ISBN 978-82-326-2926-8 (printed ver.)
ISBN 978-82-326-2927-5 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Tron Vedul Tronstad

Doctoral thesis

Doctoral theses at NTNU, 2018:68

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Electronic Systems

NTNU

Doctoral theses at NTNU, 2018:68

Tron Vedul Tronstad

Hearing Protection Goes Digital

The Next Step in Preventing Noise-Induced Hearing Loss

NTNU
Norwegian University of
Science and Technology

Tron Vedul Tronstad

Hearing Protection Goes Digital

The Next Step in Preventing Noise-Induced Hearing Loss

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2018

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

ISBN 978-82-326-2926-8 (printed ver.)
ISBN 978-82-326-2927-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:68

Printed by NTNU Grafisk senter

Abstract

Hearing loss is one of the most common occupational health issues in the world. Despite much focus on noise abatement and hearing conservation programs, still many workers suffer from noise-induced hearing loss (NIHL).

One of the challenges with many hearing conservation programs is that they are based on performing tests only once every three years, which means that a hearing loss might go undetected for long periods. Additionally, there is quite a large uncertainty in the standard hearing threshold measurement method, which leads to that a threshold shift of 15 dB is required for a conclusion that NIHL might be present. Together this leads to a very reactive hearing test regime, where large hearing threshold shifts must be present before any counteractions are initiated.

This thesis will present a new hearing monitoring regime, using much more frequent hearing measurements and statistical process control, that can detect small (<5 dB) hearing threshold shifts. A rapid automated hearing threshold measurement, implemented in a smart communication earplug, facilitates the frequent measurements. This new regime could be used to initiate individual counteractions that aim at preventing further negative developments.

Preface

This dissertation is submitted in partial fulfilment of the requirements for the degree of philosophiae doctor (Ph.D.) at the Norwegian University of Science and Technology (NTNU). My supervisor has been Professor II Odd Kr. Ø. Pettersen at the Department of Electronics and Telecommunication at NTNU, and research director at SINTEF, and co-supervisor has been Professor U. Peter Svensson at the Department of Electronics and Telecommunication at NTNU.

The work has been performed at SINTEF in the Next Step project funded by the Research Council of Norway and Statoil ASA, project no. 220667 NRC. The project started January 2013 and lasted until December 2016.

During the project I was involved in six publications. A list of these are given in Table 1. Three of the publications (Tronstad 2015; Tronstad and Gelderblom 2016; Tronstad 2017) have been used in this thesis. Femke B. Gelderblom was co-author in one of the publications. She contributed to the writing of the introduction, discussion and conclusion of the paper, in addition to proofreading the entire paper. I did the field work, the statistical analysis, and also contributed to the writing of the entire paper. Footnotes are used to indicate where text from the papers have been used in this thesis.

Table 1: List of articles and reports published during the project.

T. V. Tronstad (2015). ‘Hearing Measurements During Two Norwegian Music Festivals’. In: *58th International Conference: Music Induced Hearing Disorders*. Aalborg, Denmark: Audio Engineering Society, pp. 1–4

T. V. Tronstad and F. B. Gelderblom (2016). ‘Sound exposure during outdoor music festivals’. In: *Noise Health* 18.83, pp. 220–228

T. Gjestland, O. Kvaløy, T. V. Tronstad and A. Melvær (2016). ‘Active Hearing Protection Device Provides Unique Possibilities for Hearing Research’. In: *SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility, 11-13 April*. Stavanger, Norway: Society of Petroleum Engineers, pp. 1–7

T. Gjestland, T. V. Tronstad and O. Kvaløy (2016). *Mulig metode for tidlig deteksjon av hørselstap. Eksempler på bruk av prosesskontrollmetoder på hørselsdata fra SLASH (in Norwegian, Possible method for early detection of hearing loss. Examples on use of process control on hearing data from SLASH)*. A27862, ISBN 9788214061369. Trondheim, Norway: SINTEF

T. Gjestland and T. V. Tronstad (2016). ‘The efficacy of sound regulations on the listening levels of pop concerts’. In: *J. Occup. Environ. Hyg.* 14, pp. 17–22

T. V. Tronstad (2017). ‘Statistical tool to detect small hearing threshold shifts’. In: *Int. J. Audiol.* 56.8, pp. 596–606

Acknowledgement

This thesis is dedicated to Asle Melvær who sadly passed away June 15, 2016. Asle worked with noise issues in Statoil ASA, and in the petroleum industry in general, for several decades. In addition to always being in a cheerful mood, his professionalism reflected everything he did. He was a key person in the project and his main idea about early detection of a hearing damage is one of the fundamental pillars in the work we have continued on. Hopefully this will reduce the number of people with noise-induced hearing loss in the years to come.

I would also like to thank my supervisors Prof. II Odd Kr. Ø. Pettersen and Prof. U. Peter Svensson for their support and guidance during the entire project.

Additionally I would like to thank the members of the steering committee in the project; Solveig Engen and Kristin Brørs from Statoil ASA, Trym Holter and Viggo Henriksen from Honeywell Safety Products, Prof. Magne Bråtveit from the University of Bergen, and Truls Gjestland and Olav Kvaløy from SINTEF. Prof. II Odd Kr. Ø. Pettersen and Prof. U. Peter Svensson have also been members in the steering committee.

Last, but not least I would like to thank my family for their everlasting presence. My wife, Guro, and my three children, Vidar, Åsne and Børge, have an unique ability to make me re-focus and take a break during off-hours.

Contents

Abbreviations

Acronyms

ABR Auditory Brainstem Response

ARHL Age Related Hearing Loss

ATS Asymptotic Threshold Shift

DPOAE Distortion Product Otoacoustic Emission

EEH Equal Energy Hypothesis

HL Hearing Level

HPD Hearing Protection Device

NIHL Noise-induced Hearing Loss

NIOSH National Institute For Occupational Safety And Health

OAE Otoacoustic Emission

OHSP Occupational Health Service Provider

PTA Pure Tone Audiometry

PTS Permanent Threshold Shift

QP Quietpro® QP100Ex

SDT Speech Detection Threshold

SIN Speech In Noise

SPC Statistical Process Control

SPL Sound Pressure Level

SRT Speech Recognition Threshold

TEOAE Transient-evoked Otoacoustic Emission

TTS Temporary Threshold Shift

WHO World Health Organization

Chapter 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). The World Health Organization (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 legislation regulating occupational noise, there are many 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 health issues from manufacturing industries refers to a hearing loss in the US (NIOSH 2010).

WHO (2015) also estimated that 360 million people worldwide had *disabling hearing loss*[a] in 2012. This number included all causes of hearing loss, not only noise. Other causes include ear infections and infectious diseases, ototoxic medicine, injury to the head and wax/foreign bodies in the ear canal, in addition of congenital causes such as maternal rubella, syphilis or inappropriate use

---

[a] Disabling hearing loss refers to hearing loss greater than 40 dB in the better hearing ear in adults and a hearing loss greater than 30 dB in the better hearing ear in children, measured as an average for frequencies 0.5 kHz, 1 kHz, 2 kHz, and 4 kHz (WHO 2015).

of particular drugs during pregnancy, low birth weight, and lack of oxygen at the time of birth. WHO also state that half of all cases of hearing loss are avoidable through primary prevention[a], and reduced sound exposure, and use of hearing protection devices are two of these measures. In addition WHO points out that early detection and interventions are crucial to minimize the impact of hearing loss on a child's development.

The large number of people at risk is alarming, especially when taking into account the individual and social cost associated with hearing impairments.

In 2006 Shield did a review of literature on the social and economic costs of hearing impairments and applied this on the European population (Shield 2006). She estimated that more than 71.5 million adults (15.9 %) in Europe had more than 25 dB average hearing threshold shift over the frequencies 0.5, 1, 2, and 4 kHz, in the better ear. From this an estimate of the total economic cost associated with hearing impairment was made. Using a 'quality of life' approach, including overall effects such as psychosocial effects and loss of productivity/unemployment, a total cost of EUR 284 billion/year was estimated for Europe. Shield also stated that out of these, EUR 213 billion/year are associated with unaided hearing loss. It was also claimed that hearing impairments cost the UK alone GBP 18 billion each year, due to loss in productivity or unemployment, where GBP 13.5 billion are due to unaided hearing loss. This shows that the impact of not using hearing aids when in need of one is very large.

Kochkin (2012) has estimated that only one out of four persons with hearing impairment have a hearing aid. There are several reasons for this, but one of the most common is that the person is not aware of the hearing impairment. Kochkin states that among non-adopters, i.e. persons with a hearing impairment not using hearing aids, almost half (46.3 %) had never tested their hearing, or had not been tested during the last ten years.

Presbycusis, or age related hearing loss (ARHL), is inevitable when getting old. Even if the size of the ARHL differs between individuals, it is accepted that the hearing gets worse as we grow older (Robinson and Sutton 1979). Animal studies, measuring the hearing threshold for non-exposed individuals, further strengthens the acceptance of an inherent worsening of the hearing as function of age (Fetoni et al. 2011). One should, however, notice that not only

---

[a] Primary prevention aims to prevent disease or injury before it ever occurs. See Table 4.2 for more details.

the hearing threshold increases with age, but also the variation. This is possible to see in the international standard, ISO 7029, where hearing thresholds are described as function of age (ISO 7029 2000). Figure 1.1 shows the 10 %, 50 %, and 90 % percentiles of the hearing threshold shift at 6 kHz for the ages from 18 – 70 years according to the ISO standard. Some criticism have been given to the standard for not being accurate, but the increased variability is not questioned (Stenklev and Laukli 2004). It is unknown how large part of the ARHL that consists of accumulated NIHL, and what proportion that is associated with age alone, but different sound exposure and NIHL might explain some of this variability.

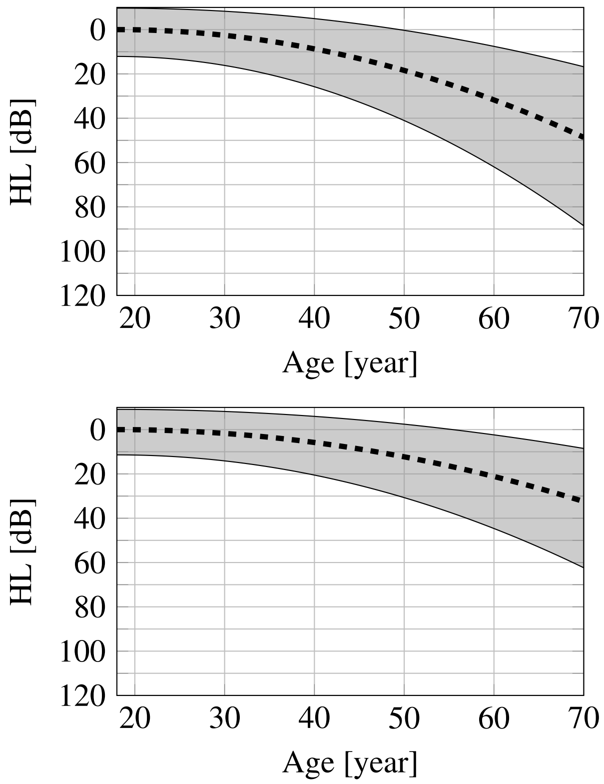

Figure 1.1: Age-related hearing loss (ARHL) for 6 kHz for males (upper) and females (lower) according to ISO 7029 (2000). The dashed lines show the median value, while the filled areas are made of the 10 % and 90 % percentiles.

Since hearing threshold shifts develop over time it is not easy to notice that the hearing is getting worse. This is also the reason why many elderly people often claim to have normal hearing; they do not notice the change in hearing since it has developed over time. The problem can be illustrated by quotes that could have been said by many persons with elevated hearing threshold shift:

> My hearing is perfectly normal, but…
>
> > …my wife/husband is slurring/mumbling more than before.
> >
> > …tv programs are so noisy these days.
> >
> > …I cannot go out on restaurants any more because it is impossible to understand what is being said.

This unawareness, combined with a lack of knowledge about the consequences of an elevated threshold, is unfortunate for the protection against noise-induced hearing loss. Lack of knowledge also leads to many myths associated with hearing, hearing damage and hearing protection. Some examples are (adapted from G. W. Hughson et al. (2002)):

- I'm already deaf so there is no point in wearing hearing protection.
- Using earplugs will cause ear infections.
- Short periods of noise exposure are not harmful.
- Earmuffs make me appear less attractive, silly, less macho.
- I'm resigned to hearing loss by having to work in a noisy job.
- The manager doesn't wear the hearing protection, why should I?

The development of NIHL is described by the international standard, ISO 1999 (2013). Figure 1.2 shows the time development for the 50 % percentile for four sound exposure levels. The figure clearly show that the first 10 years are most important for the development of NIHL. This also means that these are the most important years when it comes to identifying and acting upon a negative development of the hearing.

Recently the National Academies of Sciences, Engineering, and Medicine (2016) published a report on hearing health care for adults, highlighting some important actions that can improve the current situation. One of the things they

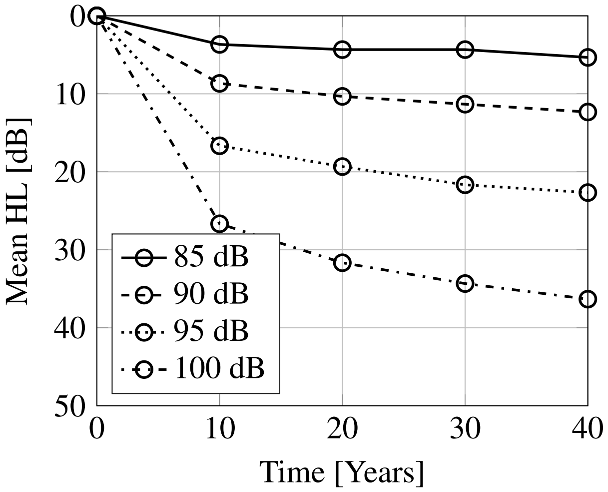

Figure 1.2: Time development of NIHL for four different sound exposures. The hearing level (HL) is found by taking the mean of 3, 4, and 6 kHz, and are based on the 50 % percentile.

mention is to improve the population-based information on hearing loss and hearing health care.

As the WHO pointed out, early detection and interventions of hearing loss are crucial for a child's development. Similarly, early detection and interventions of acquired noise-induced hearing loss is crucial to limit the enormous individual and social cost associated with an elevated hearing threshold. Currently in Norway, the detection of such shift is based on a hearing monitoring regime where the worker is tested with pure tone audiometry (see Ch. 2 for more details) ones every three year. Due to a large variation in hearing test results, a threshold shift of 15 dB, or more, must be seen to confidently say that a hearing threshold shift has occurred. This is a very reactive regime, where a large damage must be present before any intervention can be put into action.

More frequent hearing measurements, combined with a statistical process control regime, can improve this hearing monitoring and detect much smaller (<5 dB) threshold shifts. This way earlier interventions, on an individual basis, are feasible and it might be possible to prevent any further negative development.

1.1 Research Problem

The vision of this project is to prevent hearing damage. This will be achieved by implementing a realizable hearing monitoring regime using frequent hearing measurements. Additionally a secondary goal is to make individuals more aware of their hearing such that they can take necessary actions when exposed to loud sound. Such education can also contribute to the knowledge about hearing health care.

In the hypotheses below the *current hearing test regime* means the one used in Norway today (see Sec. 3.1.1 for details). The *proposed hearing test regime* means the method described in detail in this thesis.

A technology research approach has been used as underlying method when developing the hearing monitoring system presented in this thesis. The main hypothesis in such research is that the new ‘artefact’ is better than the old one. Because this hypothesis is hard to falsify, it is necessary to use sub-hypotheses to evaluate the main hypothesis.

The main hypothesis for the hearing monitoring system is therefore:

H1: The proposed hearing monitoring system is better than the current hearing test regime.

To be able to evaluate this, two sub-hypotheses are expressed:

$H1_A$: The proposed hearing monitoring system can detect smaller hearing threshold shifts than the current regime.

$H1_B$: The proposed hearing monitoring system can detect threshold shifts faster than the current regime.

In addition the practicalities around the system is considered to assess how to implement the hearing monitoring in real life, and experiences are drawn from the project where the system has been used.

1.2 Thesis Outline

Chapter 1 gives an introduction to NIHL and the importance of detecting hearing threshold shifts and acting upon them. It also presents the research problem and gives an outline of the thesis.

In Chapter 2 methods to measure hearing ability is presented. Both subjective and objective tests are discussed, and the automated hearing test used in this work is presented. This hearing test is based on a version developed by Vinay, Henriksen et al. (2014) in a previous project related to the work in this thesis. Challenges experienced with the initial version are presented together with some improvements made. A discussion around the use of temporary threshold shift (TTS) data will also be presented. The initial plan was to use knowledge about the sound exposure to estimate the TTS the exposed person might have. A study performed on music festival participants is presented and show that no simple estimate of the hearing threshold shifts can be performed.

Chapter 3 will first give a brief description of the hearing monitoring used today, then introduce a new method capable of detecting small (< 5 dB) hearing threshold shift. This can act as an early warning indicator that can be used to prevent further negative development. The chapter includes a technical description of the statistical process control used in the hearing monitoring.

Next, Chapter 4 gives a more superior view of the implementation of the proposed hearing monitoring. Several aspects are discussed for different phases of the hearing monitoring. This chapter is supposed to work as a guideline/check list if one wants to start using the hearing monitoring regime.

Then Chapter 5 shows examples from real-world data. 18 workers at an offshore installation have measured their hearing using a hearing protection device (HPD) with an embedded hearing test. In addition the sound exposures have been measured using the same HPD. Several aspects around the proposed hearing monitoring system will be elucidated.

Finally, Chapter 6 and Chapter 7 will summarize, give concluding remarks and present possible future work.

A flow chart of the hearing monitoring scheme can be seen in Figure 1.3. The red dashed boxes show where the different parts are presented in this thesis.

