in hypothesis testing. Often the notation ARL_0 is used to describe the in-control situation, and ARL_1 describes the process when it is out of control. One should also note that by decreasing ARL_1 the ARL_0 will also decrease. It is therefore necessary to consider both situations when setting control limits.

If more than one control chart is implemented for a process at the same time, the false alarm rate will be affected. The reason is that an alarm signal is typically given when at least one of the control charts indicates a process which is out-of-control. A Bonferroni correction can be applied to the ARL_0 to adjust for this (Hawkins and Olwell (1998)). If m control charts are used, the actual ARL_0 for the joint process control can be estimated by

$$ARL_{0,\text{Group}} \approx \frac{ARL_0}{m}$$

This means that when the control limits are to be set, the ARL_0 value for each single control chart must be multiplied with m to get the wanted false alarm rate for $ARL_{0,\text{Group}}$.

2.2. Run Length Percentile

In this paper another measure will also be used to evaluate the process control charts; the run length percentile. Other papers have also used such representation of the quality of the control charts (Khoo and Quah (2002), Chakraborti (2007)). The rationale for this approach is that the distribution of the run length values is often highly skewed, especially when the process is in control. This means the ARL_0 value, which is the arithmetic mean of the run lengths, can give a false impression of how good the control chart is. Percentiles can be useful for setting the control limits for the control chart. One might, for instance, specify that only 3 % of the control charts should have a run length less than 100 when in control. In this paper Monte Carlo simulations will be used to find such percentiles.

2.3. SPC Implementation

SPC is often divided into two phases; an initial phase (Phase I) where the control charts are constructed, and a control phase (Phase II) where the charts from Phase I are used to monitor the process.

In Phase I the goal is to specify the control limits to be used in Phase II. Several papers have looked at how many observations are needed in Phase I, in order to get reliable estimates of the process parameters, and how to adjust the control limits for a given number of observations. Jensen et al. did a review of such literature, and show that the recommendations vary from approx. 100 to 300 observations for individual control charts (Jensen et al. (2006)). Hawkins have, however, pointed out that many use much lower numbers, often around 25 observations, in practice (Hawkins (1987)). This can, as he also states, give wrong estimates of the parameters, especially the standard deviation, which in turn can affect the performance of the control charts.

Another procedure, called self-starting control chart exists, where Phase I can be omitted and the unknown process parameters are continuously estimated as new observations are available. This type of chart is especially suitable when the number of observations is small, and when it is cumbersome to collect more samples.

2.4. Self-starting Control Charts

Self-starting control charts do not need a Phase I, since the parameters needed are estimated continuously as new observations are available. This makes self-starting control charts especially attractive for small run-lengths, since the monitoring of the process can start immediately.

Quesenberry showed an implementation of such a control chart, defining a Q value to base the monitoring on (Quesenberry (1991)). The Q value if defined as

$$Q_i = \Phi^{-1}\left[G_{i-2}\left(\sqrt{\frac{i-1}{i}}\left(\frac{\left(X_i - \overline{X}_{i-1}\right)}{S_{i-1}}\right)\right)\right], i = 3,4,\ldots$$

where $\Phi^{-1}(\cdot)$ is the inverse of the normal standard distribution function, and $G_m(\cdot)$ is the Student's t distribution with m degrees of freedom. In addition we have

$$\overline{X}_i = \frac{\sum_{n=1}^{i} X_n}{i}, \quad S_i^2 = \frac{\sum_{n=1}^{i}\left(X_n - \overline{X}_i\right)^2}{i-1}, \quad i = 2,3,\ldots$$

Quesenberry showed that the Q_i values are independent and normal distributed.

Comparing the self-starting control charts with the control charts with known parameters, the performance is worse. One should, however, not compare these since they are based upon two completely different premises. If the process parameters are known, one should always use this information in the process control, but if the parameters are unknown one must decide whether to estimate the parameters in a Phase I part, or use a self-starting control chart. If process observations are readily available, the first is recommended, but if this is not the case, the latter will be more efficient.

2.5. Q Chart

If one uses the Q values as input to the X chart, one can make a self-starting control chart for the individual values. This is called a Q chart, and since the Q values are normal distributed it is easy to set control limits that are the same for all processes that should have the same ARL_0. A disadvantage of this chart is that the Q values are normalized, thus they can be difficult to interpret, for instance in a plot.

2.6. CUSUM Q Chart

The Q values can also be used as input to the time weighted control charts. Quesenberry proposed, for instance, both a CUSUM and an EWMA control chart (Quesenberry (1991)). Another proposed method is the adaptive CUSUM of the Q chart (ADQ) (Li et al. (2010)), which can be even more effective at detecting a range of shifts in the mean value. The simulations presented by Li and Wang are, however, not exclusively in favor of the ADQ and will not be further described in this paper. The CUSUM Q value can be found using the equations above and replacing X_i with Q_i and using the fact that Q is standard normal distributed :

$$S_i^+ = \max[0, Q_i - k + S_{i-1}^+]$$

$$S_i^- = \max[0, -k - Q_i + S_{i-1}^-]$$

where $S_0^+ = S_0^- = 0$ and k is a reference value often called an allowance, or slack value, that can be chosen for optimal response to a shift of a specified size (Hawkins and Olwell (1998)). Notice that the S_i values do not increase unless the $|Q|$ values are larger than the reference value k.

2.7. Erroneous Data

Erroneous data, or outliers, are a challenge for any type of control chart. Large outliers will be interpreted as large shifts in the process, and will be flagged as an out-of-control situation if not taken care of. Additionally, since the estimation of both the mean and standard deviation will be affected by such data points, the performance of the control chart can be deteriorated for the rest of the process control.

There are two common approaches to handle outliers; truncation and winsorization. Truncation of values means that the values are removed from the data set, while winsorization means that the suspect observation is replaced by the nearest "non-suspect" data value.

Hawkins proposed a solution for the time weighted control charts where outliers are detected using a maximum level for the Q values (Hawkins (1980)). If any Q value exceeds a preset limit, W, the value is winsorized to this limit. If W is selected wisely this method protects well against large outliers, but all data point will contribute to the control chart. One must also be aware that the control limits must be adjusted to give the same ARL when winsorization is applied.