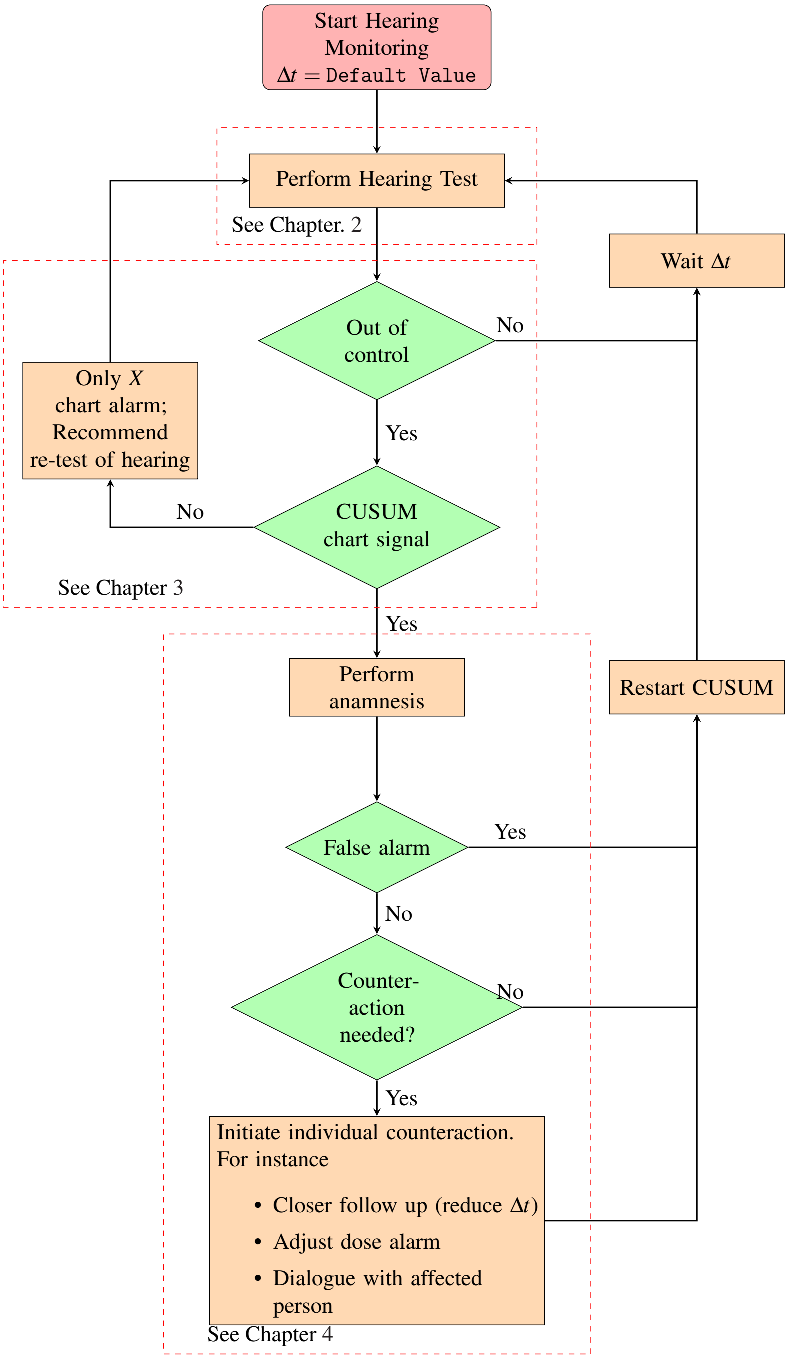

Figure 1.3: Flow chart showing the hearing monitoring scheme presented in this thesis, together with the chapters where more information about the different parts can be found.

Chapter 2

Measuring Hearing Function

To test the function of the hearing several tests exist. These can be divided into two main categories:

Subjective tests	Requires a subjective response from the test subject, e.g. performing an action or saying something.
Objective tests	No subjective response is required, but objectively measurable responses are collected.

A brief overview of such tests is given below. Furthermore, automated implementations of subjective tests are described in Sec. 2.3 and 2.4. This includes one implementation, ‘New Early Warning Test (NEWT)’, which has been developed for the Next Step research project. The last section of this chapter gives a discussion of temporary threshold shifts (TTS), which is often the result of exposure to loud sound. This last section presents results from studies of noise exposure and hearing threshold measurements at music festivals.

2.1 Subjective Tests

The subjective tests can again be divided into two main categories; threshold and supra-threshold tests. The threshold tests find the faintest sounds a person can hear, while the supra-threshold tests measures the function of the hearing, including the cognitive parts.

2.1.1 Pure Tone Audiometry

The most common, and the one considered to be the ‘gold standard’, is the pure tone air conduction audiometric test, often called pure tone audiometry (PTA).

This test is often performed using headphones on the test subject and playing test tones on one side at the time (monaural), but it can also be performed binaurally without headphones in a sound field. The first is often used for diagnostic purposes while the latter is often used to fit and assess the function of hearing aids, since the test can be performed with and without the hearing aid fitted.

Air conductive audiometry does not differentiate between sensorineural and conductive hearing losses. Conductive losses can be detected using bone conduction pure tone audiometry where a tuning fork, or a vibrating device, is placed on the skull. The vibrations will then go through the skull and directly into the cochlea without going through the outer and middle ear. If these sounds are heard normally it can be stated that the sensory cells in the inner ear are normal, and that the cause of a hearing threshold shift lies elsewhere.

Both monaural and binaural pure tone audiometry are standardized in international standards, ISO 8253-1 (2010) and ISO 8253-2 (2009), respectively.

Pros and Cons

Pure tone audiometry gives rather consistent answers, and the results can be used to determine if a hearing damage exist. It is, for instance, the test used to establish if a damage has occurred in occupational settings.

The test is, however, subject to criticism as well, especially the last few years. Kujawa and Liberman (2009) showed that neural damage could be present in the sensory cells of the cochlea in animals even if the auditory threshold had returned to normal. The complete recovery of a hearing threshold shift is therefore not a ‘proof’ that the hearing is undamaged.

2.1.2 Speech Tests

It is also possible to measure hearing function with speech material. The most common tests used for speech material is (adapted from The American Speech-Language-Hearing Association (1988)):

Speech recognition threshold (SRT): The level needed for recognition of 50 % of the speech material. Recognized means that the test person must be able to reproduce the word(s), either orally or by selecting the correct word(s) from a list.

Speech detection threshold (SDT): The level where the test subject can discern the presence of speech material 50 % of the time. The test subject does not have to recognize the word(s), just that there is something spoken.

Both these tests can be performed monaural (with headphones) or binaural (with headphones, in sound field or bone vibrator), and with or without additional background noise. If noise is added the tests are often referred to as ‘speech in noise’ (SIN) tests.

Also the speech tests are standardized by an international standard (ISO 8253-3 2012).

Pros and Cons

One of the pros of performing speech tests is that they are closer to the sound exposure experienced in the real world, and especially speech in noise is representative for the problem many experience when their hearing is getting worse. One of the most common complaints, often experienced before any other hearing challenge, is the problem of recognizing speech in noise (e.g. in restaurants or cocktail parties).

The method is, however, also criticized for not only measuring the hearing function, but also the cognitive function (Beck and Repovsch 2013). It has even been found that structural differences in the brain (gray matter volume) can predict the word recognition score (Harris et al. 2009). This means that it can be difficult to differentiate between hearing loss and cognitive decline.

Nonetheless, since there are little correlation between pure tone audiometry results and the speech-in-noise ability, one must perform SIN tests to assess how a person understands speech in noise (Beck and Repovsch 2013).

2.2 Objective tests

There also exist objective tests that can measure the function of the hearing. These are especially suitable for subjects that cannot give a response (e.g. infants).

2.2.1 Otoacoustic emission

The most common objective method is otoacoustic emissions (OAEs). OAEs are sounds of cochlear origin that can be picked up with a sensitive microphone placed in the ear canal.

The OAE tests can be divided into two main categories; transient-evoked otoacoustic emissions (TEOAEs) and distortion-product otoacoustic emissions

(DPOAEs). The first is evoked by sending clicks into the ear triggering a broad band of frequencies, while the latter is evoked by the nonlinear intermodulation between two stimulus tones. Even if the click triggers a broad band of frequencies, it is possible to get frequency information from the response by splitting the results into frequency bands after recording. The DPOAE is more frequency specific since the intermodulation occurs at unique frequencies (e.g. $2f_1 - f_2$ where f_1 and f_2 are the frequencies of the two different tones).

Pros and Cons

As mentioned, the best thing about the objective tests is that they do not need any subjective response from the test subject. This reduces the uncertainty in the measurement and also makes automated testing easier.

The problem is that the results are harder to interpret. Even if the presence of a strong OAEs is a good indicator of a functioning ear, weak or absent OAE does not have to indicate a damage (Kemp 2002). It is therefore difficult to use this test to classify the magnitude of a damage.

2.2.2 Electrocochleography

Electrocochleography is a method for recording the electrical potentials of the cochlea. Stimulus-related potentials are measured, either for the whole nerve or compound action potentials for the auditory nerve (Ferraro 2010). These electrical signals can be measured different places, both inside the tympanic membrane (transtympanic), in the ear canal (extratympanic), or with electrodes attached to the skull. The first two methods are especially suited to measure cochlear potentials, e.g. cochlear microphonic, while the latter is better for auditory brainstem response measurements.

Auditory brainstem responses (ABRs), or auditory evoked potentials, are electrical field potentials generated by stimulation of the auditory pathways (Berger and Blum 2007). These are recorded with scalp electrodes and display the time course of the electrical signal occurring in the ear and brain following a sound stimuli. The ABR show a predictable time pattern with seven characteristic waves that can be used diagnostically.

Berger and Blum (2007) points out that clicks are often used as sound stimuli to evoke the potential, and masking noise is presented on the other ear to avoid co-activation. Other stimuli, such as chirps, are however also possible to use (Dau et al. 2000). Because of the high background activity in evoked potentials, it is also essential to perform repeated measurements to maintain a good signal-to-noise ratio. Around 1000 – 4000 stimuli are given, with approx. 8 – 10 stimuli per second (Berger and Blum 2007).

Pros and Cons

Electric signals going from the cochlea can give a lot of information about the hearing, but the interpretation of the results need careful attention. It is also stated that not all waves in an auditory brainstem response are present in normal hearing individuals, hence one must interpret the results with care. Often laboratories establish their own normative database to base decisions on, but published tables of normal absolute and interpeak latencies are also available (Berger and Blum 2007).

2.3 Automated Hearing Tests

Automated hearing tests have been suggested and used for several decades. Margolis and Morgan (2008) wrote in a popular science article that already in 1963 Jerger expressed in the book *Modern Developments in Audiology*, that

> the number of audiometric examinations made today has grown to such a magnitude that it is only natural that some of the techniques of measurement should become automated. (p. 30)

and continued with

> . . . it appears only natural that those features of audiometry which can be automated will be, and the audiologist will find himself fully occupied with the task of analysing and interpreting the data. The routine work can be done by the machine. (p. 31)

Margolis and Morgan (2008) also point out that the first automated hearing test was implemented by George von Békésy already in 1947. This was also the most used automated hearing test used for research purposes until it was replaced by more accurate evoked potential tests in the 1970s.

It was, however, not until 2001 that the first automated instrument was commercially available for diagnostic purposes. Until then most of the automated tests were used for industrial purposes to screen and/or monitor the hearing of the employees, or in research.

The most common procedure in pure tone audiometry is the Hughson-Westlake (HW) procedure, described already in 1944 by Walter Hughson and Harold D.

Westlake (W. Hughson and Westlake 1944). The HW procedure describes how the sound level should be adjusted from one stimuli to another according to how the test subject responds to the stimulus. This procedure, or modified versions of it, is currently the most used method to perform pure tone audiometry, both manually and automatically (Song et al. 2015), They are, however, rather ineffective and therefore more refined tests have been suggested. Three general methods for determining psychophysical responses, such as subjective hearing results, have been described (adapted from Leek (2001)):

Parameter estimation by sequential testing (PEST):
PEST is characterized by an algorithm for threshold searching that changes both step sizes and direction (i.e., increasing and decreasing level) across a set of trials. The PEST algorithm is designed to place trials at the most efficient locations along the stimulus axis in order to increase measurement precision while minimizing the number of trials required to estimate a threshold.

Maximum likelihood estimation (MLE):
In this procedure sets of stimulus-response trials are fit with some type of ogival function and subsequent trial placement and threshold estimation is taken from those fitted functions. A new psychometric function is generated after each trial or set of trials, and subsequent trials are placed at a targeted performance level on the most up-to-date function. A maximum-likelihood fitting algorithm is typically used with this type of procedure.

Staircase procedures:
These methods generally use the previous one or more responses within an adaptive track to select the next trial placement, then provide a threshold estimate in a variety of ways, most commonly by averaging the levels at the direction reversals in the adaptive track. No assumptions about the underlying psychometric function is needed, which the other procedures need.

Especially the two first are effective methods, and the MLE utilize all the available information to find the most probable threshold shift.

Another method is the Ψ method. Kontsevich and Tyler (1999) proposed this Bayesian adaptive method where both the threshold and slope of an unknown psychometric function can be found. Other approaches have also been proposed, for instance a procedure using machine learning to find the most informative sound stimuli (Song et al. 2015).

The hearing test used in the studies presented in this thesis is a MLE method. This will be described next.

2.4 New Early Warning Test (NEWT)

An automated hearing test has been used in the work presented in this thesis. The underlying method is the *New Early Warning Test* (NEWT) presented by Vinay, Henriksen et al. (2014), but during the project some improvements have been made. An adaptive maximum likelihood estimation (MLE) psychophysical procedure is used to find the most probable HL for the test subject. This procedure was first proposed by Green (1990).

Test speed has always been very important for the NEWT method, and the current implementation can find the HL estimate, shown to be at least as accurate as normal audiometry (Vinay, Svensson et al. 2015), after six sound stimuli on each frequency. This means that testing three frequencies on both ears takes approx. 2 – 2.5 min. The results in the current realization also have a higher resolution than normal 5 dB bins. The exact resolution differs slightly within the dynamic range, but lies around 2 dB.

2.4.1 Challenges

One of the problems with the first version of NEWT was the limited dynamic range (30 dB) of the system. Since the difference in hearing level between ears and individuals can be much larger than 30 dB, prior information was needed to first place this 'window' close to the real hearing level. This meant that individual seed values for each test frequency were needed and that these were estimated based on previous measurements. The median of the last three results was chosen as the seed value.

Correct placement was, in the project, defined to be from -5 dB to 25 dB around the seed value. This placement meant that, provided that the seed value was estimated close to the real HL, a hearing threshold shift of 25 dB could be measured in one test. Larger shifts would require more than one test to be found.

If a true shift of 40 dB was to be measured this would require three measurements. The first measurement would give a 25 dB HL, but the seed value would not be adjusted. The second measurement would also give a 25 dB HL, and the seed value would be adjusted to 25 dB. Finally the third hearing test would give the correct HL. This sequence does not take into account the test-retest variability of the test subject, hence more than three measurements could be needed.

Since this issue was expected to be only a start-up challenge, when the person's seed value was not yet tuned into the correct value, this was not considered a problem in the initial version of the test.

Another challenge, which made the seed problem more severe, was the lack of redundant information. Since test speed was, and still is, considered most important, no redundant information was prioritized. A consequence of this was that the first version of the test was sensitive to lapse of attention and background noise. This was especially critical for the first test stimuli on each frequency. If the test subject did not answer on this stimuli, the method immediately estimated the HL to be worse than 25 dB. Since lapse of attention is more likely to happen at the first stimuli, when the test subject might not be aware that the test has started, this was a recurrent problem in the test results from the early stage of the project. Combined with the seed value estimation mentioned above, this had a detrimental effect on the hearing monitoring. Since the seed value could jump 25 dB if the test subject failed to answer correctly on two consecutive tests, large unwanted shifts occurred. Additionally this issue became even more problematic since the seed value could only jump in 5 dB steps in the other direction, since the window was placed from -5 dB to 25 dB around the seed value. An unwanted shift of 25 dB would therefore need ten tests(!) to return to normal.

2.4.2 Improvements

To cope with these challenges a revised version of the NEWT was developed. This method was also presented at the Hearing Across the Lifespan (HEAL) conference, in Cernobbio, Italy, June 2 – 4, 2016.

First of all the dynamic range was doubled from 30 dB to 60 dB. This was done by defining the start stimuli, presented at 25 dB to be in the middle of the dynamic range instead of at the bottom as it was in the first version. This simple change immediately doubled the dynamic range, without changing anything else in the method. Monte Carlo simulations, not shown in this thesis, verified that the accuracy of the method was unaffected. Increasing the dynamic range to 60 dB meant that the seed value became less important. By placing the 'window' such that it goes from -5 dB HL to 55 dB HL most of the population will be covered and more than 30 dB hearing threshold shifts can be measured for persons with normal hearing ($\text{HL} < 25$ dB).

Second, NEWT were made more robust against lapse of attention at the first stimuli. Since this error was especially detrimental for the test results, solving this problem was important. Instead of changing the NEWT method itself, which would require a new verification, the solution was to re-start the test if the test-subject did not respond to the first stimuli. The only negative consequence was an increased test duration for the individuals with HL > 25 dB, since they always would need to re-start the test. The increased time was, however, only around 5 s per frequency.

In Figure 2.1 a flow chart of the NEWT process can be seen. The part within the green dashed box illustrates how the more robust method was implemented. If the components inside the box is removed, the flow chart represents the initial version of the test.

2.4.3 Implementations

Three different implementations of the hearing test have been used in this project, all based on the same methodology. The intention has all the time been to embed the hearing test into the hearing protection device[a] (HPD), but this was not done in the first version of the HPD used in the project. Instead the QP-users had to go to a dedicated computer where they could connect their headset to a hearing test device. This also meant that they had to go to the computer and stand beside it while performing the hearing test. Figure 2.2 shows a person in front of this computer.

Later in the project the hearing test was embedded into the QP, hence the workers could perform the hearing test wherever and whenever they would like, with some limitations associated to background noise. Even if the HPD has a verified attenuation, i.e. it tests if the earplug has been inserted correctly, background noise levels above 40 dB to 50 dB can affect the hearing test results. The NEWT version implemented into the QP use the improvements presented, and gives calibrated HL as output.

Additionally another computer based version of the hearing test has been used for the testing performed at the music festivals. A computer version of NEWT used Sennheiser HDA 200 Audiometric closed-back headphones connected to an external sound card and were implemented in Matlab (MathWorks 2014). This version was not calibrated. For more details see Sec. 2.5.3.

---

[a] The hearing protection device used in this project was the Quietpro® QP100Ex. This HPD is produced and sold by Honeywell. In this thesis the device will be called QP.

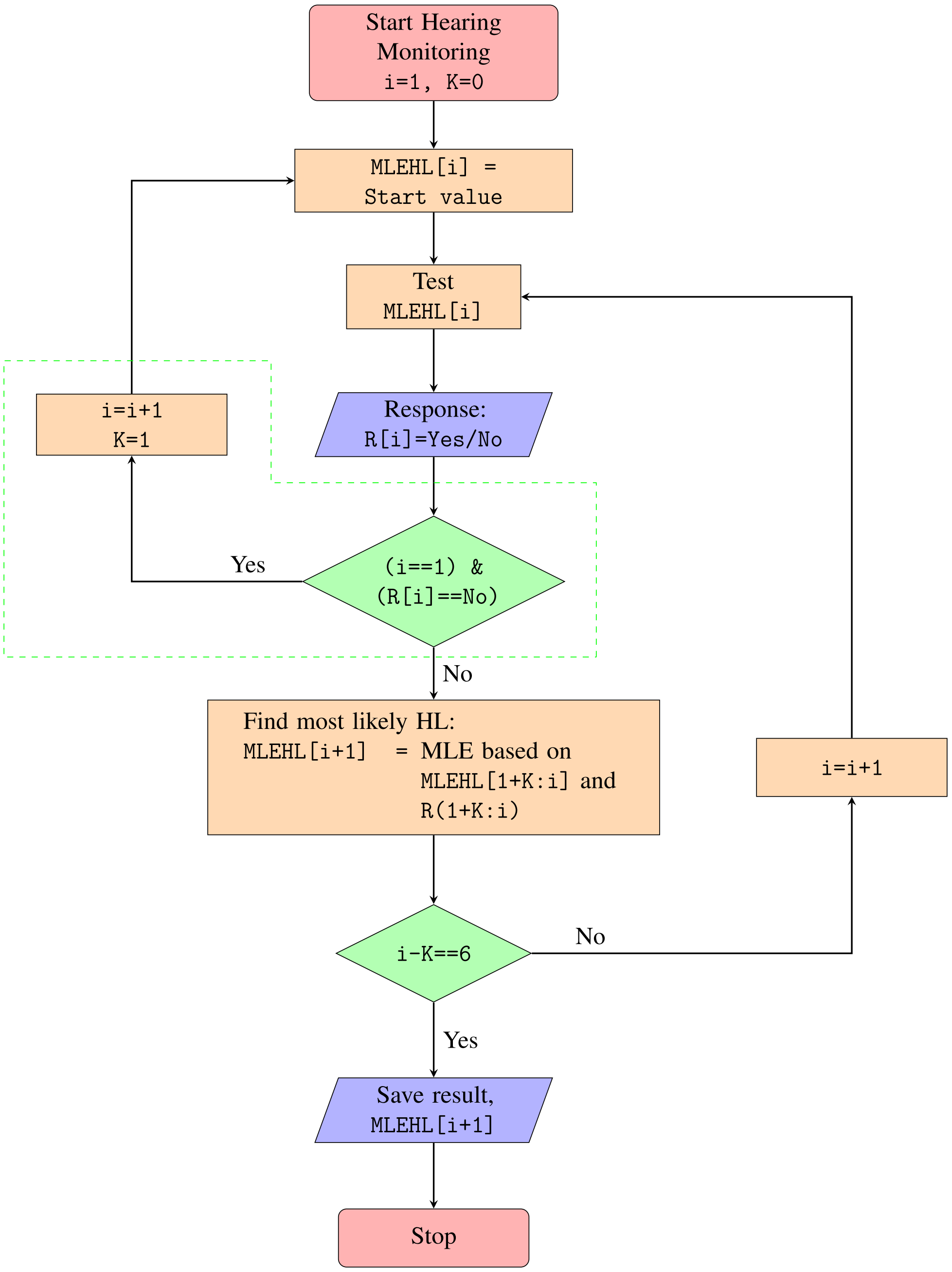

Figure 2.1: Flow chart of the NEWT hearing test. `MLEHL` stands for the most likely estimation of the hearing level.