It is, however, not enough to winsorize the Q values. As mentioned, if the outlier value in the underlying dataset is not handled, it can seriously deteriorate the estimation of especially the standard deviation, but also the mean value. A possible solution is to use more robust estimators, e.g. the median of the data as mean estimator, and the median of the moving range as estimator for the standard deviation. Using more robust estimators will, however, in general deteriorate the performance of the control chart. Another solution is to either truncate or winsorize the data used in the estimation as well.

3. SPC for the Hearing

As mentioned above, one might either use a Phase I to estimate the process parameters, or use a self-starting control chart. A challenge for the monitoring of the hearing is to get enough data points in a Phase I stage to get good estimates for the HL and variability. If 100 measurements should be used, as some authors recommend, and the test person performs one test each day, then 20 work weeks will pass before Phase II can start. A possible solution could be to have an initial test period where the test person performs several hearing tests each day to reduce the length of Phase I. It is, however, a laborious task to perform many hearing tests, hence such an intensive test period could lead to higher variability due to exhaustion. A self-starting regime will therefore be elaborated in this paper.

3.1. Specifying In-Control Parameters

One of the most important parts of the control chart procedure is to establish the in-control parameters. If the estimate of the process mean or standard deviation is wrong, then the control chart performance can be severely deteriorated (Jensen et al. (2006)). This is especially true for the time-weighted control charts that are used to detect small changes.

There are two values needed for the hearing process control chart; the HL (process mean), and the test-retest variability (process standard deviation).

Several studies have looked at the test-retest variability of both manual and automatic hearing tests. In general the findings are that the standard deviation is below 5 dB, and often reported around 2-3 dB (Jerlvall and Arlinger (1986), Stuart et al. (1991), Henry et al. (2001), Smith-Olinde et al. (2006), Vinay et al. (2015)). For the estimation of the variability there are therefore two possibilities. Either one can use a common standard deviation value, or one can estimate the intrasubject variability. Since the variation is an individual value, estimation of the standard deviation will probably increase the reliability of the control chart. In addition, if a person is very consistent when performing hearing tests, it will be possible to detect small changes faster than if a person's responses vary a lot. Estimation of the intra-subject variability will therefore be used in this paper.

The process mean, i.e. the HL baseline, will also be an individual value. It is possible to estimate this value based on ordinary audiometric data, measured by, for instance, an audiologist. However, since the test-retest variability, as mentioned above, is similar for manual audiometry done by professionals and automatic tests, using ordinary audiometric data as HL baseline, rather than using the data points that are used in the control chart, would not increase the reliability of the control chart.

Presbycusis, or age-related hearing loss (ARHL), must also be taken into account. As a person gets older the hearing inevitably gets worse. Whether this should be corrected for, or flagged as a process which is out-of-control, is something that must be decided. One solution would be to adjust the expected HL with an estimation of the ARHL. The progress of ARHL is described in ISO 7029 (ISO 7029 (2000)), hence one could use this estimation as input to the control chart. It is, however, difficult to give an accurate estimate for an individual’s level since the variation in expected threshold values also increases with age. This increased variation suggests that ARHL should preferably not be corrected for in the control chart itself, but left to the medical personnel in follow-up evaluations, once an out-of-control signal has been detected. Using a self-starting control chart will, however, take ARHL into account to some extent. Since the HL is estimated continuously throughout the monitoring, a slow shift in the true HL will lead to a drift of the estimated value. This effect will be elucidated in the following Monte Carlo simulations.

3.2. Proposed Control Chart

Even if the self-starting control charts can have slightly worse performance than a thoroughly executed Phase I/Phase II regime, arguments could be made that it is more important to start the hearing monitoring fast than to detect very small hearing threshold shifts.

A chart of the HL's will be presented together with the estimation of the mean value. If one measurement is more than 3 standard deviations above the mean value the test person must be prompted to perform a new measurement. If the consecutive measurement also is more than 3 standard deviations above the mean value, it is stated that a large hearing threshold shift has occurred and an out-of-control signal is given. This approach is very similar to the recommended procedure given by the Norwegian Labour Inspection Authority (2013) if a standard deviation of 5 dB is assumed. The difference is that instead of using a common standard deviation, the individual's variability is used to determine the limit. This means that for an individual with consistent answers (i.e. small standard deviation) smaller hearing threshold shifts will be detected, and vice versa. In addition a CUSUM Q chart is used to detect small persistent changes. As mentioned above, an EWMA chart could also been used, but CUSUM is preferred because of its ability to estimate the point where the process went out of control. This can be useful when performing an anamnesis after a hearing threshold shift has been detected.

It will also be important that the control chart is robust against outlier data points. The winsorizing Q value approach will therefore be used, as described above, together with the proposed rejection of outliers in the parameter estimation.

3.3. Deciding Control Limits

Before the control charts can be constructed, one must decide which control limits should be used for the CUSUM Q chart. The control limits are affected by two choices that have to be made: ARL_0 and ARL_1.

The first choice depends, among other things, on the cost associated with flagging an out-of-control signal. When the control charts detect a possible change, the test person must be sent to the occupational health service. As mentioned above, most companies test their employees once every three years, hence it might be a reasonable choice that the control charts should not give false alarms more frequently than this. Otherwise the control chart will lead to an increased expense for the company since the employees must visit the occupational health service more often than before.

To find a control limit one must also decide on how many observations (hearing tests) the employees should perform per year. In this paper it is assumed that hearing tests are performed approximately once per week. If one decides that only 5 % of the employees should have a false alarm rate less than 100 (leading to approx. three years between each false alarm) the control limits become 5.9 and 4.2 for $k = 0.5$ and $k = 0.75$, respectively.

The second choice will be affected by the control limit settings. If many false alarms cannot be tolerated, the ability to detect changes will be affected adversely. However, by changing the allowance value the performance of the control chart can be adjusted to some extent. This will be elucidated in the next section where Monte Carlo simulations of different shifts of mean value will be tested. Increasing the allowance value, k, will make the control chart more robust against variations in the observation, but will reduce the ability to detect small changes.

4. Simulations

Monte Carlo simulations, implemented in Matlab (MathWorks (2016)), were used to evaluate the performance of the CUSUM Q chart. The following situations were simulated:

- No shift (NS)
- Step shift (SS)
- Ramp shift (RS)
- Presbycusis (P)
- Comparison with current regime (C)

NS simulates the non-damaged ear, and will illustrate the false alarm rate (ARL_0). SS will simulate a sudden hearing loss that could occur due to a loud sound exposure. The RS will simulate a progressive worsening of the hearing. Even if such a linear approach most likely is too simplistic to describe an actual developing hearing loss, it will give insight into the effect of a progressing hearing loss with different rates of progression. Finally P will simulate how age related hearing loss is detected by the control charts.

Both SS and RS used data sets with 2000 observations in each. The NS situation used 20 000 observations to get better estimates of ARL_0. The individual observations were random values from a normal distribution with zero mean and a standard deviation of 5. These values were chosen since they are typical values for a hearing test as mentioned above. If an out-of-control signal was flagged during the simulation, the run length was saved and the simulation stopped. If no out-of-control signal was detected, the run length was set to 2000. One should be aware that this will lower the estimated ARL value somewhat.

For the SS and RS situation the simulation was rejected and replaced with a new simulation if the out-of-control signal was flagged before the shift was introduced (false alarm). This was considered acceptable since information about the false alarm rate is found from the NS simulations.

The development of presbycusis is described in the international standard ISO 7029 [23], and is expressed by equations of the same form:

$$HL(\mathrm{Age}) = \alpha(\mathrm{Age} - 18)^2$$

where α has different values for the different frequencies and genders, and Age is the person's age in years. The equation is valid for ages between 18 and 70 years.

Two cases were used as input to the Monte Carlo simulations; 6 kHz for males, and 3 kHz for females. These are the cases with the most extreme values, $\alpha_{\mathrm{male,6}\ kHz} = 0.018$ and $\alpha_{\mathrm{female,3}\ kHz} = 0.0075$, respectively, so simulations of the other frequencies would end up between these two curves. Different observation rates, from once every third year to 200/year, were simulated, and the observations were distributed evenly throughout time. A random distribution of the observations was also simulated, but results are not shown because they were almost identical to those for an even distribution in time. Such a randomizing would correspond to hearing tests that are not performed with regular time intervals.

Erroneous data points, or outliers, have been simulated by adding 25 dB to observations in a randomized matter. No such outliers were, however, allowed during the first 10 observations. The reason is that early outliers are quite detrimental for the estimation of the process parameters, and it is rather easy to prevent such errors in practice. By using a training session where the test person performs 5-10 tests under close supervision, large outliers can be prevented. The outliers were also only applied as an increased HL, i.e. an apparent worsening of the hearing. The reason is that the most probable reasons for outliers: background noise and lapse of attention, as well as other interruptions, all lead to an elevated threshold. Outliers in this direction will also have the worst effect on the process control since they will increase the mean and standard deviation, leading to a less sensitive process control.

How quickly threshold shifts will be detected, that is, how low values of ARL_1 can be reached, depends on how well estimated the parameters are. If the shift arises early in the monitoring, the detection ability will be worse since the parameters in general become more precisely estimated as more observations are available. Observation number 50, 100, 150 and 200 will be used as starting points of the SS and RS to show this effect.

To compare the presented method with the current regime three large hearing threshold shifts are simulated; 10 dB, 15 dB, and 20 dB.

4.1. No Shift, NS

Figure 1 shows the run length distributions for two sets of selected control parameters for the CUSUM Q chart, together with the ARL_0 values, when no threshold shift is introduced. As can be seen, the two distributions are far from normal, but approximately the same. The ARL_0 values are also almost identical, which is expected since the control parameters have been selected such that they should be similar. It could be noticed that the non-normal distribution leads to that the mean value and median value are quite different. Here, this implies that around 63 % of the tested persons will have a false alarm go off more often than after 2700 measurements, which is the mean value.

[Insert Figure 1 here]

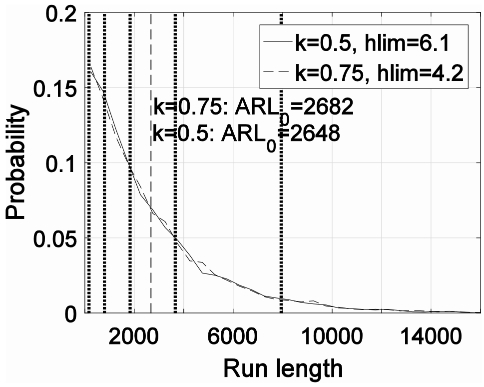

Figure 1: CUSUM Q chart run length distribution for the no-shift situation. The ARL_0's, shown with the dashed line, are almost the same for the two sets of parameters. The dotted lines show, from left to right, 5 %, 25 %, 50 %, 75 % and 95 % percentiles.

4.2. Step Shift, SS

The results from the SS can be seen in Table 1. Focusing on the 95 % percentile, which describes the situation where most test persons are included, we see that a shift smaller than one standard deviation will almost never (run length 133 (k = 0.5) and > 200 (k = 0.75)) be detected by the CUSUM Q chart, if the shift occurs early in the process (after 50 observations). The same shift will, however, be detected after only 12 (k = 0.5) and 16 (k = 0.75) observations for half of the test subjects (the 50 % percentile). If a step shift of one standard deviation occurs later in the monitoring, using k = 0.5 proves most efficient, and the shift is detected for 95 % test subjects before 34 observations have been made. A step shift of two standard deviations is detected before 8 observations are made for all situations simulated.

[Insert Table 1 here]

4.3. Ramp Shift, RS

Table 2 shows the run length percentiles for the RS scenario. The table only shows results for two RS rates; 0.1 dB/obs. and 0.4 dB/obs, respectively. The table shows that the onset time for the ramp shift is not as critical for the detection ability as for the step shift since the results are similar for all onset points.