Figure 2.2: Photo of a QP user in front of the computer where the exposure data could be uploaded to the database, and the hearing test could be performed before it was embedded in the HPD. Photo: Kari Aasbø/Statoil ASA.

2.5 Temporary Threshold Shifts

Temporary threshold shift (TTS) have been researched for several decades, but still many of the underlying mechanisms are unknown. There are also conflicting evidence on the link between TTS and PTS. Melnick (1991) concludes that the asymptotic threshold shift (ATS), which is a plateau reached after 8 h to 10 h of sound exposure, can be an indicator of the maximum magnitude of a PTS for the same sound, if a person is exposed to it for several years. He does, however, also state that the relationship is more complicated for intermittent sound exposure, dependent on level, frequency, and duty cycle of the sound. This is a challenge for real sound exposures, that often are intermittent, with changing level, frequency, and duty cycle (see Figure 2.3 for an example of a real music festival exposure).

More recently it has also been stated that TTS might be more dangerous than previously expected. Kujawa and Liberman (2009) introduced the term ‘hidden hearing loss’ after they identified permanent damage on the nerve synapses (cochlear synaptopathy) even when the hearing threshold returned to normal after loud sound exposure. One should be aware that this study was performed on animals and one does still not have any studies indicating the same damaging mechanism in humans.

The problem for the hearing monitoring is that TTS can affect the process control and that such temporary changes can give an ‘out of control’ signal. Even if it is not obvious that such detection is negative since a TTS can be associated with loud sound exposure, the risk is that the system will ‘cry wolf’, and the alarms might not be trusted. TTS can be caused by other factors than noise, for instance a stuffy nose or otitis media, and such temporary conditions is not necessarily associated with an increased risk of hearing impairment.

One possibility would be to use a known relationship between a sound exposure and hearing threshold shift to estimate the TTS in a hearing measurement. Ward et al. (1958) presented a mathematical model for the onset and offset of TTS as function of sound exposure level several decades ago. Later others have also proposed such models (e.g. Keeler 1968; Mills et al. 1979; Patuzzi 1998; Ordoñez and Hammershøi 2011). Simple or multi-exponential functions are used to determine the size of the TTS, but the problem is that intermittent noise with varying SPL complicates the situation.

To see if the hearing test used in the project could detect TTS, and to see if a simple relationship between sound exposure and hearing threshold shift could be found, a study on the hearing of music festival participants was performed. Eight persons wore sound level meters and tested their hearing before and after each day with concert exposure.

This work has been published in the Noise & Health (Tronstad and Gelderblom 2016) and in the conference proceedings for the 58th Audio Engineering Society conference in June 2015, named ‘58th International Conference: Music Induced Hearing Disorders’ (Tronstad 2015).

In the following subsections a brief summary of the most important findings are given together with the main conclusion.

2.5.1 Music Festival Study: Introduction [a]

Concert attendees often complain about ringing sounds and that everything sounds ‘muffled’ after an event (Bogoch et al. 2005). These phenomena generally disappear within the next day or days. The muffling sensation is

---

[a] The text in the introduction is taken from the paper presenter in Noise & Health in 2016 (Tronstad and Gelderblom 2016).

more formally known as a temporary threshold shift. If the hearing threshold does not restore completely, it becomes a permanent threshold shift, also known as noise-induced hearing loss (NIHL). The ringing is called tinnitus, and is a sound experienced by the person in absence of any external stimulus. Previous research shows that loud music can cause TTS, PTS/NIHL, and tinnitus (K. Kähäri et al. 2004; Opperman et al. 2006), although conflicting evidence exists (Zhao et al. 2010).

Most of the research on the impact of loud music on hearing focusses on concerts, discotheques, and portable media players (Vogel, Verschuure et al. 2010; European Commission 2008; Vogel, Brug et al. 2007; Maassen et al. 2001). Music festivals have lately become increasingly popular throughout the world. Several hundred large and small festivals all over the country fill the air with loud music each year, mostly during the summer months from May to October. Several hundreds of thousand people attend these events, just in Norway. These festivals have multiple stages and often last for several days. With only one published study on the sound exposure of a festival's audience (Mercier et al. 2003), knowledge of this topic is severely limited. Lack of information on the unique dose received by festivalgoers inhibits evaluation of the impact of these events on hearing. If the sound level at the concerts during a music festival is similar to the sound levels used at single concerts, the risk of getting a hearing damage is even higher during music festivals. This is due to the fact that most participants attend more than one concert during such festivals.

Current regulations for festivals (if existent) are, in many European countries, often based on the international standard ISO 1999 (2013) and/or the European Directive 2003/10/EC – noise (European Commission 2003) that regulate occupational noise exposure. The ISO standard states that an employee can be exposed to 85 dBA for eight hours, each day, for his/her entire work career, without increasing the risk of suffering from noise-induced hearing damage. The European Directive limits the exposure of employees to 87 dBA per 8 h workday, and sets lower ('hearing protection must be made available') and upper ('hearing protection must be worn') action limits at 80 dBA and 85 dBA, respectively. The Norwegian occupational noise legislation (The Norwegian Labour Inspection Authority 2015) is based on ISO 1999, but additionally sets an action limit of 80 dBA for an eight hour work day. This means the employer must take action if the noise level at work exceeds this limit.

ISO 1999 also relies on the equal energy hypothesis (EEH), which assumes that an equal amount of sound energy always has the same damaging potential. A consequence of this hypothesis is that one can change the noise exposure's distribution and/or increase the level while reducing the exposure time, or vice versa, without affecting the damaging potential, as long as the energy remains constant.

The Norwegian Directorate of Health has made a guideline for local authorities in Norway to help them set limits for concerts and festivals (The Norwegian Directorate of Health 2011). To prevent hearing damage among the visitors, the guideline sets critical limits of $L_{p,\mathrm{A},30\,\mathrm{min}} = 99\,\mathrm{dB}$ and $L_{\mathrm{C,peak}} = 130\,\mathrm{dB}$. $L_{p,\mathrm{A},30\,\mathrm{min}}$ is for the A-weighted equivalent level over a 30 minute period, and the limit applies to the loudest 30 minutes of the concert. $L_{\mathrm{C,peak}}$ is the C-weighted peak-level, and this limit is equal to the peak limit used in the Norwegian occupational noise legislation. These limits are derived from occupational noise exposure regulations using the EEH (The Norwegian Directorate of Health 2011).

The Worlds Health Organization (WHO) also gives recommendations regarding the sound level exposure at ceremonies, festivals and entertainment events (Berglund et al. 1999). The WHO sets the limit at $L_{p,\mathrm{A},4\,\mathrm{h}} = 100\,\mathrm{dB}$, and also restricts the number of such exposures to less than five per year. They also recommend that the sound level never should exceed $L_{\mathrm{A,Fmax}} = 110\,\mathrm{dB}$.

Although the WHO recommendation is slightly more liberal for single concerts, it restricts the number of events per year, where the Norwegian guideline does not.

Following the Norwegian guideline, and assuming 1.5 h long concerts with the allowed equivalent level, the total dose over the entire year will exceed the WHO recommendation if you attend more than 13 concerts.

There are several differences between single concerts and festivals. Most notably, the length of the exposure differs. A single concert can last from less than an hour to perhaps three hours. The sound exposure during such a concert is rather constant, possibly with a warm-up band before the main attraction. Music festivals are different. Many artists play rather short concerts, often less than an hour, but they typically play one after another. This gives a completely different exposure pattern, lasting for almost twelve hours, with periods of loud music and pauses that depend on how many concerts the participant choose to attend. In addition, many festivals cause sound exposure on sequential days. This is rarely true for attendants of single concerts, unless they are extremely dedicated spectators.

2.5.2 Music Festival Study: Exposure[a]

Two music festivals held in Norway during the summer 2014 were selected for the study, because of their long lengths. Hove festival outside Arendal lasted 7 days, and Øya festival outside of Oslo lasted 5 days.

Hove festival

Hove festival, or just ‘Hove’, lasted from 28 June to 4 July 2014, with an increase in the number of artists during the last four days. It was one of the largest music festivals in Norway, but it went bankrupt in September 2014. The festival had several camping sites near the concert area where a majority of the participants could stay. This made the festival popular for people from all over Norway and even Northern Europe.

Hove was arranged on an island called Tromøya, outside Arendal. This island has a bridge connection to the main land and is a recreational area for people living nearby.

The concerts started around 1 pm each day, but the number of stages used increased throughout the day and night, with the big headliners at the end. Since the festival area was rather isolated from private homes and other noise sensitive buildings, concerts did not have to end at 11 pm, but continued until between 2 am and 3 am each night.

Øya festival

In 2014, the Øya festival, also ‘Øya’, lasted from 5 to 9 August 2014. The first day’s concerts took place in clubs and discotheques in Oslo, while the last four days’ concerts were held in a park. Only the concerts in the park were used in this study. There are no camping cites associated with the concert area at this festival. Most of the visitors are therefore from the Oslo region.

At Øya the concerts started around 2 pm, except for the last day when it started at 1 pm. The festival was located in downtown Oslo, and surrounded by residential buildings. The local authorities therefore put restrictions to the organizers to follow the Norwegian guideline for concerts and festivals. This meant that all concerts had to end at 11 pm, and that there were sound level restrictions as mentioned above.

---

[a] The text in this section is taken from the paper presented in Noise & Health in 2016 (Tronstad and Gelderblom 2016).

Participants

Participants for the study were recruited from students of the Norwegian University of Science and Technology (NTNU). Posts on the university's intranet and posters around campus asked people already planning to attend the festivals to participate. No restrictions applied to the participants' age, sex or hearing ability.

It was important, from an ethical point of view, to only recruit persons already planning to go to the festivals, since the sound exposures at these festivals are potentially damaging to their hearing. The participants were informed about the risks involved in attending concerts and were allowed to wear hearing protection if they wanted to. All participants received NOK 1000 after the festivals as compensation for their participation.

Eight persons were recruited, five male and three female. Four males went to Hove and three females and one male went to Øya. The age was rather equal among all participants, with a mean of 20.8 ± 0.5 years.

The participants signed an informed consent form before the measurements began. They were instructed to act as normal festival participants. In addition to the dose measurements they also performed audiometric tests that will be presented in the next section (Sec. 2.5.3). Figure 2.3 shows an example of a measurement series for one participant from one day at the Hove festival. The corresponding 30 minute and four hour equivalent levels are plotted in the same figure.

Table 2.1: Statistical measures from the two festivals in the study.

Festival	Minutes	Minutes > 100 dBA	L10 [dBA]	L50 [dBA]	L90 [dBA]
Hove	13 155	879 (6.7 %)	97.0	76.8	58.1
Concerts	1683	804 (47.7 %)	107.8	99.6	90.7
Øya	9291	398 (4.3 %)	97.8	86.1	72.8
Concerts	3143	372 (11.8 %)	100.4	95.6	87.9
Both	22 446	1277 (5.7 %)	97.5	81.8	60.9
Concerts	4826	1176 (24.3 %)	103.8	96.7	88.3

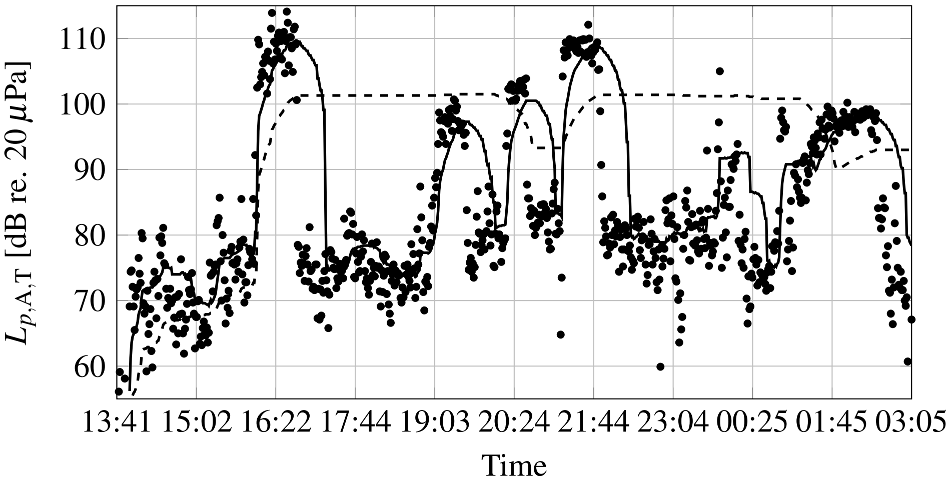

Figure 2.3: Example of one minute equivalent level measurements. Dots: One minute equivalent levels. Solid line: Calculated 30-minute equivalent levels. Dashed line: Calculated four hour equivalent levels. Both calculated levels use a sliding time window.

Based on the measurements from each participant the statistical measures L10, L50 and L90 were calculated. Table 2.1 shows these values and Figure 2.4 shows the distribution of the measurements. The figure shows that the data are not normally distributed and one has to take this into account when performing statistical analysis on the data.

Table 2.2 shows the sound dose in Pa^{2}h for each person. An event with

$L_{p,\mathrm{A,4h}} = 100\,\mathrm{dB}$, as the WHO allows, gives 16 $\mathrm{Pa}^2\mathrm{h}$. This recommendation is exceeded seven times at Hove, and none at Øya. Following the WHO recommendation with maximum four events with such level, the total yearly festival dose becomes 64 $\mathrm{Pa}^2\mathrm{h}$. This is exceeded by two persons at Hove. P1 might also have been overexposed, since two days of exposure are missing.

There was no statistical difference between Hove ($M = 11.1$ $\mathrm{Pa}^2\mathrm{h}$, CI=[6.8 18.2] $\mathrm{Pa}^2\mathrm{h}$) and Øya ($M = 7.0$ $\mathrm{Pa}^2\mathrm{h}$, CI=[5.0 9.7] $\mathrm{Pa}^2\mathrm{h}$) when looking at the daily exposures; $t(27) = 1.68, p = .105$. Nor the persons showed a significant difference ($F_{7,25} = .65, p = .70$) when looking at the daily doses.

The mean daily dose for all the participants was 8.9 $\mathrm{Pa}^2\mathrm{h}$. This corresponds to a four hour equivalent level of 97.5 dBA. It is, however, clear that some of the participants had a considerable higher exposure dose during some days. The fourth day for P4 and fifth day for P2 gave, for instance, total dose around 46 – 47 $\mathrm{Pa}^2\mathrm{h}$. This corresponds to a four hour equivalent level of approx. 104.6 dBA.

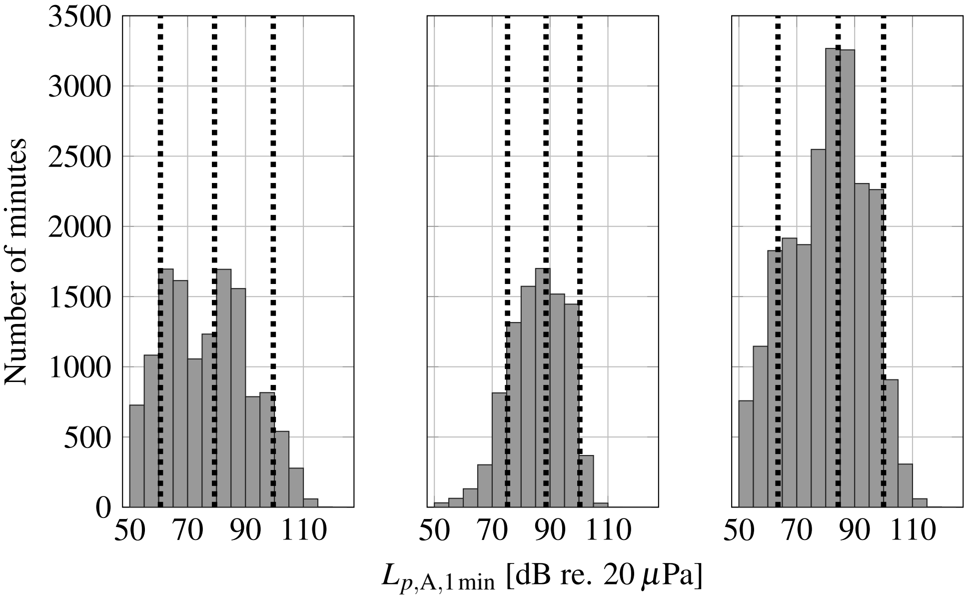

Figure 2.4: Distribution of one minute equivalent sound pressure levels during the festivals. The dotted lines in each plot correspond to L90, L50 and L10 from left to right. Left: Hove festival. Middle: Øya festival. Right: Both festivals.

Table 2.2: Daily sound dose, in Pa^{2}h, during the music festivals for each participant. *: Participant was exposed to loud sound, but data is missing. The total dose is therefore lower than the actual. **: Participant was not exposed to loud sounds, hence not wearing the dose meter. The total dose should be correct.

	Person	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Total dose**
Hove	**P1**	*	19.7	5.7	*	36.5	61.9*
	P2	**	16.8	3.3	12.3	47.0	79.4
	P3	3.4	4.2	5.8	13.8	19.3	46.5
	P4	9.7	2.1	9.0	46.3	25.4	92.5
Øya	**P5**	11.0	14.0	10.8	5.3		41.1
	P6	11.9	5.5	4.2	5.9		27.5
	P7	7.9	10.5	4.4	4.1		26.9
	P8	9.4	17.0	1.4	6.6		34.4

The equivalent sound level during each concert was also calculated. Individual concert lengths were used in the calculation. Focusing on the concert exposures only, the difference between Hove ($M = 101.4\,\mathrm{dBA}, \mathrm{SD} = 4.4\,\mathrm{dBA}$) and Øya ($M = 95.8\,\mathrm{dBA}, =3.5\,\mathrm{dBA}$) becomes highly significant; $t(60) = 6.65, p < .001$. At Hove neither the day ($F_{4,31} = 1.6, p = .20$) nor the persons ($F_{3,32} = .31, p = .82$) had any significant differences. At Øya the persons ($F_{3,59} = 1.31, p = .28$) did not differ significantly, but day one ($M = 97.4\,\mathrm{dBA}, \mathrm{SD} = 1.9\,\mathrm{dBA}$) and two ($M = 97.6\,\mathrm{dBA}, \mathrm{SD} = 3.3\,\mathrm{dBA}$) was significantly louder than day four ($M = 93.5\,\mathrm{dBA}$,
$\mathrm{SD} = 3.3\,\mathrm{dBA}$); $p_{\text{day one}} = .002, p_{\text{day two}} = .004$. The multiple comparison was performed using the Games-Howell Method.

2.5.3 Music Festival Study: Hearing Measurements[a]

The next section will show how the festival music affected the hearing of the participants, both on a short and long term. Both hearing thresholds and otoacoustic emissions were measured.

Equipment

A laptop with Matlab was used to measure the hearing thresholds with pure-tone audiometry. Sennheiser HDA 200 headphones and an Edirol UA-25EX sound card were connected to the laptop. The sound attenuating properties of the headphone can be seen in Table 2.3.

In addition distortion product otoacoustic emissions (DPOAE) were measured with an Otodynamics Echoport ILO292 USB-I.

Method

An automatic hearing test method presented in Sec. 2.4 was used in this study. The method was originally made for an active hearing protection device, but the algorithm works for other equipment as well. A Matlab GUI was used to play sounds through a sound card over standard audiometric headphones. The hearing test was not calibrated according to ISO 389-1 (ISO 389 1998), but since pre- and post-tests were performed with the same equipment, a relative status of the hearing level could be found. The test/re-test variability for the method, in laboratory, has been measured to be 4.0 dB, 4.2 dB, and 3.9 dB for

Table 2.3: Passive attenuation for the Sennheiser HDA 200. Data taken from the data sheet found on the Sennheiser webpage (Sennheiser Electronic Corporation 2016).

Frequency [Hz]	**Passive attenuation** [dB]
125	14.3
250	15.9
500	22.5
1000	28.6
2000	32.0
4000	45.7
8000	43.8

---

[a] The text in this section is taken from the paper presented in the AES conference proceedings in 2015 (Tronstad 2015).

3 kHz, 4 kHz, and 6 kHz (Vinay, Henriksen et al. 2014). These values have been used in this study.

The DPOAE measurements were done with $L1 = 65\,\mathrm{dB}$, $L2 = 55\,\mathrm{dB}$, and $f_2/f_1 = 1.22$. Because of the varying background noise level, the participants had ear muffs on when performing the test. The cable for the DPOAE probe was carefully placed inside the muff. This have been suggested as a possible solution when performing DPOAE measurements in noisy environments (Nielson et al. 2011). The test/re-test variability for DPOAE measurements have been reported to be less than 5 dB for the frequencies of interest (Poole 2011; Wagner et al. 2008; Sockalingam et al. 2007). Since the variability increases with the signal to noise ratio (SNR), and the SNR was varying during the measurements, a conservative limit of 5 dB was used for all frequencies in this study.

A four hour equivalent sound pressure level ($L_{p,\mathrm{A},4\,\mathrm{h}}$) was calculated for all the participants each day by using

$$L_{p,\mathrm{A},T_0} = L_{p,\mathrm{A},t} + 10 \cdot \log_{10}\left(\frac{t}{T_0}\right), \tag{2.1}$$

where $T_0 = 240\,\mathrm{min}$ (4 h), and t is the entire measurement period for each individual. Such equivalent level can be used as a sound dose measure, and in this study it was used to look into the exposure/response relationship. $L_{p,\mathrm{A},4\,\mathrm{h}}$ was used since this is a value recommended by the World Health Organization (WHO) in their guideline for events with loud sound (Berglund et al. 1999). Even if the exposure patterns in this study is above ten hours, the calculation of a four hour equivalent level is used to compare exposures between individuals. An eight hour time window could have been used (in accordance to most occupational legislations), but this would just result in a reduction of three decibels from the calculated values.

To characterize the exposures a statistical noise measure have been used. L10, L50 and L90 have been calculated for the one minute equivalents levels ($L_{p,\mathrm{A},1\,\mathrm{min}}$). This means that L10 is the level that are exceeded 10 % of the time, and so on.

Results

In Figure 2.5 an example of a time series of both the hearing level and the DPOAE measurements can be seen. The sawtooth pattern of the time series clearly show how the sound exposure during the festival days affect the hearing.

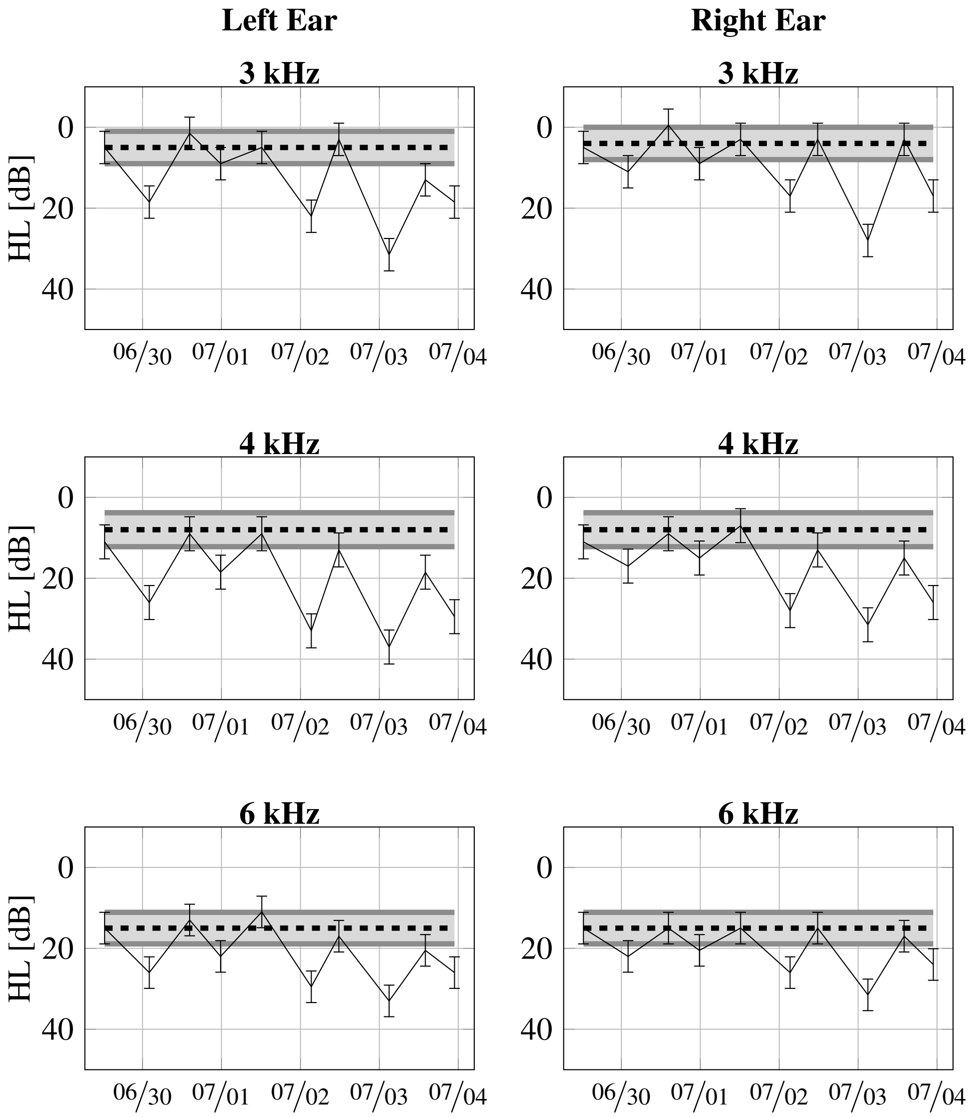

Figure 2.5: Example of a hearing level time series for one of the participants. The pre- and post-measurements are clearly shown in the sawtooth pattern, with higher hearing level at the post-tests. The shaded area is an estimated baseline plus/minus one standard deviation, using the first measurement from the festival and one post-measurement performed several weeks later with recovered hearing. The x-axis holds the date on format MM/DD.

Not all participants showed such distinct patterns, but all developed hearing threshold shifts and reduced emissions during the festivals.

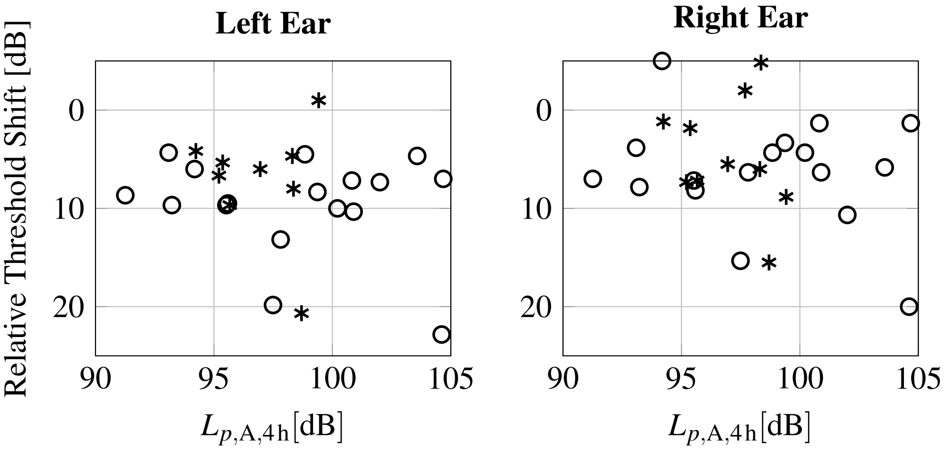

Figure 2.6: Relative change in average hearing threshold as a function of four hour equivalent sound pressure level. The arithmetic means of the measurements at 3 kHz, 4 kHz, and 6 kHz are used as average values. Stars: Øya festival. Circles: Hove festival.

Since all participants have differences in their exposure pattern, the hearing threshold shifts from a pre- to a post-test have been plotted as a function of the individual sound dose, represented by the four hour equivalent level (see Figure 2.6). A positive value means that there was a larger threshold shift at the post measurement.

Figure 2.7 shows a similar plot for the otoacoustic emissions. The relative change in DPOAE levels can be seen as a function of four hour equivalent level.

A linear regression was conducted on both the hearing level and DPOAE measurements to look into the EEH. No statistical significance were found between these levels and the $L_{p,\mathrm{A,4h}}$.

To elucidate the importance of restitution a pair of exposure patterns were selected from the measurements. These show two exposures with almost identical sound exposure dose, but with completely different effect on the hearing. Figure 2.8 shows the exposure patterns, and Figure 2.9 shows the corresponding hearing measurements.

All participants did a new hearing test several weeks after the festival. The re-test showed that the hearing levels were back to the values measured before

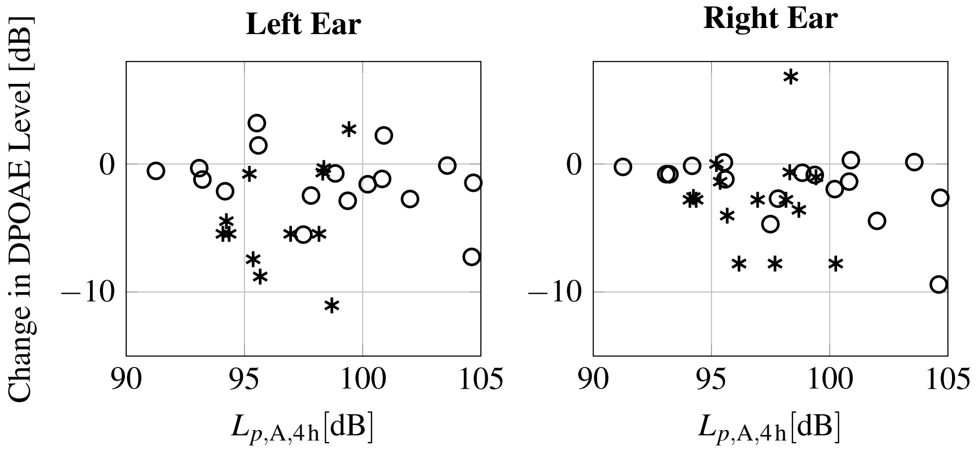

Figure 2.7: Relative change in the average distortion product otoacoustic emission (DPOAE) levels as a function of four hour equivalent sound pressure level. The average is the mean squared pressure for all the frequencies tested. Stars: Øya festival. Circles: Hove festival.

the festivals began. This indicates that their hearing have tolerated the concert exposure, but is not a guarantee. Resent studies have shown that it is possible to have what has been called a *hidden hearing loss* (Kujawa and Liberman 2009; Schaette and McAlpine 2011). This is a condition where the audiogram looks normal, but still there is a permanent damage to the cochlea. Since the DPOAE levels also were measured to be back to normal, this further indicates that their hearing did cope with the festival exposures. No further tests were conducted to verify the hearing status.

Music Festival Study: Discussion

The statistical analysis of the sound exposures for the participants clearly show that the concert exposures were different at the two festivals. The concerts at Hove had 47.7 % of the time with $L_{p,\mathrm{A,1\,min}}$ above 100 dBA while Øya had only 11.8 % of the time above this limit. The difference could have several explanations. From the table it is possible to see that the Hove participants went to less concerts than the Øya participants. This could mean that they only went to their favorite bands, and therefore stood closer to the stage when they did attend. The Øya participants, on the other hand, might have gone to more concerts with bands they did not know or particularly like, and therefore stood further from the stage in case they wanted to leave. In addition, only Øya had

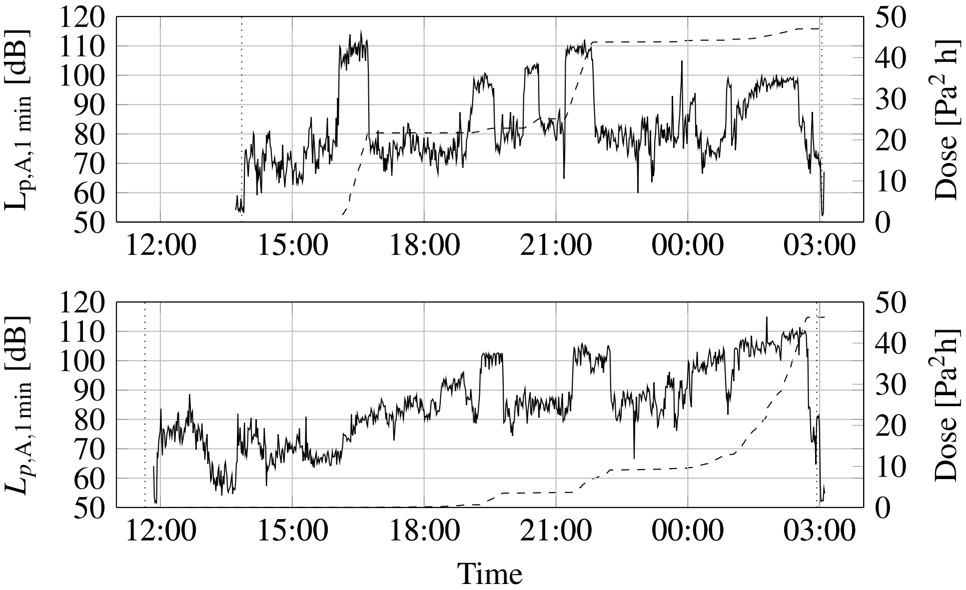

Figure 2.8: Plot of two exposure patterns with $L_{p,\text{A},4\text{h}}$ levels of 104.7 dB (upper) and 104.6 dB (lower). The solid lines show to the $L_{p,\text{A},1\,\text{min}}$, the dotted, vertical lines show the time of the hearing tests, and the dashed lines show the accumulated dose (in Pa^{2}h) on right y-axis.

to follow the Norwegian guideline for concerts and festivals, with limitations to the allowed sound pressure level. The festival monitored the sound level at each concert stage to make sure they did follow the guideline. Since Hove did not have any restrictions to the allowed sound level, this could also explain the differences in exposure. This does not, however, have any effect on the results in this study since each participant had a personal sound dose meter.

It is worth mentioning that the participants were free to drink alcohol during the festivals. Especially at the Hove festival the participants had high alcohol consumption. In addition the hearing tests were performed around 2 am and 3 am most nights, a time where the brain might not be performing at its best. Since testing of the hearing threshold requires high concentration, these two facts could have affected the variability of the post-measurements. A breathalyser was used to monitor the alcohol level for all the participants. Such method is, however, not suitable for testing high alcohol levels, and is also unreliable when testing close after the last drink. The levels recorded are therefore thought to be erroneous and have not been included in the analysis.

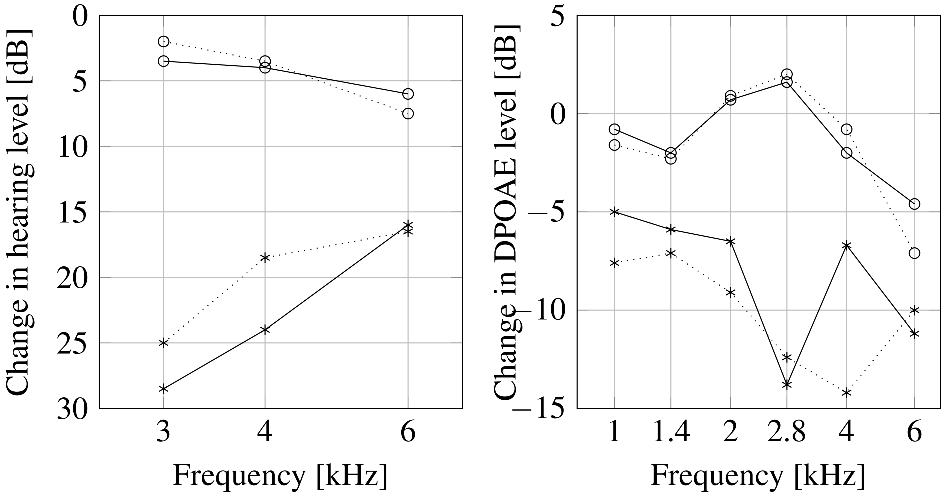

Figure 2.9: Relative change in hearing level for the exposures seen in Figure 2.8 (Circles: Upper exposure. Stars: Lower exposure.). Solid line: Left ear. Dotted line: Right ear.

Only the hearing threshold measurements, not the DPOAE levels, are thought to be affected by this possible laps of attention.

Also at Øya the variability in the automatic hearing threshold test might have been larger than expected. The reason for this was the background noise level during the tests. Even if the pre-tests were moved from the barracks to a nearby park, road traffic, birds and people made the background noise level clearly audible. The problem was most severe for the automatic hearing threshold test since the project manager could not control when the tones were played. If a noise event did coincide with a test tone, this could have lead to a false answer. The headset used does, however, have sound attenuating properties, especially at the test frequencies 3 kHz, 4 kHz, and 6 kHz. As can be seen in Table 2.3 the attenuation for this frequency region is around 40 dB. The lower frequencies (<500 Hz), that are less attenuated, is assumed to have minor masking effect on the test tones. The background noise level was not measured.

For the DPOAE measurements the background noise problem was minor both since the test indicates when the noise level is too high, and since the participants wore ear muffs above the DPOAE test probe. The DPOAE measurements were also continued until the results stabilized.