[Insert Table 2 here]

Figure 2 shows an estimation of how large the threshold shift is before it is detected with the different rates. The ARL value is simply multiplied with the rate to calculate this estimate. Even if the numbers of observations are larger for the small rates than for the large rates, the accumulated threshold shift is actually smaller for the smaller rates. So, it seems like the SPC approach can be efficient for detecting such small but steadily progressing changes in measured values, especially if the hearing is tested frequently.

[Insert Figure 2 here]

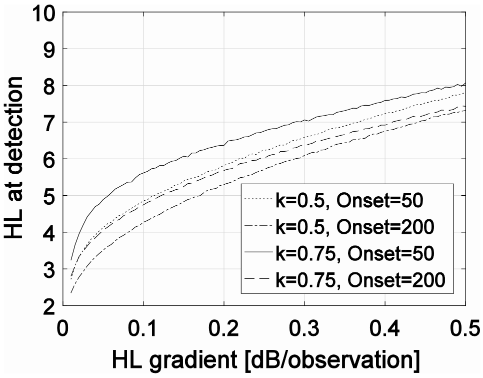

Figure 2: Estimation of amount of hearing loss accumulated before a ramp shift is detected.

4.4. Presbycusis

The results from the P simulations are shown in Figure 3. If a male tests his hearing more often than 10 times per year, the ARHL will typically be detected before 15 years have passed. For females the same detection ability requires approximately 30-40 observations per year. If 10 hearing tests are performed per year for a woman, the ARHL will be detected after 20 years, i.e. around the age of 38.

[Insert Figure 3 here]

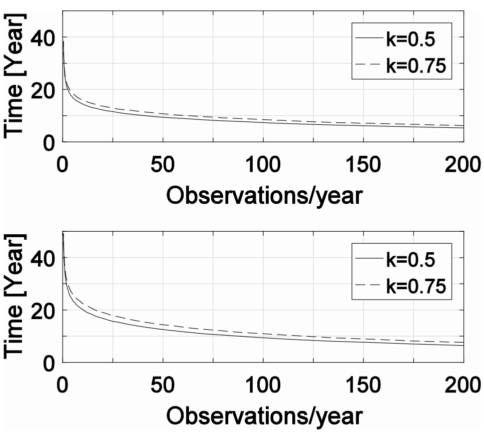

Figure 3: Plot showing when a typical presbycusis will lead to flagging of an out-of-control process, if the process control is started at age 18, for different numbers of observations per year. Upper: Male, 6 kHz ($\alpha = 0.018$). Lower: Female, 3 kHz ($\alpha = 0.0075$).

4.5. Erroneous Data Points

To illustrate how outliers can affect the process control, a simulation with erratic erroneous data points inserted, without the outlier detection method was performed. Figure 4 shows how the process control collapses when no outlier detection is applied, even with as few as five outliers randomly distributed among the 200 first observations.

[Insert Figure 4 here]

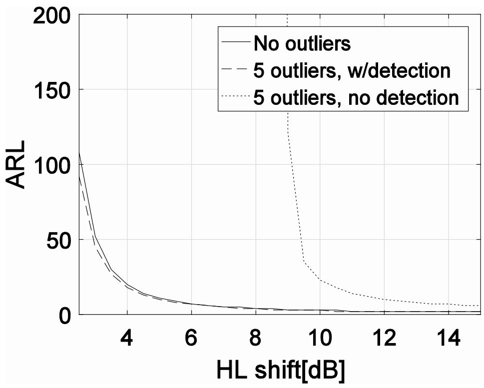

Figure 4: Illustration of outlier sensitivity. The plot shows the average run length (ARL) for different shifts in hearing level (HL). Three situations are shown; no outliers, 5 outliers with outlier detection and winsorizing of outliers, and 5 outliers without any outlier detection or counteractions.

4.6. Comparison with Current Regime, C

The focus so far has been the detection of relative small hearing threshold shift, but it is also important that large shifts can be detected quickly. To assess this a comparison between the presented hearing monitoring system and the current regime is performed. Before this can be done there is a few assumptions that has to be made. First of all it is assumed that all hearing measurements are normally distributed with 3 dB and 5 dB standard deviation (both situations are compared). Second, it is assumed that the common step size of 5 dB is used in the pure tone audiometry. Further the procedure described by the Norwegian Labour Inspection Authority is followed (The Norwegian Labour Inspection Authority (2013)). This means that when a shift of at least 15 dB is measured it has to be verified by a second measurement. It is also assumed that the minimum rate of one measurement every three year is used.

Figure 5 shows the probability distribution of measurements performed with pure tone audiometry using the mentioned assumptions. The figure shows that it is, for instance, a 60 % ($P_{\sigma=3\ \mathrm{dB}}(0\ \mathrm{dB}|0\ \mathrm{dB}) = 0.60$) chance of measuring the correct HL when the standard deviation is 3 dB, and a 38 % chance ($P_{\sigma=5\ \mathrm{dB}}(0\ \mathrm{dB}|0\ \mathrm{dB}) = 0.60$) when the standard deviation is 5 dB. The probability of measuring larger values than those presented in the figure is so small that it is neglected in the following discussion.

[Insert Figure 5 here]

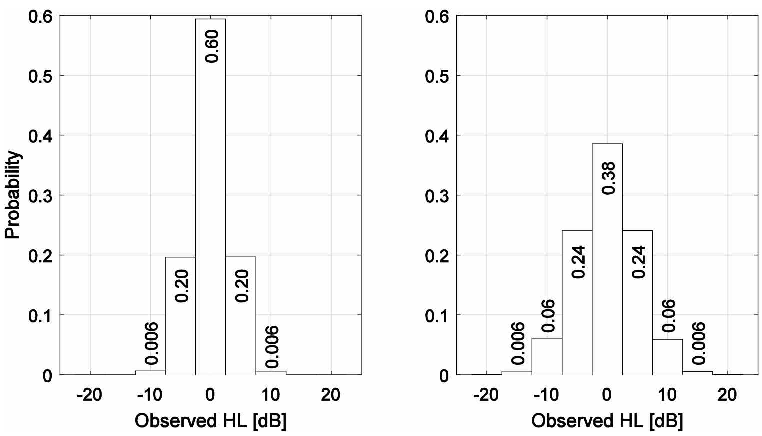

Figure 5: Distribution of the results from hearing measurements with 3 dB (left) and 5 dB (right) standard deviation, and 5 dB step size in the audiometry test. The observed HL is the difference between the measured and the actual hearing level.