The amount of elevated threshold is known to increase with the sound pressure level. However, the exposures found at festivals are difficult to express with single values because of the time variation. Figure 2.6 and 2.7 show one interpretation where the relative change in hearing threshold is plotted as function of total daily sound dose. The figures show a large variation of threshold shifts and no significant relation was found. This is not surprising since several studies have been unable to find such relation (e.g. Strasser et al. (2003)). The lack of correlation with temporary changes does not mean that there is no correlation with permanent damage, something that also have been found in several studies (e.g. Hamernik et al. (2007)). Restitution time should, nonetheless, be included when considering the risk potential from such exposure.

The time series in Figure 2.5 show a possible accumulative effect of the sound exposure on the hearing threshold. Even if most measurements are within two times the standard deviation of the baseline one might see a negative trend where the pre-measurements does not fully recover back to normal values within the next day. Since this effect is not statistically significant (not shown), one should not make conclusions, but it seems reasonable that repeated exposures to sounds that give TTS, could have such accumulated negative effect. Since the recovery time after TTS is dependent on the amount of threshold shift, it is obvious that if the restitution time is not long enough, the hearing will not return to normal before the next exposure begins. Whether this would increase the risk of PTS is not as obvious, especially since it is known that sound toughening and TTS can have protecting effects (Ahroon and Hamernik 1999; Henderson et al. 1999; Housley et al. 2013). Since the other participants had even less pronounced accumulated effect, no further analysis was performed.

The selected case shows two exposures with almost identical energy, but different effect on the hearing, both DPOAE and hearing threshold levels. The person experiencing the largest change gets most of the sound dose during the last hours of exposure (lower plot in Figure 2.8), while the other gets most of the exposure almost six hours earlier. This will give large differences in restitution time, hence the measured change in hearing will also be large. Whether the risk of permanent damage is the same for the two exposures is an open question, since none of the participants experienced any measurable permanent changes in hearing.

All participants did a post-screening of their hearing to see if the thresholds and emission levels did recover completely. Since both the hearing thresholds and the emission levels did return back to normal values, it was concluded that no permanent threshold shift was obtained during the festivals.

2.5.4 Conclusion

The measurements from the music festival study did not find any linear correlation between the measured temporary hearing changes and the equivalent exposure level. This supports the finding from other studies where no correlation between energy and temporary threshold shift have been found.

All the participants did, however, get temporary hearing changes (both DPOAE and hearing level) during the festivals. The fact that their hearing did recover, mostly back to normal values, within the next day, show how remarkable the ears are at recovering after high sound exposures. Repeated temporary threshold shifts are, nonetheless, probably not beneficial for the hearing. Since there still does not exist any test to find those of us with more susceptible hearing, it is recommended that all participants at concerts and music festivals protect their hearing, either by standing further from the stages or by using hearing protection devices.

The music festival study was performed with the intention of finding a simple dose-response relationship. Even if all participants showed elevated hearing thresholds during the festivals, no simple relationship was found between the daily dose and the hearing ability. The main reason is probably the large variation in sound level distribution throughout the days, but individual susceptibility might also influenced the results. Nonetheless, it seems difficult to use sound exposure measurements to estimate the TTS a person might end up with.

For the hearing monitoring system this means that TTS will add to the personal variability in hearing measurements, something that must be taken into account when analysing the results.

Chapter 3

Hearing Monitoring

The following chapter will present hearing monitoring as a tool in the systematic approach to prevent noise-induced hearing loss. The current legislation is presented together with the proposed statistical process control approach.

3.1 Introduction

3.1.1 Legislation

Noise-induced hearing loss (NIHL) is one of the possible negative results of noise exposure. The current regime used in Norway states that employees with daily, i.e. 8 h, equivalent sound exposure levels above 80 dBA and/or peak levels above 130 dBC, should have their hearing checked as part of a systematic health examination, in addition to wearing compulsory hearing protection. The Norwegian Labour Inspection Authority says that the health examination should be carried out by, or under supervision of, an authorized medical doctor, and that the doctor should specify, based on a risk judgement, how frequently the examination should be performed (The Norwegian Ministry of Labour and Social Affairs 2016). However, since the associated guideline also states that employees in the mentioned group should get their hearing tested *at least* every three years, this is often the rate that is being used (The Norwegian Labour Inspection Authority 2013).

It is also specified in the above mentioned guideline that the hearing should be tested with pure tone audiometry at 0.5, 1, 2, 3, 4, 6, and 8 kHz, according to the standard NS-EN ISO 8253-1 (ISO 8253-1 2010), to find the hearing level

(HL). A permanent hearing damage is determined if there is a HL $\geq$ 25 dB for at least one of those frequencies, or if there is a HL $\geq$ 20 dB for 3, 4, and 6 kHz.

In addition the guideline uses an indicator of hearing damage caused by noise. If the HL is elevated more than 15 dB at 3, 4 or 6 kHz it is assumed that the hearing has been damaged by occupational noise, unless other reasons can explain the change. The worker should then be informed about the situation and be referred to a medical doctor for further medical examination and anamnesis. Individual counteractions can then be administered to reduce the risk of further negative development of the hearing.

3.1.2 Hearing Measurements

Most audiometric tests being performed as part of the health examination use a 5 dB step size (The Norwegian Labour Inspection Authority 2013; Franks 2001). Jerlvall and Arlinger (1986) did a comparison of the test-retest variability of pure tone audiometry using 2 dB and 5 dB step sizes, showing that no overall improvement was obtained by increasing the resolution. The standard deviations were found to be in the range from 2.1 dB to 7.2 dB, and similar values have been reproduced in later studies (Stuart et al. 1991; Henry et al. 2001; Smith-Olinde et al. 2006; Vinay, Svensson et al. 2015). Stuart et al. (1991) states that because of this large variation, an audiologist must see a 10 dB to 15 dB change in HL to confidently say that the hearing threshold has changed. This is probably the reason why the Norwegian Working Environment Act uses 15 dB in the noise damage indicator.

Hearing conservation 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. 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.

3.2 Using Statistical Process Control in Hearing Monitoring [a]

The next sections of this thesis will introduce a statistical framework for detection of small threshold shifts using frequent measurements of the hearing threshold level. Such an approach can result in earlier detection of a negative development of the hearing and can become a new barrier against hearing loss.

First the concept of statistical process control (SPC) will be described, 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. Finally a discussion of the findings is presented.

3.2.1 Statistical Process Control

Statistical process control (SPC) has been a field of research since Walter A. Shewhart started his work around 1924 (Best and Neuhauser 2006). At present it is, for instance, used in the process industry to monitor manufacturing of products, as well as in financial and administrative processes. Often it is beneficial to have an early detection of when a process is out of control, meaning that some parameter has changed and the production is not fulfilling some predefined requirement. This can for instance be the production of nails that need to be 40 mm long. A certain variance in the production must be allowed, e.g. ± 0.1 mm, but if the nails suddenly are 41 mm long, they cannot be sold as the product they intended to be.

Several different SPC techniques exist, depending on what parameter is monitored, how many variables are measured, how large changes one wants to detect and what the underlying probability distribution of the process is. This paper will only consider the situation where a process is monitored by measuring one normally distributed variable (so called univariate control charts), and

---

[a] Some of the text in this section was presented in the International Journal of audiology in 2017 (Tronstad 2017), but additional details have been added here for completeness.

where detection of both small and large shifts in the average value is desired.

A control chart is simply a diagram which shows a measured value as a function of measurement number or time. The chart might show the ‘raw’ data points or some processed values, such as running average etc. SPC can then be implemented by using such control charts where it is possible to determine if the process is in or out of control.

Often subgroups are used as input to the control charts. By using an average of several measurements the variability is reduced, which in turn makes the control chart more efficient. There are, however, situations where sub-grouping are impractical. For these situations it may be beneficial to use individuals control charts (Amin and Ethridge 1998).

To monitor the process average 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 (2013) for a more thorough description of all the different control charts.

Figure 3.1 shows two examples of control charts where the process mean value and the variability is monitored. The mean value is monitored using an X chart, and the variability is monitored using an mR chart. This combination is also known as an XmR chart. Since none of the measurements are outside the control limits the process is judged to be ‘in control’.

Average Run Length

One measure of evaluating the performance of different control strategies is the Average Run Length parameter (*ARL*). This value describes the average number of measurements needed to detect a certain shift in a parameter value. Even though some authors disagree about the use of *ARL* as a performance measure, it is quite commonly applied as a means of comparing control charts (Montgomery 2013, p. 200).

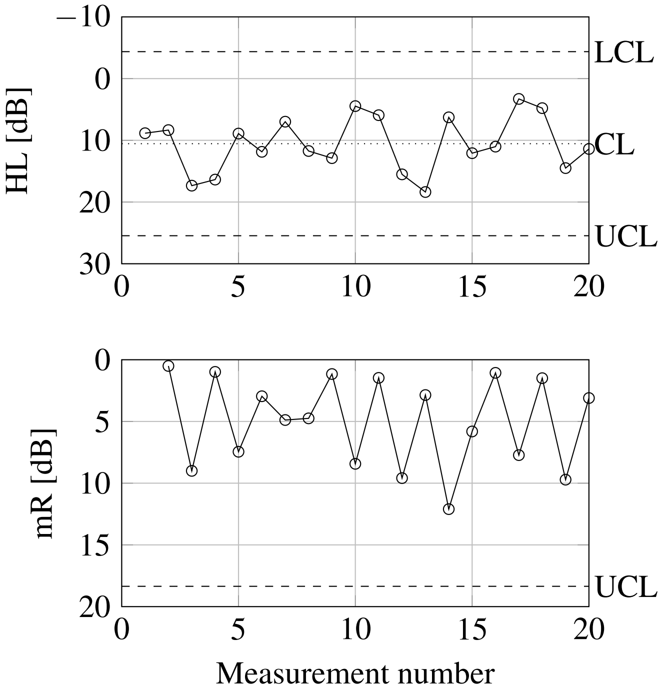

Figure 3.1: Illustration of an XmR control chart with hearing level, HL, as primary data points. Upper: X chart showing the measurements (circles), together with the upper- (UCL) and lower control limits (LCL), and the centre line (CL). Lower: Calculation of the two-span moving range, mR, (circles) for the observations in the X chart together with the upper control limit (UCL). Notice that the y-axis is flipped to correspond with normal presentation of hearing levels.

A sensitive control chart exhibits a small ARL value when the process is 'out of control'. This means that when a change in the process has occurred the error is detected quickly. Control charts should exhibit 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 applied in hypothesis testing. Commonly, the notation ARL_0 is used to describe the 'in control' situation, while ARL_1 is used when the process is 'out of control'. It is important to note that by decreasing ARL_1, the ARL_0 value 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 for this 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 value to adjust for this (Hawkins and Olwell 1998). If a total of m control charts are used, the actual ARL_0 value for the joint process control can be estimated by

$$ARL_{0,\text{Group}} \approx \frac{ARL_0}{m} \tag{3.1}$$

This means that when one is setting control limits, the ARL_0 value for each individuals control chart must be multiplied by m in order to obtain the desired false alarm rate for $ARL_{0,\text{Group}}$.

Run Length Percentile

This thesis also utilises another measure, the run length percentile, to evaluate process control charts. Other authors have also used such representations of the control chart quality (Khoo and Quah 2002; Chakraborti 2007). The rationale for this approach is that the distribution of 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, may give a false impression of how good the control chart really is. Percentiles can also be useful for setting control limits for the control chart. For example, it may be possible to specify that only 3 % of the control charts should have a run length less than 100 when ‘in control’. In this thesis, Monte Carlo simulations are used to identify such percentiles.

SPC Implementation

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

During Phase I, the goal is to specify the control limits to be applied in Phase II. Several authors have looked into the issues of the number of observations required in Phase I to obtain reliable estimates of the process parameters, and how to adjust the control limits for a given number of observations. Jensen et al. (2006) conducted a literature review and demonstrated that recommendations vary between 100 to 300 observations for individuals control charts. Hawkins (1987) pointed out that many practitioners use a much lower number of observations, often around 25. However, as the author points out, such low numbers are liable to provide incorrect parameter estimates, especially for the standard deviation, which in turn can affect the performance of the control charts.

Another important point is that the process should be in control during Phase I. If not, the control limits can be biased and the performance of the control chart compromised. There are, however, techniques to handle measurements which deviate from the process in control. One of the techniques is to use trial control limits to assure oneself that the process is actually in control. If some measurements fall outside the trial control limits, it is possible to look for assignable causes and exclude these measurements and calculate new control limits for Phase II. Such methods must, however, be used with caution.

When the control charts have been setup Phase II can start. During this period the goal is to detect an out of control process as soon as possible (i.e. a small ARL_1), without giving false alarms (i.e. a large ARL_0). This is done by finding a centre line (CL) and use control limits and decision rules to determine if the new data points are in or out of control. The control limits are often called upper control limit (UCL) or lower control limit (LCL), depending on which side of the parameter they are, as presented in Figure 3.1.

There is also a procedure known as a self-starting control chart. Here, Phase I can be omitted and estimates of the unknown process parameter values are generated continuously as new observations become available. This type of chart is particularly suitable when the number of observations is small, and when it is cumbersome to collect more samples.

If one compares the self-starting control charts with control charts with known parameters, the performance will be worse. This means that the control chart will be less sensitive to changes in the underlying process and/or have more frequent false alarms. However, this comparison is not useful because the charts are based on two completely different premises. If the process parameters are known, one should always use this information during process control. However, if the parameters are unknown one must decide whether to obtain estimates of the parameters during a Phase I (see above), or use a self-starting control chart. The first option is recommended if process observations are readily available. If this is not the case, the latter will be more efficient.

In the next section two different categories of control charts are shown; control charts which base the decision on single observations, and time-weighted control charts, which base the decision on several consecutive observations.

All the observations in this paper are assumed to come from a normal distribution with mean value μ_0 and standard deviation σ, i.e. $X_i \sim N(\mu_0, \sigma^2)$.

3.2.2 X Chart

The X chart, or Shewhart individuals chart, is a plot of the individual measurements, X_i, and each observation is used to evaluate if the process is in or out of control as exemplified in Figure 3.1. This chart is good at detecting large shifts quickly. Often the control limits are set to be three times the standard deviation of the data, also called three-sigma control limits (Montgomery 2013). For data from a normal distribution, approx. 99.7 % of the samples will be within $\mu_0 \pm 3\sigma$. The ARL_0 can be calculated to be around 370 when such limits are used. The control limits can be expressed

$$UCL_X = \mu_0 + 3\sigma \tag{3.2}$$

$$LCL_X = \mu_0 - 3\sigma \tag{3.3}$$

Since the standard deviation, σ, often is unknown, it must be estimated. Several possible solutions exist for the individuals control chart, but Wheeler (2010) has pointed out that only two are appropriate; the mean and the median of the moving range, mR (see below for more details). It can be shown that the mean value of mR (denoted ($\overline{\text{mR}}$) overestimates σ and that this can be corrected for by multiplying with 0.886, and the median (denoted $\widetilde{\text{mR}}$) underestimates σ and must by multiplied with 1.047 (Bryce et al. 1997). The control limits then become

$$\text{UCL/LCL} = \begin{cases} \mu_0 \pm 3 \cdot (0.886\overline{\text{mR}}) = \mu_0 \pm 2.66 \cdot \overline{\text{mR}} & \text{(Mean based)} \\ \mu_0 \pm 3 \cdot (1.047\widetilde{\text{mR}}) = \mu_0 \pm 3.14 \cdot \widetilde{\text{mR}} & \text{(Median based)} \end{cases} \tag{3.4}$$

mR Chart

The mR values can be used not only for estimating the standard deviation, but plotted in the so-called mR chart which can complement the X chart in order to monitor the development of the variance in the process. Whether or not to include the mR chart is, however, debatable. Some authors argue that little or no information is added to the X chart by including the mR chart, but Amin and Ethridge (1998) point out that the combination will be better at detecting a shift in only the process variability. Wheeler (2010) also states, in a popular science article, that the mR chart has interpretative benefits and that it can be used to get information about how the control limits have been calculated.

The so-called two-span mR is used as input to the mR chart. The term two-span indicates that two measurements are included in the calculation of each mR value. The mR can be calculated by taking the absolute value of the difference between the current measurement and the previous one:

$$mR_i = |X_i - X_{i-1}| = \max(X_i, X_{i-1}) - \min(X_i, X_{i-1}), \quad i = 2, \ldots, n \qquad (3.5)$$

The span of the moving window can be larger than two, but it can be shown that a larger span will only increase the uncertainty of the estimator if the measurements are contaminated by erroneous data (Woodall and Montgomery 2000).

When setting up an XmR chart the control limits for the X- and mR chart must be set together to get the desired ARL values. Amin and Ethridge (1998) present a possible procedure to find control parameters using an iterative approach with curve read-out and computer program simulations.

Often one is trying to detect if the variability increases, i.e. only an UCL is used in the mR chart, but LCL is also possible to implement.

Decision Rules

To detect whether a process is in or out of control different decision rules exist. A set of such rules are the Western Electric Rules first introduced in 1956 (Western Electric Company 1956). These rules use three different zones around μ, also referred to as the centre line (CL), to base the decision on. These can be seen in Figure 3.2.

The Western Electric Rules suggest the following to signal a process out of control:

1. One measurement outside the UCL or LCL limits, or
2. Two out of three consecutive points in Zone A, or
3. Four out of five consecutive points in Zone A or B, or
4. Eight consecutive points in Zone A, B or C

All the points in each rule must be on the same side of the centre line, and the process is judged to be out of control if one of the rules is broken.

The introduction of such detection rules will increase the probability of detecting a process out of control, especially for small changes in the process mean, but it will also affect the false alarm rate. Applying all four Western

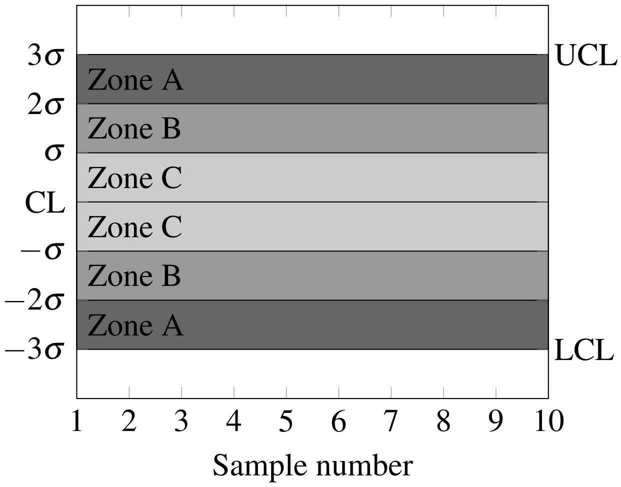

Figure 3.2: Illustration of the different zones used in the Western Electric Rules for an $\overline{X}$ or X chart. CL: Centre Line, UCL: Upper Control Limit, LCL: Lower Control Limit. The control chart use the three sigma control limits.

Electric Rules will, for instance, reduce ARL_0 to 91.75, instead of 370 when only rule 1 is used (Champ and Woodall 1987). One therefore has to adjust the control limits, similarly to the Bonferroni correction mentioned above, to get the desired ARL_0 value. This adjustment will, however, also change the performance to detect larger shifts, and it is therefore not obvious if it is beneficial to include more than the first rule, or if another control chart should be used. To detect small changes in the process one might instead use time-weighted control charts, such as CUSUM or EWMA. This will be discussed next.

3.2.3 Time Weighted Control Charts

Where the X chart only uses information about the current observation (when decision rule 1 is applied) to decide whether the process is in or out of control, the time-weighted control charts use information from several of the preceding samples to base the decision on. Montgomery (2013) suggests that either the CUSUM, a CUmulative SUM of the deviation from the mean value, or the EWMA, an Exponential Weighted Moving Average, can be used to detect small shifts in the process mean.

CUSUM

If one has a process where it is important to detect both a positive and a negative shift of the process mean, the CUSUM chart is constructed by making one cumulative sum for the negative shifts (C_i^-), and one for the positive (C_i^+).

The CUSUM values are calculated with the following equations (adapted from Montgomery (2013, p. 418)):

$$C_i^+ = \max\left[0, X_i - (\mu_0 + K) + C_{i-1}^+\right] \tag{3.6}$$

$$C_i^- = \max\left[0, (\mu_0 - K) - X_i + C_{i-1}^-\right] \tag{3.7}$$

In the equations X_i is the current observation, μ_0 is the process mean, K is usually called a reference value (or the allowance, or the slack value) that can be chosen for optimal response to a shift of a specified size (Hawkins and Olwell 1998), and $C_0^+ = C_0^- = 0$. Note that only observations that deviate from μ_0 with more than K will add to the C_i^+ and C_i^-. The accumulated values are also reset (set to zero) when the value becomes negative. This is done to make the CUSUM values stable over time. If this resetting was not done, the variation of the CUSUM value would increase as more observations were included, and the possibility of extreme values would also increase.

Control limits similar to those used for X charts can be used to decide when the process is out of control. This approach is called a tabular, or algorithmic, CUSUM (Montgomery 2013). To set the limit one must first choose a value for K. This is done by using the following equation:

$$K = k\sigma \tag{3.8}$$

The value of k can be selected, and is often chosen as 0.5δ, where δ is the shift, in number of standard deviations, one wants to be able to detect. For instance if one has data with $\sigma = 5$, and want to be able to detect a shift of one standard deviation ($\delta = 1$), then $K = 0.5 \cdot 1 \cdot 5 = 2.5$.

Next the control limit, $H = h\sigma$, must be decided. For a given value K, a value of h can be found that gives a desired ARL_0 value. Using $k = 0.5$ and $h = 4.77$ will give an $ARL_0 = 370$, which is approximately the same as a Shewhart control chart with 3σ control limits. Often h values around 5 are used (Montgomery 2013).

The tabular method can also be made more sensitive by introducing a *headstart* or *fast initial response* (FIR). This means that C_0^+ and/or C_0^- is given a non-zero value, typically $H/2$. This will make the CUSUM more sensitive to a change early in the process.

Another approach to decide when the process is out of control is the so-called V-mask procedure. This method places a lying V-shaped mask around the last observation and the process is considered in control as long as all the previous observations are inside the arms of the mask. However, if any of the preceding observations is outside the arms, the process is out of control. The lower plot in Figure 3.3 shows an example of a V-mask CUSUM.

The CUSUM value to be used as input is slightly different from the tabular version presented above:

$$C_m = \sum_{i=1}^{m} (X_i - \mu_0) = (X_i - \mu_0) + C_{m-1} \tag{3.9}$$

One might notice that the reference value, k, is omitted, and the C value is never reset but is a 'true' cumulative summation of all deviations from the mean.

The V-mask can be constructed to behave identically to the tabular CUSUM. This is done by defining (adapted from Montgomery 2013, p. 429)

$$k = A \tan\theta \tag{3.10}$$

and

$$h = A\,d \tan\theta = dk \tag{3.11}$$

where θ is the angle of the V-mask lines, and d is the distance from the last observation to the point where the two lines in the V-mask meet. A is the horizontal distance on the V-mask plot between successive points in terms of unit distance on the vertical scale. The parameters k and h are the same as defined above.

From Figure 3.3 it is also possible to see that the performance of the two CUSUM versions are identical. Both flag the process to be out of control at observation 36.

A strength of the CUSUM is that it can be used to estimate the point where the process went out of control. When this point is found, it can be easier to identify the cause of the change.

For the tabular version the CUSUM value is expected to increase when the process is out of control (i.e. a shift in the expected value). It is therefore

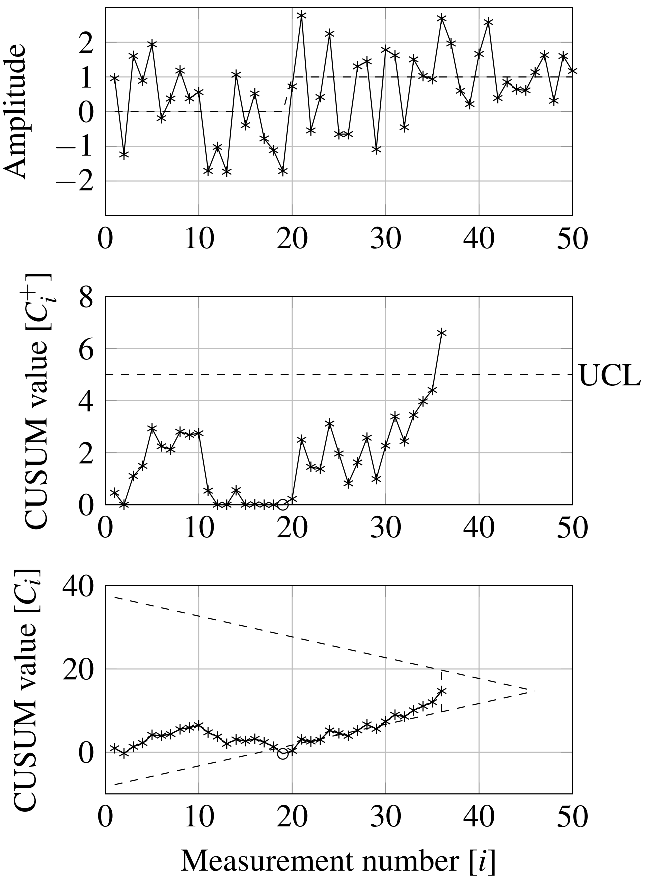

Figure 3.3: Example of a set of individual measurement points (upper), a tabular (middle) and V-mask (lower) CUSUM. UCL: Upper control limit.

possible to go backwards in the control chart to find the last point where the CUSUM value was zero. This point is considered to be the last observation where the process was in control. For the example in Figure 3.3 (middle diagram) it is possible to see that this point is observation 19 (marked with a circle).

For the V-mask CUSUM one must use a different technique. When the process is considered out of control one must go back and find the first observation that went outside the V-mask. For the example in Figure 3.3 (lower diagram) it is possible to see that several observations are outside the lower V-mask line, but the first observation is number 19 (marked with a circle), the same as for the tabular version.

The data used in the example in Figure 3.3 had a one standard deviation shift introduced at observation 20.

Johnson (1961) has proposed a method to design the V-mask, and Montgomery (2013, p. 430) has later reformulated this method. The parameters d and θ can be calculated by using α (α_0 in Johnson's paper) and β (α_1 in Johnson's paper). α is a type I error, i.e. 2α is the greatest allowable probability of a false alarm, and β is a type II error, i.e. the probability of not detecting an actual shift of size δ. Montgomery rewrote Johnson's equations to

$$\theta = \tan^{-1}\left(\frac{\delta}{2A}\right) \tag{3.12}$$

and

$$d = \left(\frac{2}{\delta^2}\right)\ln\left(\frac{1-\beta}{\alpha}\right) \tag{3.13}$$

Montgomery does however strongly advise against the use of the V-mask procedure, and has three points to support his statement (Montgomery 2013, p. 430):

1. The headstart feature, which is very useful in practice, cannot be implemented with the V-mask.
2. It is sometimes difficult to determine how far backward the arms of the V-mask should extend, thereby making interpretation difficult for the practitioner.
3. Perhaps the biggest problem with the V-mask is the ambiguity associated with α and β in the Johnson design procedure.

It has also been stated (NIST/SEMATECH 2012) that the V-mask procedure is actually a carry-over from the pre-computer era, and that standard statistical software easily can implement the tabular method.

Exponentially Weighted Moving Average

The exponentially weighted moving average (EWMA) is defined as

$$z_i = \lambda X_i + (1-\lambda)z_{i-1} \tag{3.14}$$

where $\lambda \in (0,1)$ and $z_0 = \mu_0$. The variance of z_i can be found to be

$$\sigma_{z_i}^2 = \sigma \left(\frac{\lambda}{2-\lambda} \right) \left(1 - (1-\lambda)^{2i}\right) \tag{3.15}$$

This means the control limits can be expressed as

$$\mathrm{UCL} = \mu_0 + L\sigma \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i}))} \tag{3.16}$$

$$\mathrm{LCL} = \mu_0 - L\sigma \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i}))} \tag{3.17}$$

The control limits are centred around μ_0 and L is a width factor that can be chosen to adjust the sensitivity of the control chart. For long run lengths the variance, and thus the control limits, approach a stable value:

$$\mathrm{UCL} \underset{i \gg 1}{\approx} \mu_0 + L\sigma \sqrt{\frac{\lambda}{2-\lambda}} \tag{3.18}$$

$$\mathrm{LCL} \underset{i \gg 1}{\approx} \mu_0 - L\sigma \sqrt{\frac{\lambda}{2-\lambda}} \tag{3.19}$$

Figure 3.4 shows an example of an EWMA control chart.

A strength of the EWMA control chart is that it is very insensitive to the normality assumption (Borror et al. 1999). This makes it popular for processes with short run-lengths and/or rational subgroups of size $n = 1$ (i.e. individual data points), where the underlying probability distribution is uncertain.

Montgomery also mentions one potential concern about the EWMA. If the EWMA value is on one side of the centre line when an out of control shift occurs in the opposite direction, then the detection could take several observations (Montgomery 2013, p. 437). This is called the inertia effect.

Self-starting Control Charts

As mentioned earlier, self-starting control charts do not require a Phase I, since the parameters needed are estimated continuously as new observations become available. This makes self-starting control charts particularly attractive in connection with short run-lengths, since monitoring of the process can start immediately.

Quesenberry (1991) demonstrated the implementation of such a control chart, in which a value Q is defined as a basis for monitoring. The Q value is defined

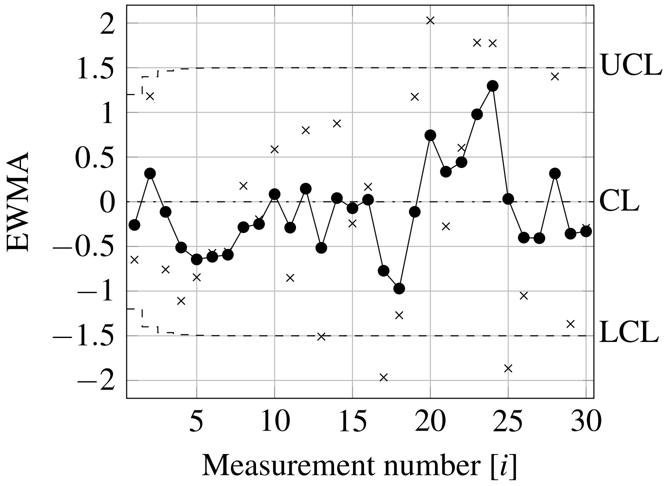

Figure 3.4: Example of an exponentially weighted moving average (EWMA) control chart. The solid line with dots represents the EWMA values, and the individual observations are marked with x's. The control parameters used in the control chart are $\lambda = 0.4$ and $L = 3$. UCL: Upper control limit. CL: Centre line. LCL: Lower control limit.

as

$$Q_i = \Phi^{-1}\left[G_{i-2}\left(\sqrt{\frac{i-1}{i}}\left(\frac{X_i-\overline{X}_{i-1}}{S_{i-1}}\right)\right)\right], \quad i = 3,4,\ldots \tag{3.20}$$

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

$$\overline{X}_i = \frac{\sum_{n=1}^{i} X_n}{i}, \quad S_i^2 = \frac{\sum_{n=1}^{i}(X_n-\overline{X}_i)^2}{i-1}, \quad i = 2,3,\ldots \tag{3.21}$$

It is also possible to calculate $\overline{X}_i$ and S_i^2 by updating their values as new observations are available, using

$$\overline{X}_i = \overline{X}_{i-1} + \frac{X_i - \overline{X}_{i-1}}{i}, \quad S_i^2 = S_{i-1}^2 + \frac{(i-1)(X_i-\overline{X}_{i-1})^2}{i}, \quad i = 2,3,\ldots \tag{3.22}$$

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

Q Chart

If Q values are used as input to the X chart, it is possible to construct a self-starting control chart for individual values. This is called a Q chart, and since the Q values are normally distributed it is a straightforward matter to set control limits that are identical in all processes that should have the same ARL_0 value. A disadvantage of the Q chart is that the Q values are normalised, and can thus be difficult to interpret, for instance in a plot.

CUSUM Q Chart

Q values can also be used as input to the time-weighted control charts. For example, Quesenberry (1991) proposed using both a CUSUM and an EWMA control chart. Another suggested method is the adaptive CUSUM of the Q chart (ADQ), which can be even more effective at detecting a range of shifts in the mean value (Li and Wang 2010). However, the simulations presented by Li and Wang are not exclusively in favour of the ADQ and will not be discussed further in this thesis. The CUSUM Q value can be found by using the equations above and replacing X_i with Q_i and exploiting the fact that Q is a standard normally distributed variable:

$$S_i^+ = \max\left[0, Q_i - k + S_{i-1}^+\right] \tag{3.23}$$

$$S_i^- = \max\left[0, -k - Q_i + S_{i-1}^-\right] \tag{3.24}$$

where $S_0^+ = S_0^- = 0$, and k is the reference value.

Self-starting EWMA

Other self-starting versions of the EWMA has also been proposed. Jones (2002) has developed a procedure to calculate control limits when the unknown parameters are estimated during the process control. If such a control limit adjustment is not performed, the false alarm rate (ARL_0) will be higher than expected. Adjusting the control limits with the proposed method will result in the expected ARL_0 value, at the expense of a higher ARL_1 value.

Li, Zhang et al. (2010) propose a different method they call a self-starting EWMA likelihood ratio test (SSELR). The unique characteristic of this method is that both the mean and the variance are monitored simultaneously in one control chart.

Yet another method has been proposed by Tsiamyrtzis and Hawkins 2008, where an EWMA is used in a Bayesian framework. This means that prior information about the values in the in-control situation is used to estimate the distribution of the variables. As more observations are gathered, this information is also used as prior knowledge in the next step of the monitoring. This method is especially good at detecting jumps in the start-up of a process.

Erroneous Data

Erroneous data, or outliers, represent a challenge for any type of control chart. Large outliers will be interpreted as major shifts in the process, and will be flagged as an 'out of control' situation if they are not otherwise dealt with. Furthermore, since the estimation of both the mean and standard deviation will be affected by such data points, the performance of the control chart may be negatively impacted for the remainder of the process control.

There are two common approaches to handling outliers; truncation and winsorisation. Truncation means that values are simply removed from the data set, while winsorisation involves replacing 'suspect' values with the closest 'non-suspect' value. Other methods to detect and treat outliers exist, but the reader is referred to other literature for these (e.g. Ghosh and Vogt 2012).

Hawkins (1980) proposed a solution for time-weighted control charts where outliers are detected. This involved the application of maximum Q values. If any Q value exceeds a preset limit, W, the value is winsorised to this limit. If W is selected wisely, this method provides effective protection against large outliers, but it also means that all data points will contribute to the control chart. One must also be aware that when winsorisation is applied, the control limits must be adjusted to provide the same ARL value.

However, it is insufficient simply to winsorise the Q values. As previously mentioned, if an outlier value in the underlying dataset is not dealt with, it may have a serious negative impact on the estimation of parameters such as the standard deviation, in particular, and the mean. A possible solution here is to use more robust estimators such as the median of the data as a estimator of the expected value, and the median of the moving range as an estimator of the standard deviation. However, in general, the use of more robust estimators will have a negative impact on the performance of the control chart. Another solution is to apply truncation or winsorisation to the data used in the estimation as well.

3.2.4 SPC for the Hearing

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 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 project an automatic hearing test, described by Vinay et al. (Vinay, Svensson et al. 2015), is used as the underlying test for the monitoring of the hearing. The process control regime is, however, possible to apply to any hearing test that is carried out frequently enough.

When setting control limits for the control charts it is reasonable to take a look at the occupational legislations. In Norway most workplaces with high sound exposures test the hearing of their workers, as required by the The Norwegian Labour Inspection Authority (The Norwegian Labour Inspection Authority 2013). Even if their guidelines recommend annual testing, often the minimum rate of one test every three year is used.

In these guidelines a person is defined as hearing impaired if the HL is worse than 25 dB for one, or worse than 20 dB for all, of the test frequencies 3, 4, and 6 kHz. When a threshold shift is found, an anamnesis should be performed to determine a probable cause of the change. The threshold shift is considered work-related until it has been identified that other causes are responsible for the change.

The guidelines also use an indicator for NIHL, which can be used to implement counteractions for individuals showing signs of elevated HL. This indicator is defined as a change of at least 15 dB at 3 kHz, 4 kHz or 6 kHz between two consecutive hearing measurements. If a change is found in one measurement, it must be verified by a re-test before any further conclusions are made.

If a control chart should be useful in the context of the Norwegian guidelines 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.

For the hearing process one wants to detect if a person's HL is increasing, i.e. the hearing is getting worse. Only one-sided control limits should therefore be considered.

Presbycusis, or age-related hearing loss (ARHL), must also be taken into account. As people get older their hearing inevitably deteriorates. A decision must be taken as to whether this should be corrected for, or flagged as an ‘out of control’ process. One approach is to adjust the expected HL value with an estimation of the ARHL. The progress of ARHL is described in ISO 7029 (2000), and it may be possible to use this estimation as input to the control chart. However, it is difficult to provide an accurate estimate for an individuals threshold level because 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 medical personnel as part of their follow-up evaluations, once an ‘out of control’ signal has been detected. Nevertheless, the use of a self-starting control chart will take ARHL into account to some extent. Since the HL value is estimated continuously throughout the monitoring process, a slow shift in true HL values will lead to a drift of the estimated value. This effect will be elucidated in the following Monte Carlo simulations.

Specifying In-Control Parameters

One of the most important parts of the control chart procedure is to establish the so-called ‘in control’ parameters. If estimates of the process mean or standard deviation are incorrect, control chart performance may be severely deteriorated (Jensen et al. 2006). This is especially true for time-weighted control charts used to detect small changes.

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

Several studies have been conducted to investigate the test-retest variability of both manual and automatic hearing tests. In general, it has been shown that the standard deviation is below 5 dB, and is commonly reported to be about 2 – 3 dB (Vinay, Svensson et al. 2015; Jerlvall and Arlinger 1986; Stuart et al. 1991; Henry et al. 2001; Smith-Olinde et al. 2006). There are thus two possible options for the estimation of variability. One can either employ the same standard deviation value for all individuals, or estimate individual standard deviation values. Since the standard deviation can vary substantially between individuals, the use of individually estimated values will probably increase the reliability of the control chart. Moreover, if the person performing the hearing tests is very consistent, it will be possible to detect small changes faster than if a person’s responses exhibit high levels of variability. An estimation of intrasubject variability will thus be applied in this thesis.