To compare the two methods three different step shifts are assessed; 10 dB, 15 dB, and 20 dB. The probability of detecting the shifts with the current regime can be found by calculating the probability of measuring a value larger than 15 dB, two times in a row, using the probability distribution in Figure 5.

The probability of measuring 5 dB, or more, than the actual threshold shift is the sum of $P(5\ \mathrm{dB}|0\ \mathrm{dB})$, $P(10\ \mathrm{dB}|0\ \mathrm{dB})$, and so on. This means, for example, that the probability of measuring a single value at, or above 15 dB when the actual hearing level is 10 dB is 30.9 % ($0.24 + 0.06 +$

$0.006 + \cdots = 0.309$), when the standard deviation is 5 dB. The probability of measuring two consecutive values at, or above 15 dB is therefore only 9.6 % ($P(\geq 15\ \mathrm{dB}|10\ \mathrm{dB})^2 = 0.309^2 = 0.096$).

Table 3 shows the probability of detecting a true threshold shift of 10 dB, 15, dB, and 20 dB with the current regime.

[Insert Table 3 here]

The probability of detecting the same shifts using the presented method above is found by performing Monte Carlo simulations. Figure 6 shows the results from 10 000 simulations of the different threshold shifts.

[Insert Figure 6 here]

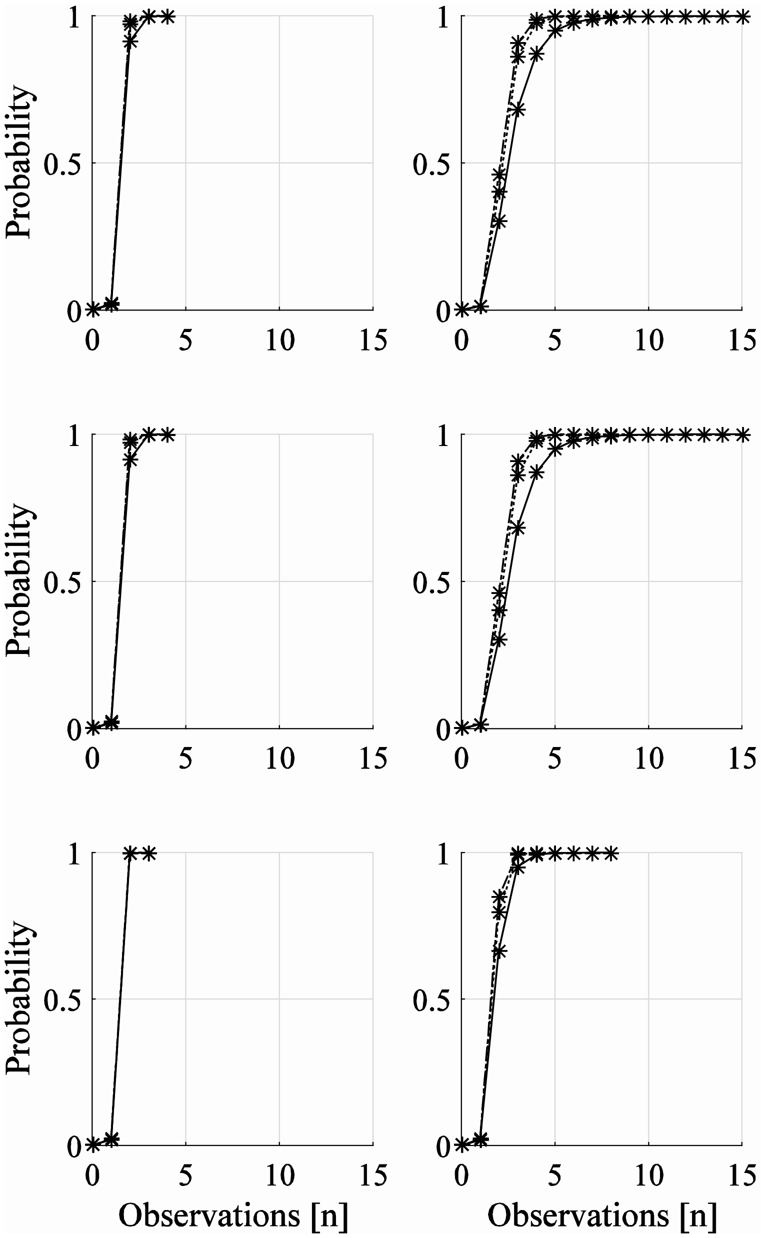

Figure 6: Probability of detecting a 10 dB (upper), 15 dB (middle), and 20 dB (lower) hearing threshold shift using the presented method. Left column: 3 dB standard deviation. Right column: 5 dB standard deviation. The hearing threshold shifts are introduced at three different points; observation 20 (solid lines), 50 (dotted lines) and 200 (dashdotted lines).

Figure 6 shows that the probability of detecting a large shift quickly increases and exceeds the probabilities found by the current regime (see Table 3) after two (3 dB standard deviation) or three (5 dB standard deviation) observations. This means that if more than one hearing measurement is performed per year, the performance will be better with the presented method. It is also possible to observe that the probabilities quickly approaches 100 % reliability as more than 5-10 observations are made.

5. Real Data

Currently the hearing monitoring regime is being used by Statoil ASA at two offshore installations in the petroleum's industry. Figure 7 shows an example of a time series of the hearing level measurements and the corresponding process control from one of the users. The figure show the three frequencies being tested (3, 4 and 6 kHz) at both ears. As can be seen the user has been measuring the hearing for approximately three years, and a total of 54 measurements have been performed during this period.