The process mean, i.e. the HL baseline, will also be an individual value. It is possible to estimate this parameter based on standard audiometric data measured by an audiologist. However, since as previously mentioned, the test-retest variability resulting from manual audiometry tests carried out by professionals is similar to that for automatic tests, the use of standard audiometric data as the HL baseline, rather than the data points in the control chart, will not increase the reliability of the control chart.

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 thesis.

Proposed Control Chart

Even if self-starting control charts may exhibit slightly worse performance than a thoroughly executed Phase I/Phase II regime, there are arguments in favour of starting the hearing monitoring quickly rather than focus on the detection of very small hearing threshold shifts. Therefore, a self-starting chart is chosen in the proposed control chart. Furthermore, both an X chart, and a CUSUM Q chart will be employed as described in the following.

An X chart using HL values as input with '3 σ' control limits will be presented, together with the running estimate of the mean value. This means that if a measurement deviates more than three standard deviations from the mean value, it is flagged to be 'out of control'. The person undergoing the test must then be prompted to perform a new measurement to verify the results. If the following measurement is also 'out of control', a notification will be made that a large hearing threshold shift seems to have occurred, and health personnel should be noticed. This approach is very similar to the recommended procedure issued by The Norwegian Labour Inspection Authority (2013) in situations where a standard deviation of 5 dB is assumed. The difference is that instead of using the same standard deviation for all individuals, the individual standard deviation estimate is used to determine the limit. This means that for an individual who provides consistent responses (i.e. with small standard deviation values), smaller hearing threshold shifts will be detected, and vice versa. Moreover, a CUSUM Q chart is used to detect small, persistent changes. As mentioned previously, an EWMA chart could also have been used, but CUSUM is preferred because of its ability to estimate the point at which the process goes 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 in relation to outlier data points. Thus, the approach involving winsorization of the Q value, as described above, will be used together with the proposed rejection of outliers in the parameter estimation.

Alarms given by the X chart are triggered by sudden large data points (i.e. a large increase in hearing level). Even if it is possible to suddenly have an increased hearing level, e.g. as response to extremely loud sound exposures, there is a large probability that such data point is an outlier. Therefore alarms given by the X chart should not be used directly to determine if the process is ‘out of control’. This further means that false alarms from the X chart will not contribute to the overall false alarm rate, hence they are not included in the simulations below. This is considered acceptable since a persistent shift will in most cases quickly be detected by the CUSUM Q chart anyway. One should, however, be aware that especially shifts occurring early in the process control can go undetected by the CUSUM Q chart if the estimation of the mean value and/or standard deviation is wrong. Alarms from the X chart should therefore be visually inspected by personnel to determine the cause of the alarm.

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 influenced by the choices of ARL_0 and ARL_1 values.

The first choice, target value ARL_0, depends, among other things, on the cost associated with flagging an ‘out of control’ signal. When the control charts detect a possible change, the person undergoing the test must be sent to the occupational health service provider (OHSP). Most companies test their employees once every three years, so it may be a reasonable option to stipulate that control charts should not provide false alarms more frequently than every three years. Otherwise, adherence to the charts will result in increased costs to the company since employees will be obliged to visit the OHSP more often than before.

Figure 3.5 shows the percentage of processes with a critical run length (CRL) less than three selected values (100, 150, and 200) as function of CUSUM control limit. Each point along the curves are found by running Monte Carlo simulations of 10000 control charts and counting the process controls with a run length less than the CRL. The curve showing a CRL of 100 indicates how large percentage of the control charts have false alarms more frequently than once per 100 observations. This line can be used to set a control limit for the CUSUM control chart for a selected false alarm rate. For instance, if the requirement is that less than 10 % of the process control charts should give a false alarm more frequently than once per 200 observation, it is possible to read out that the control limit for the CUSUM chart should be approximately 5.9 and 4 when the allowance value, k, is 0.5 and 0.75, respectively.

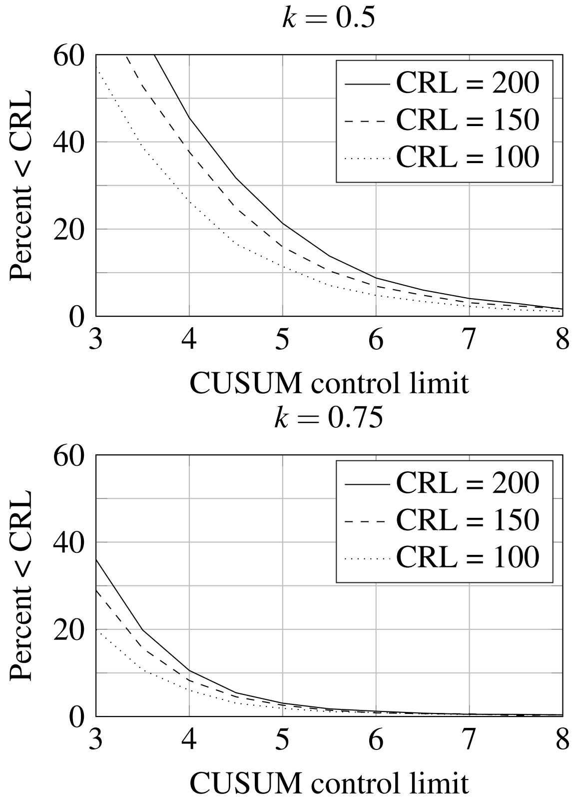

Figure 3.5: Percentage of process controls that give a run length below the Critical Run Length (CRL) as function of CUSUM control limit for allowance values of (upper) $k = 0.5$, and (lower) $k = 0.75$.

Identifying a control limit also requires a determination of how many observations (hearing tests) the employees should perform each year. In this thesis it is assumed that hearing tests are performed approximately once a week (i.e. ≈ 40 measurements per year). If a decision is made that only 5 % of employees should have a false alarm rate of less than 100 (leading to approximately three years between each false alarm), the control limits will be 6.1 and 4.2 for $k = 0.5$, and $k = 0.75$, respectively.

The second choice, target value ARL_1, will be affected by the control limit settings. If a large number of false alarms cannot be tolerated, the ability to detect changes will be affected adversely. However, by changing the allowance value, as mentioned in Sec. 3.2.3, some adjustment of the performance of the control chart can be achieved. This will be elaborated in the following section describing the use of Monte Carlo simulations to test different shifts of the mean value. By increasing the allowance value k, the control chart will be made more robust against variations in the observed data points, but the ability to detect small changes will be reduced.

3.3 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)
- Presbyacusis (P)
- Comparison with the current regime (C)

Figure 3.6 shows an illustration of the three first situations, and presbyacusis can be seen in Figure 1.1.

NS simulates the non-damaged ear, and provides an illustration of the false alarm rate (ARL_0). SS simulates a sudden hearing loss that may occur following an exposure to a loud sound. RS simulates a progressive deterioration of hearing. Even if such a linear approach is probably too simplistic to describe a real case of progressive hearing loss, it will provide an insight into the effect of such loss at different rates of progression. P simulates how ARHL is detected by the control charts, and finally C will compare the presented method with

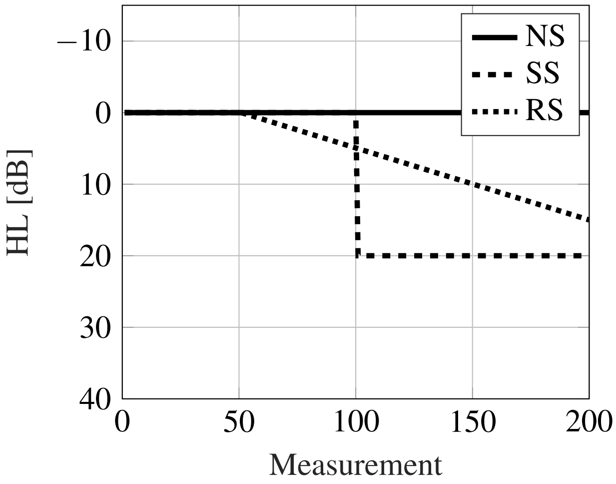

Figure 3.6: Illustration of the situations simulated. No shift (NS) shows the situation where no change in the hearing is simulated. The 0.1 dB/obs ramp shift is introduced at measurement number 50 and the 20 dB step shift is introduced at measurement 100.

the current control regime used in e.g. Norway.

Both SS and RS used data sets each containing 2000 simulated observations. The NS example employed 20 000 observations in order to achieve better estimates of ARL_0. The individual observations were random values derived from a normal distribution with zero mean and a standard deviation of 5 dB. These values were chosen because they are typical for a hearing test as described 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 result in a somewhat lower *ARL* estimate than the theoretically correct one.

For the SS and RS examples, if the 'out of control' signal was flagged before the shift was introduced (false alarm), the simulation was rejected and replaced with a new one. This was considered acceptable since information about the false alarm rate can be derived from the NS simulations.

The development of presbycusis is described in the international standard ISO 7029 (2000). Figure 3.7 shows the 50 % percentile estimations for males and females for the frequencies 3, 4, and 6 kHz. The curves are expressed by equations on the form:

$$\mathrm{HL}(\mathrm{Age}) = \alpha(\mathrm{Age} - 18)^2 \tag{3.25}$$

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

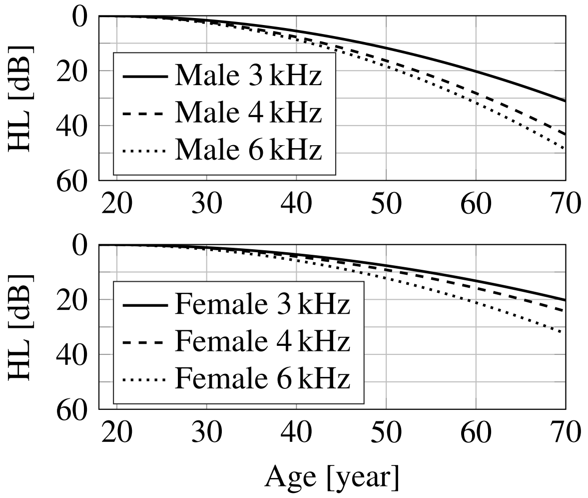

Figure 3.7: Estimation of hearing thresholds as a function of age and gender according to ISO 7029 (2000).

Two cases were used as input to the Monte Carlo simulations; 6 kHz for males, and 3 kHz for females. These are the cases containing the most extreme values ($\alpha_{\text{male,6kHz}} = 0.018$ and $\alpha_{\text{male,3kHz}} = 0.0075$, respectively), so simulations of the other frequencies will generate results between these two curves. Different observation rates, from once every third year to 200 per year, were simulated, and the observations distributed evenly over time. A random distribution of the observations was also simulated, but the results of this are not shown because they were almost identical to those for an even distribution over time. Such randomising corresponds to hearing tests that are not performed at regular intervals.

Erroneous data points, or outliers, have been simulated by randomly adding 25 dB to some observations. However, no outliers were permitted as part of the first ten observations. The reason for this is that early outliers are quite detrimental to the estimation of the process parameters, and it is easy to prevent such errors in practice. By conducting a training session during which the person in question performs between five and ten tests under close supervision, large outliers can be prevented. Moreover, the outliers were applied only as an increase in HL, i.e. an apparent deterioration in hearing, since the most probable reasons for outliers, such as background noise, loss of concentration and other interruptions, all result in an elevated threshold. Furthermore, outliers in this direction have the most negative impact on process control because they increase the values of the mean and standard deviation, resulting in less sensitive process control.

How soon a threshold shift is detected, i.e. how low values of ARL_1 can be reached, depends on the accuracy of the parameter estimation. In general, parameters are more precisely estimated as more observations become available. So if a shift arises early in the monitoring, the detection capability will be worse than if a shift arises later. Observation numbers 50, 100, 150, and 200 are employed as starting points for the SS and RS examples in order to demonstrate this effect.

Last, a comparison with the current regime is conducted. This is done to be able to answer the hypothesis whether the proposed hearing monitoring system is as good as, or better, than the current hearing test regime used in occupational settings. In order to compare the method presented here with the current regime, three large hearing threshold shifts are simulated; 10 dB, 15 dB, and 20 dB.

3.3.1 No Shift, NS

Figure 3.8 shows the run length distributions for two sets of selected control parameters for the CUSUM Q chart, together with the ARL_0 values (from an X chart), when no threshold shift is introduced. As can be seen, the two distributions are far from normal, but are approximately identical. The ARL_0 values are also almost identical, which is expected because the control parameters have been selected so that they should be similar. It can also be seen that the non-normal distribution means that the mean and median values are quite different. This implies that about 63 % of persons tested will experience a false alarm more than once every 2700 measurements, which is the mean value. Since we are using six control charts to monitor the hearing, the $ARL_{0,\text{Group}}$ can be found to be (using Eq. 3.1) to be around 450.

3.3.2 Step Shift, SS

The results from the SS example are shown in Table 3.1 and Figure 3.9. If we examine the 95 % percentile we see that a shift smaller than one standard deviation [run length 133 ($k = 0.5$) and > 200 ($k = 0.75$)] will almost never be detected by the CUSUM Q chart if the shift occurs early in the process (during

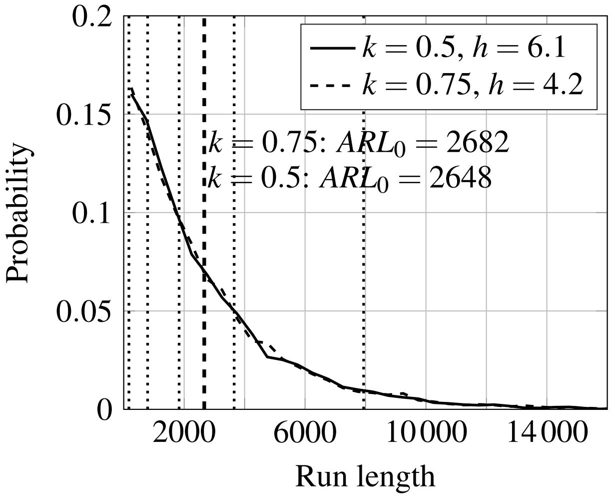

Figure 3.8: CUSUM Q chart run length distribution for the no shift situation. The ARL_0 values (vertical dashed lines) are almost identical for the two parameter sets. The vertical dotted lines, displayed from left to right, show the 5 %, 25 %, 50 %, 75 %, and 90 % percentiles.

the first 50 observations). However, the same shift will 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 $k = 0.5$ which has proved to be the most efficient), the shift is detected for 95 % of the 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.

Table 3.1: CUSUM Q chart run lengths after a step size 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
Step shift onset: after 100 observations	5 %	4	3	2	1
	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

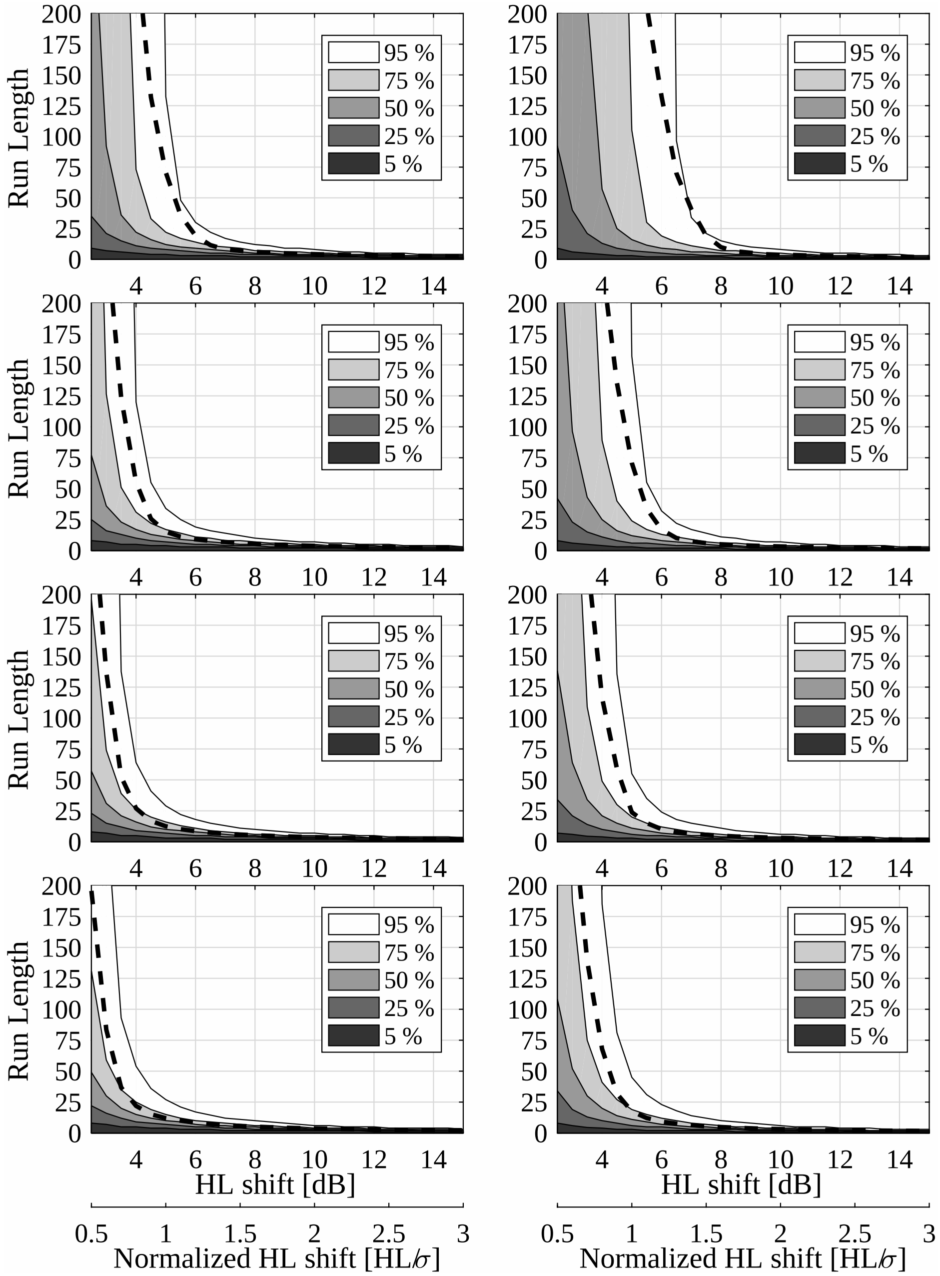

Figure 3.9: CUSUM Q chart run lengths after a step shift has been introduced. The change in hearing level (HL) is on the x-axis. The normalized change, in HL per standard deviation, is also shown. Left column: $k = 0.5$, $h = 6.1$. Right column: $k = 0.75$, $h = 4.2$. The four rows show different points of onset for the step shift. Onset point, from upper to lower: 50, 100, 150, and 200. The dashed line is the estimated *ARL* (which is somewhat underestimated for large run length values).

3.3.3 Ramp Shift, RS

Table 3.2 and Figure 3.10 show the run length percentiles for the RS scenario. The table only shows the results for two RS rates (0.1 dB/obs and 0.4 dB/obs, respectively), whereas Figure 3.10 shows curves for different RS rates.

Both the table and the figure show that the onset time for the ramp shift is not as critical for the detection capability as for the step shift scenario because the results are similar for all onset points. It is also possible to see from the figure that the distribution of run lengths is symmetrical since the *ARL* is close to the 50 % percentile.

Figure 3.11 shows an estimation of the magnitude of the threshold shift before it is detected at the different rates. To calculate this estimate, the *ARL* value is simply multiplied by the rate. Even though the numbers of observations are greater for the small than for the large rates, the accumulated threshold shift is actually smaller for the smaller rates. Thus it appears to be the case that the SPC approach may be efficient in detecting such small, but steadily progressing, changes in measured values, especially if hearing is tested frequently.