[Insert Figure 7 here]

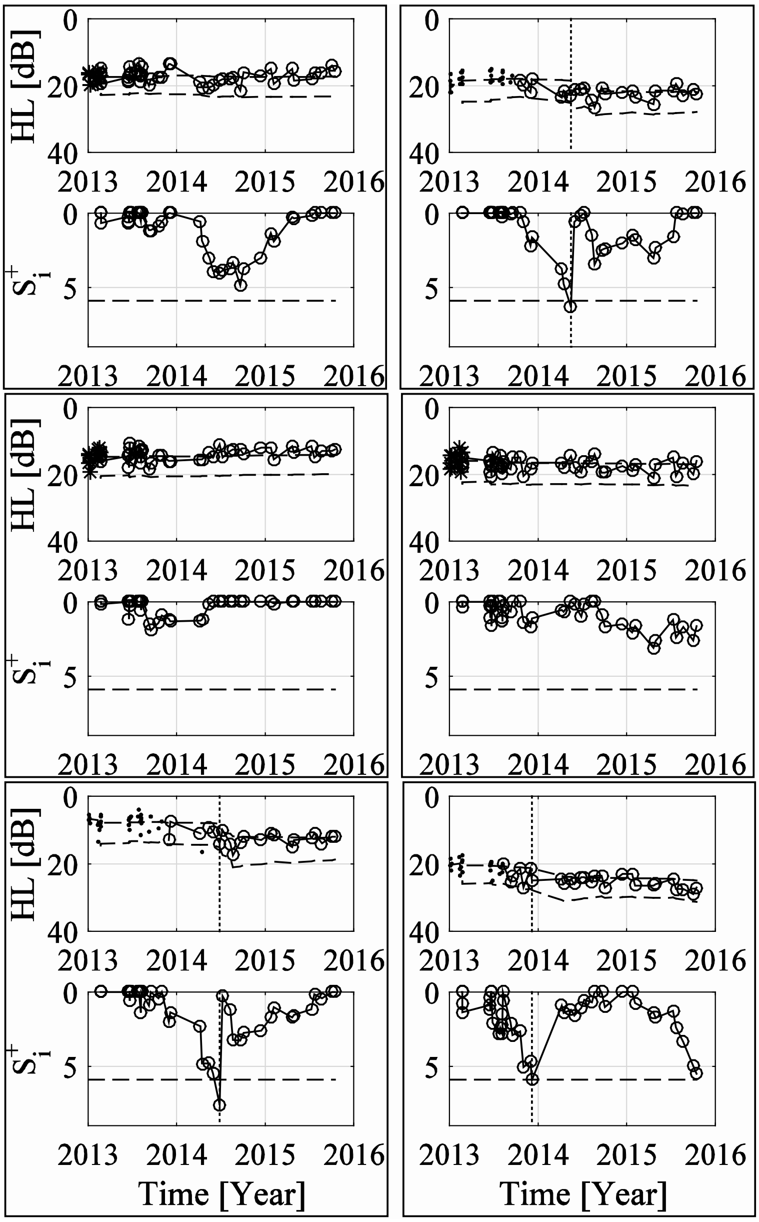

Figure 7: Plot of the hearing level (HL) measurements and the statistical process control performed on these from one person performing regular measurements for almost three years. Left column: Left ear. Right column: Right ear. Upper rectangles: 3 kHz. Middle rectangles: 4 kHz. Lower rectangles: 6 kHz. The lower plot in all the six boxes in Figure 7 show the CUSUM Q value from the statistical process control presented above, using $k = 0.5$, and $h = 5.9$. In three of the boxes one might see a dotted vertical line. This indicates the point where the CUSUM Q value, S_i^+, crosses the control limit signalling that the process is "out of control". By calculating the difference between the mean value of the last five observations before the processes gave a signal, and the mean value of

the observations preceding these five, it is possible to estimate the hearing threshold shifts that are detected. Table 4 shows a summary of the details from the process control.

[Insert Table 4 here]

6. Discussion

The process control proposed in this paper can be tuned to detect shifts in the hearing threshold that are larger than a specified level, for instance 5 dB. For individuals with consistent hearing test answers, i.e. smaller variation, it would also be possible to detect even smaller threshold shifts. The problem is that shifts below 5 dB can be difficult for an audiologist to verify. Even if it is possible to use a smaller step size than the 5 dB normally used in an audiometric test, e.g. 1 or 2 dB, the standard deviation in these tests is still around 5 dB (Jerlvall and Arlinger (1986)). If statistical process control of the hearing proves accurate at detecting hearing threshold shifts, then it would be possible to implement counteractions based on the outcome from the control charts. More practical experience is needed before this can be concluded.

If the process control should function optimal, many hearing measurements must be performed. This means that hearing measurements must be made readily available for the test subjects and that it is necessary to move the testing out from the occupational health service office to new platforms. Since the process control does not need calibrated input it is possible to use computer based tests with off-the-shelf soundcard and headphones, or app-based hearing tests using smart-phones or tablets. As long as the same equipment is used, the method will detect changes in hearing threshold. Calibrated measurements can be collected in the traditional way through the occupational health service. Such data can be used in the interpretation of the results from the process control.

Another observation for the RS situation is that even if the run length increases as the gradient gets smaller (see Table 2), the average accumulated hearing loss decreases (see Figure 2). This means that it is beneficial to increase the number of hearing measurements, which will lower the gradient for a given ramp shift, if the goal is to detect a shift as early as possible. Since increasing the number of observations also will affect the estimation of the mean value, this benefit was not obvious.

The simulations of presbycusis also showed that if one measures the hearing more often than 10 times per year, from the age 18, the hearing will be flagged out of control after around 10-15 years. This must be taken into account when considering a detected threshold shift. If the process control is initiated when the person is older the situation might change since the threshold shift gradient increases with age. This was not further explored.

The international standard, ISO1999, estimates that the time development of NIHL is greatest the first ten years of exposure (ISO 1999 (2013)). The expected time development of the hearing must therefore be taken into consideration when the control chart is constructed.

It has also been shown that outliers can have a detrimental effect on the process control, if outlier counteraction measures are not used, see Figure 4. Even as few as five large outliers among the 200 first observations (that is, 2.5 % erroneous data points) will make the process control unable to detect hearing threshold shifts that are smaller than 10-12 dB if no counteractions are implemented. Using the presented winsorizing approach all the large outliers are detected, and the performance of the control chart is obtained.

The comparison between the presented method and the current regime also showed that large shifts will be efficiently detected. Especially important is the fact that the reliability of the new method quickly approaches 100 % if more than 5-10 observations are performed. This will improve the sensitivity of the hearing monitoring. It also shows that it is not necessary to perform many hearing measurements to outperform the current regime, but fewer measurements will lead to a reduced ability to detect small hearing threshold shifts.

Using the process control on real data also showed that small (≈ 3 dB) hearing threshold shifts can be detected. Detection is, however, only the first step in the prevention of hearing loss. After a possible hearing threshold shift has been detected a multi-step process must be initiated. These steps include that it must be assessed if it is a false alarm or not, if there are any natural causes for the shift (e.g. a common cold), etc. If it still is reason to believe that the threshold shift is noise-induced, then appropriate counteractions must be considered to reduce the risk of further negative development. The different steps are not thoroughly discussed here, but is presented in a corresponding paper looking at the use of statistical process control on the hearing on a more superior level (Tronstad (Forthcoming 2017)).

A possible improvement of the method described, is to implement a multivariate control chart. Since several frequencies are tested on both ears, it is possible to exploit the probable covariation between these tests. If a person is experiencing a NIHL it is likely that more than one frequency and/or ear is affected at the same time. This can be used to either lower the detection limit, or make the control chart more robust. Such an approach will be explored in a future paper.

7. Conclusion

This paper has elucidated a possible monitoring scheme that can be used to detect small hearing threshold shifts. The Monte Carlo simulations showed that it is possible to detect small step shifts in the HL, but that the onset point is of some importance. Early onset means that the hearing monitoring has less time to estimate the actual mean value and standard deviation, and this is reflected by a decreased ability to detect changes.

Ramp shifts does not show this behaviour, and similar performance is seen for all onset points simulated. Small ramp shift gradients are also easier to detect if the total accumulated shift is used as criteria for comparison. This means that when it comes to performing hearing tests the more the merrier.

It was also found that presbycusis will eventually be detected, but that several years will go until this happens.

Finally the importance of outlier detection and counteractions were shown. Without high quality input data to the control charts the performance will be dubious. A possible detection rule and counteraction were presented and was shown to perform well for large outliers.

Real data from an offshore installation also show that it is possible to get individuals to perform frequent hearing measurements if the test is made readily available. The hearing monitoring regime presented can then be used as an early warning indicator and individual counteractions can be implemented if a hearing threshold shift commences. By doing this all persons will be better protected, even those who are more susceptible to loud sound.

Funding information: This work was supported by the Research Council of Norway [grant number 220667 NRC] and Statoil ASA.

References

R. W. Amin and R. A. Ethridge. A note on individual and moving range control charts. *J. Qual. Technol.*, 30: 70–74, 1998.

M. Best and D. Neuhauser. Walter A Shewhart, 1924, and the Hawthorne factory. *Qual. Saf. Health Care*, 15: 142–143, 2006.

S. Chakraborti. Run length distribution and percentiles: the shewhart chart with unknown parameters. *Qual. Eng.*, 19: 119–127, 2007.

M. Concha-Barrientos, D. Campbell-Lendrum, and K. Steenland. Occupational noise: assessing the burden of disease from work-related hearing impairment at national and local levels, 2004.

C. Crandell, T. L. Mills, and R. Gauthier. Knowledge, behaviors, and attitudes about hearing loss and hearing protection among racial/ethnically diverse young adults. *J. Natl. Med. Assoc.*, 96 (2): 176–186, Feb 2004.

Douglas M. Hawkins. *Identification of outliers*. Monographs on Statistics and Applied Probability. Springer Netherlands, 1st edition, 1980.

Douglas M. Hawkins. Self-starting cusum charts for location and scale. *J. Roy. Statist. Soc.*, 36 (4): 299–316, 1987.

Douglas M. Hawkins and David H. Olwell. *Cumulative Sum Charts and Charting for Quality Improvement*. Statistics for Engineering and Physical Science. Springer -Verlag New York, 1998.

J. A. Henry, C. L. Flick, A. Gilbert, R. M. Ellingson, and S. A. Fausti. Reliability of hearing thresholds: computer-automated testing with ER-4B canal phone earphones. *J. Rehabil. Res. Dev.*, 38 (5): 567–581, 2001.

ISO 1999. Acoustics – Estimation of noise-induced hearing loss, 2013.

ISO 7029. Acoustics - Statistical distribution of hearing thresholds related to age and gender, 2000.

Willis A. Jensen, Allison Jones-Farmer, Charles W. Champ, and William H. Woodall. Effects of parameter estimation on control chart properties: A literature review. *J. Qual. Technol.*, 38 (4): 349–364, 2006.

L. Jerlvall and S. Arlinger. A comparison of 2-dB and 5-dB step size in pure-tone audiometry. *Scand. Audiol.*, 15 (1): 51–56, 1986.

Hannah Keppler, Dhooge Ingeborg, Degeest Sofie, and Vinck Bart. The effects of a hearing education program on recreational noise exposure, attitudes and beliefs toward noise, hearing loss, and hearing protector devices in young adults. *Noise Health*, 17 (78): 253–262, 2015. doi: 10.4103/1463-1741.165028. URL http://www.noiseandhealth.org/article.asp?issn=1463-1741;year=2015;volume=17;issue=78;spage=253;epage=262;aulast=Keppler;t=6.

M. B. C. Khoo and S. H. Quah. Computing the percentage points of the run-length distributions of multivariate cusum control charts. *Qual. Eng.*, 15: 299–310, 2002.

Zhonghua Li, Jiujun Zhang, and Zhaojun Wang. Self-starting control chart for simultaneously monitoring process mean and variance. *Int. J. Prod. Res.*, 48 (15): 4537–4553, 2010. doi: 10.1080/00207540903051692. URL http://dx.doi.org/10.1080/00207540903051692.

Arve Lie, Marit Skogstad, Tore Tynes, Håkon A Johannessen, Karl-Christian Nordby, Ingrid Sivesind Mehlum, Line Arneberg, Bo Engdahl, and Kristian Tambs. Støy i arbeidslivet og helse, 2013.

MathWorks. MATLAB R2016a (64-bit), 2016. URL http://www.mathworks.com.

Douglas C. Montgomery. *Statistical Quality Control*. Hoboken, NJ: John Wiley & Sons, 7th edition, 2013.

NIOSH. Occupationally-induced hearing loss, 2010. URL http://www.cdc.gov/niosh/docs/2010-136/-pdfs/2010-136.pdf. Fact sheet. Cincinnati, Ohio:

National Institute for Occupational Safety and Health.

Charles Quesenberry. SPC Q charts for start-up processes and short or long runs. *J. Qual. Technol.*, 23: 213–224, 1991.

M. Sliwinska-Kowalska and M. Pawelczyk. Contribution of genetic factors to noise-induced hearing loss: A human studies review. *Mutat. Res.*, 752 (1): 61–65, 2013.

Laura Smith-Olinde, Nannette Nicholson, Courtney Chivers, Patricia Highley, and D. K. Williams. Test-retest reliability of in situ unaided thresholds in adults. *Am. J. Audiol.*, 15 (1): 75–80, 2006. URL http://search.proquest.com/docview/204420288?accountid=12870.

C. Spankovich and C. G. Le Prell. Healthy diets, healthy hearing: National Health and Nutrition Examination Survey, 1999-2002. *Int. J. Audiol.*, 52 (6): 369–376, Jun 2013.

A. Stuart, R. Stenstrom, C. Tompkins, and S. Vandenhoff. Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults. *Audiology*, 30 (2): 82–90, 1991.

S. Tak, R. R. Davis, and G. M. Calvert. Exposure to hazardous workplace noise and use of hearing protection devices among US workers NHANES, 1999–2004. *Am. J. Ind. Med.*, 52 (5): 358–371, May 2009.

The Norwegian Labour Inspection Authority. Veiledning om Hørselskontroll av støyeksponerte arbeidstakere, Dec 2013. URL http://www.arbeidstilsynet.no/binfil/download2.php?tid=77943.

T. V. Tronstad. The next step in preventing noise-induced hearing loss. *Int. J. Audiol.*, Forthcoming 2017.

Vinay, U. Peter Svensson, Olav Kvaløy, and Tone Berg. A comparison of test-retest variability and time efficiency of auditory thresholds measured with pure tone audiometry and new early warning test. *Appl. Acoust.*, 90: 153–159, 2015.

Joos Vos. Auditory tests for the early detection of noise-susceptible individuals – A literature study. Technical report, TNO Human Factors, 2005. URL https://www.sto.nato.int/publications/-STO%20Meeting%20Proceedings/RTO-MP-HFM-123/MP-HFM-123-05.pdf. Last accessed: 8 Sept 2016.

Table 1: CUSUM Q chart run lengths after a step shift has been introduced.

		Step shift size = σ		**Step shift size = 2σ**	
	Perc.	$k = 0.5$	$k = 0.75$	$k = 0.5$	$k = 0.75$
Step shift onset: after 50 observations	5 %	4	3	2	1
	25 %	8	7	3	2
	50 %	12	16	4	3
	75 %	22	>200	5	5
	95 %	133	>200	8	8
	5 %	4	3	2	1

Step shift onset: after 100 observations	25 %	7	6	3	2
	50 %	11	12	4	3
	75 %	17	24	5	4
	95 %	34	157	7	7
Step shift onset: after 150 observations	5 %	3	3	2	1
	25 %	7	6	3	2
	50 %	10	11	4	3
	75 %	16	20	5	4
	95 %	29	55	7	6
Step shift onset: after 200 observations	5 %	4	3	2	1
	25 %	7	6	3	2
	50 %	10	11	3	3
	75 %	15	19	4	4
	95 %	27	45	6	6

Table 2: CUSUM Q chart run lengths after a ramp shift has been introduced.

		Ramp shift rate = 0.1 dB/obs		**Ramp shift rate = 0.4 dB/obs**	
	Perc.	$k = 0.5$	$k = 0.75$	$k = 0.5$	$k = 0.75$
Ramp shift onset: after 50 observations	5 %	23	24	11	12
	25 %	39	43	15	16
	50 %	49	57	18	19
	75 %	59	69	21	22

	95 %	72	86	25	27
Ramp shift onset: after 100 observations	5 %	22	24	10	10
	25 %	35	40	14	15
	50 %	43	51	17	18
	75 %	51	62	19	21
	95 %	62	75	23	25
Ramp shift onset: after 150 observations	5 %	23	23	10	10
	25 %	35	39	14	15
	50 %	44	49	17	18
	75 %	52	59	20	21
	95 %	63	72	23	24
Ramp shift onset: after 200 observations	5 %	22	23	10	10
	25 %	35	38	14	15
	50 %	43	48	17	17
	75 %	51	57	20	20
	95 %	62	70	23	24

Table 3: Probability of detecting a given shift using the procedure described by the Norwegian Labour Inspection Authority (2013).

	Probability	
Threshold shift	$\sigma = 3$ dB	$\sigma = 5$ dB
10 dB	4.1 %	9.5 %
15 dB	63.6 %	47.8 %
20 dB	98.8 %	87.1 %

Table 4: Estimation of the hearing threshold shifts at the test frequencies giving an out-of-control signal. The "Warning signal" is the observation where the control charts give the signal.

Ear	Freq	Warning signal	Post HL-value	Pre HL-value	Difference
Right	6 kHz	34	23.7 dB	20.9 dB	2.8 dB
Right	3 kHz	37	21.6 dB	18.0 dB	3.6 dB
Left	6 kHz	39	11.1 dB	7.8 dB	3.3 dB