Table 3.2: 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

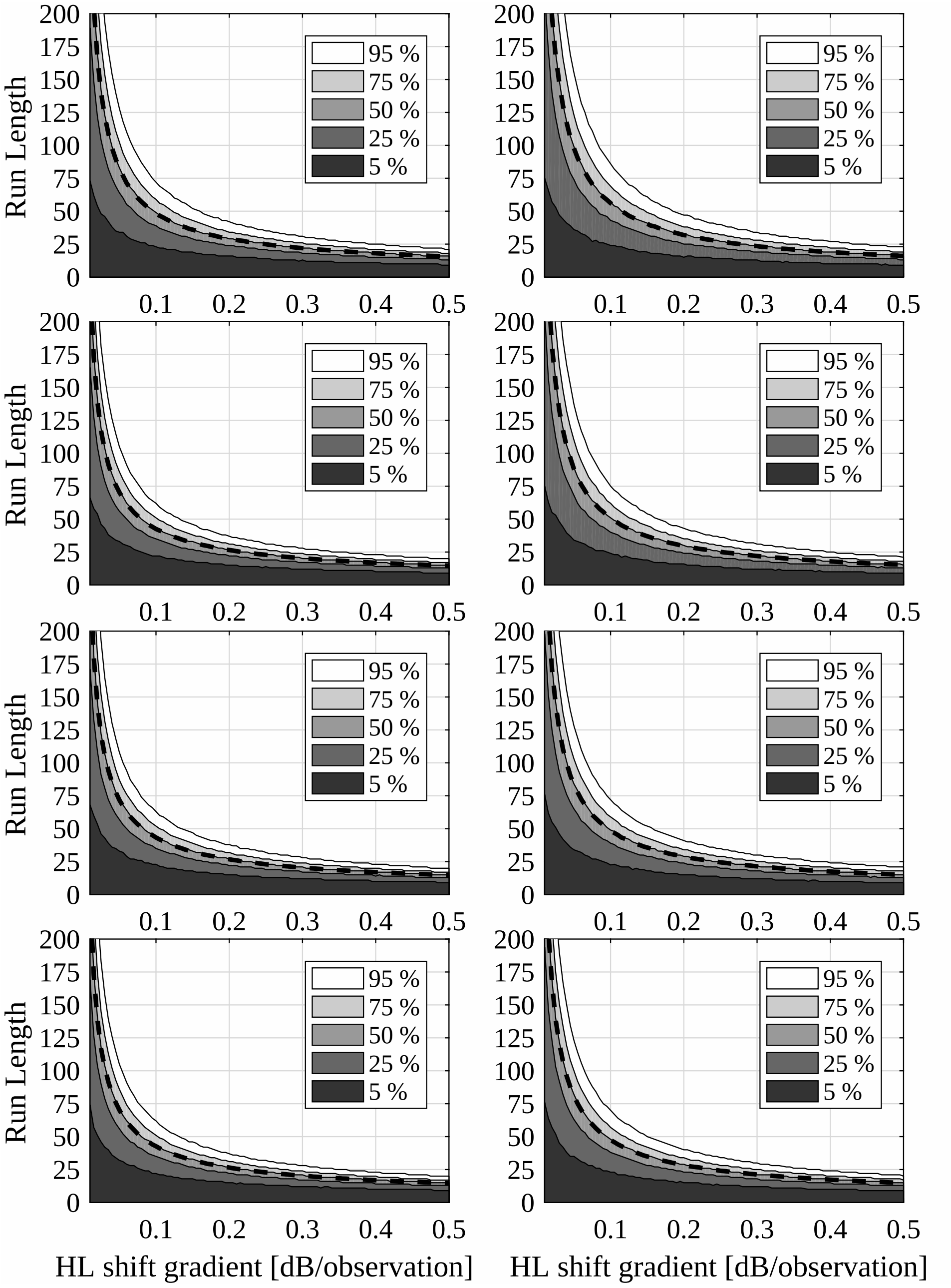

Figure 3.10: Run lengths for different ramp shift rates. The change in hearing level (HL) is on the x-axis. Left column: $k = 0.5$, $h = 6.1$. Right column: $k = 0.75$, $h = 4.2$. The different rows show different points of onset for the ramp shift. Onset point, from upper to lower: 50, 100, 150, and 200. The dashed line is the estimated *ARL*.

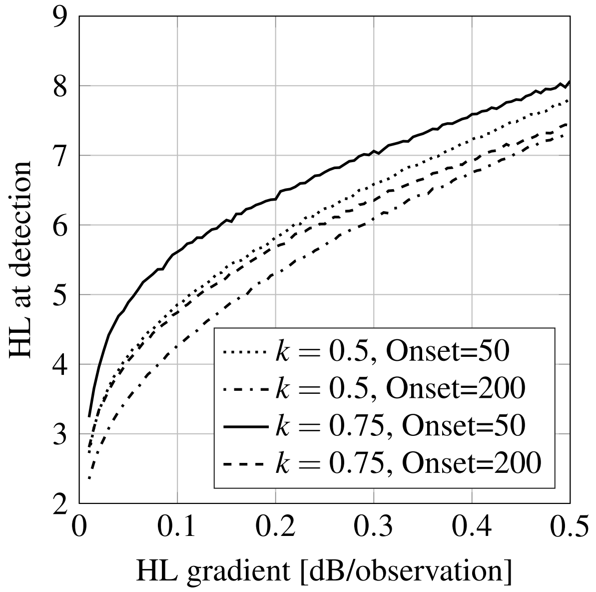

Figure 3.11: Estimation of degree of hearing loss accumulated before a ramp shift is detected.

3.3.4 Presbyacusis, P

The results from the P simulations are shown in Figure 3.12. If a male tests his hearing approx. 40 times per year, the ARHL will typically not be detected before 10 years have passed. For females using the same test rate, approx. 30 years will go before the ARHL it is detected. This means that such ‘false’ alarm will not affect the monitoring until after several years. It is also possible to see that more frequent hearing tests will also detect ARHL earlier. This can be a challenge since it is difficult to differentiate between NIHL and ARHL. It is, however, possible to use ISO 7029 (2000) to determine if the shift is related to normal ageing, and this will be shown later in Ch. 5.

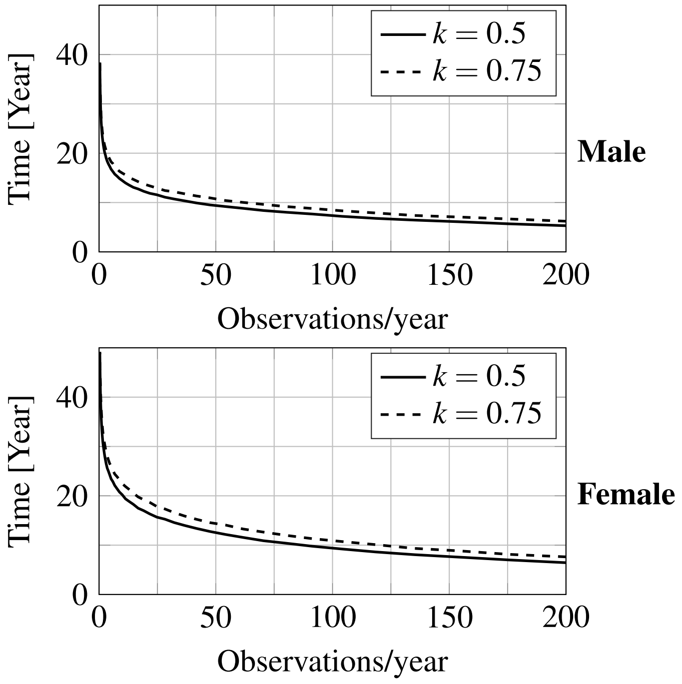

Figure 3.12: Plots showing when a typical presbycusis will result in flagging of an ‘out of control’ process, given that the process control is started at age 18, for different numbers of observations per year. The upper plot is for a male, 6 kHz ($\alpha = 0.018$), and the lower for a female, 3 kHz ($\alpha = 0.0075$).

3.3.5 Erroneous Data Points

To illustrate how outliers can affect a process control, a simulation was performed involving the insertion of erratically distributed erroneous data points with and without application of the outlier detection method. Figure 3.13 shows how the process control collapses when no outlier detection is applied, even with as few as five outliers randomly distributed among the first 200 observations.

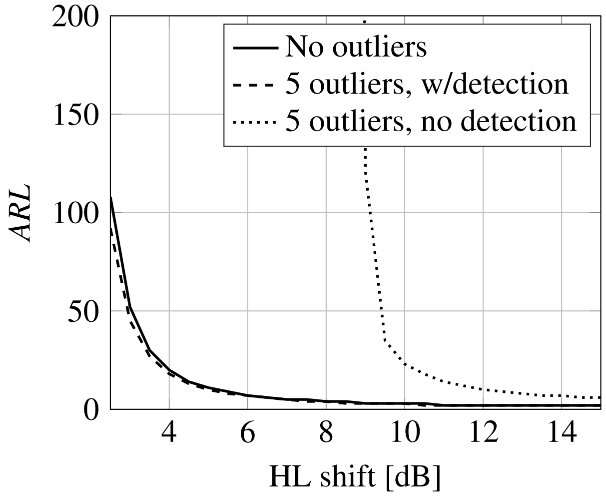

Figure 3.13: Illustration of outlier sensitivity. The plot shows the average run length (*ARL*) from the CUSUM *Q* chart for different shifts in hearing level (HL). Three situations are shown; no outliers, five outliers with outlier detection and winsorising of outliers, and five outliers without any outlier detection or counteractions applied.

3.3.6 Comparison with Current Regime, C

The focus in the previous subsections was on the detection of relatively small hearing threshold shifts. However, it is also important that large shifts can be detected quickly. In order to assess this, a comparison between the proposed hearing monitoring system and the current regime in Norway has been performed. Before this can be done, a number of assumptions must be made. First of all, it is assumed that all hearing measurements are normally distributed with standard deviations of 3 dB and 5 dB (both situations are compared). Secondly, it is assumed that the commonly used step size of 5 dB is applied during pure tone audiometry. Furthermore, the procedure described by the The Norwegian Labour Inspection Authority (2013) is followed. 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 of one measurement is made every three years.

Figure 3.14 shows the probability distribution of measurements performed using pure tone audiometry as constrained by the assumptions described above. For example, the figure shows that the probability of measuring the correct HL when the standard deviation is 3 dB is 60 % ($P_{\sigma=3\,\text{dB}}(0\,\text{dB}|0\,\text{dB}) = 0.60$), and 38 % ($P_{\sigma=5\,\text{dB}}(0\,\text{dB}|0\,\text{dB}) = 0.38$) when the standard deviation is 5 dB. The probability of measuring larger values than those presented in the figure is so small that this is omitted in the following discussion.

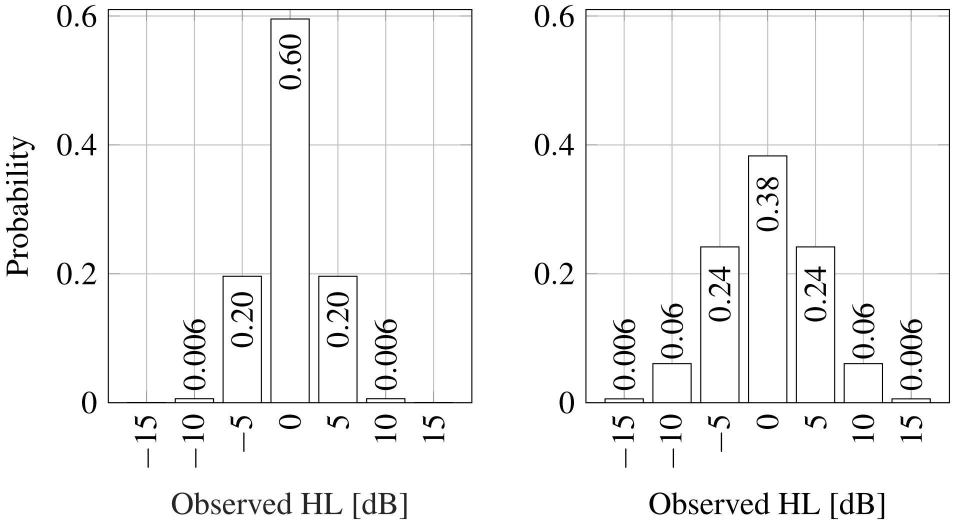

Figure 3.14: Two plots showing the distribution of results from hearing measurements using a 3 dB (left) and a 5 dB (right) standard deviation, combined with a 5 dB step size applied during the audiometry test. The observed HL is the difference between the measured and the actual hearing level.

In order to compare the two methods, three different step shifts have been assessed; 10, 15, and 20 dB. The probability of detecting these shifts under the current regime can be found by calculating the probability of measuring two consecutive values larger than 15 dB, using the probability distributions shown in Figure 3.14.

The probability of measuring values at 5 dB or greater than the actual threshold shift is the sum of $P(5\,\text{dB}|0\,\text{dB})$, $P(10\,\text{dB}|0\,\text{dB})$, and so on. For example, this means that the probability of measuring a single value at or above 15 dB, when the actual hearing level is 10 dB, is 30.9 % ($0.24+0.06+0.006+\cdots = 0.309$), given a standard deviation of 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 3.3 shows the probability of detecting true threshold shifts of 10 dB, 15 dB, and 20 dB under the current regime.

Table 3.3: Probability of detecting a given shift using the procedure described by The Norwegian Labour Inspection Authority (2013).

	Probability	
Threshold Shift	$\sigma = 3\,\mathrm{dB}$	$\sigma = 5\,\mathrm{dB}$
10 dB	4.1 %	9.6 %
15 dB	63.6 %	47.8 %
20 dB	98.8 %	87.1 %

The probability of detecting the same shifts using the method presented above can be estimated by performing Monte Carlo simulations. Figure 3.15 shows the results from 10 000 simulations of the different threshold shifts.

The plots in Figure 3.15 show that the probability of detecting a large shift quickly increases as more observations are performed, and exceeds the probabilities found under the current regime (see Table 3.3), after two (3 dB standard deviation) or three (5 dB standard deviation) observations. This means that if more than one hearing measurement is performed each year, performance will be better using the method presented in this paper. We also observe that probabilities rapidly approach 100 % reliability once more than five to ten observations have been made. This is in large contrast to the current regime which will only detect 4.1 % ($\sigma = 3\,\mathrm{dB}$) or 9.6 % ($\sigma = 5\,\mathrm{dB}$) of the individuals with 10 dB PTS.

Figure 3.16 shows the results from the simulation of a progressive hearing threshold shift. Three different rates are simulated; 2 dB/obs, 5 dB/obs, and 10 dB/obs. As can be seen, the most challenging rate is the largest progression. This might seem counter-intuitive, but is a consequence of the decision of making the control chart robust against outliers. If the progression rate is 10 dB/obs the hearing threshold shift will become 40 dB before it is reliably ($P > 0.9$) detected by the CUSUM Q chart. The X chart will, however, most likely give a warning such that more frequent hearing measurements can be initiated, leading to a lower progression rate.

It is difficult to compare these situations since they use two different measurement schemes, but one might notice that frequent hearing measurements (i.e. small hearing threshold growth rates) are beneficial under any circumstances. It is also possible to read out from the figure that it is necessary to have a progression rate of 2 dB/observation (or less) to reliably detect a 15 dB hearing threshold shift. If we assume that such threshold shift is developed during a period of three year this would correspond to at least 3 hearing measurements per year.

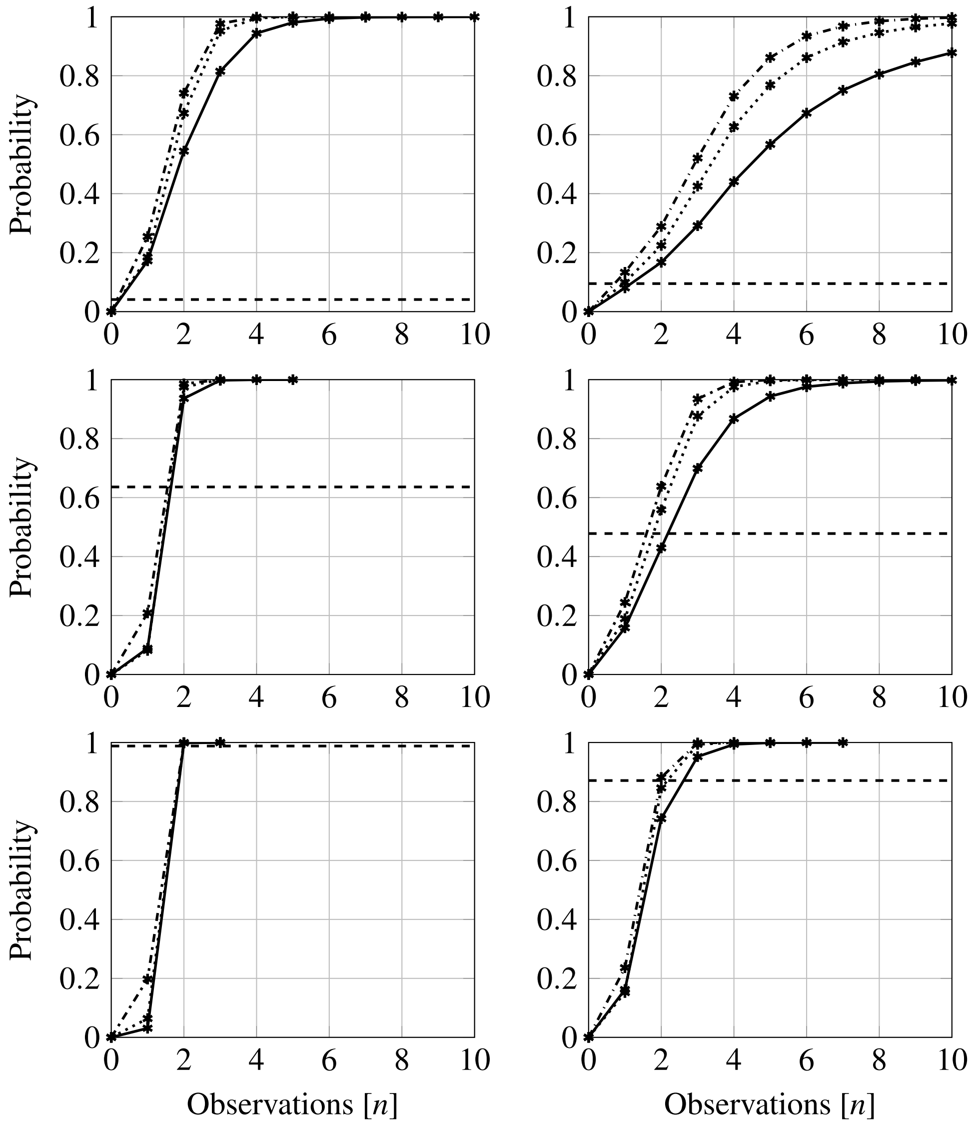

Figure 3.15: Plots showing the probabilities of detecting a 10 dB (upper), 15 dB (middle), and 20 dB (lower) hearing threshold shift using the method proposed in this thesis. The plots on the left show the results for a standard deviation of 3 dB, those on the right for 5 dB. The hearing threshold shifts are introduced at three different points; observation 20 (—), 50 (·····) and 200 (-·-·). The dashed lines (- - -) show the probabilities of the current regime.

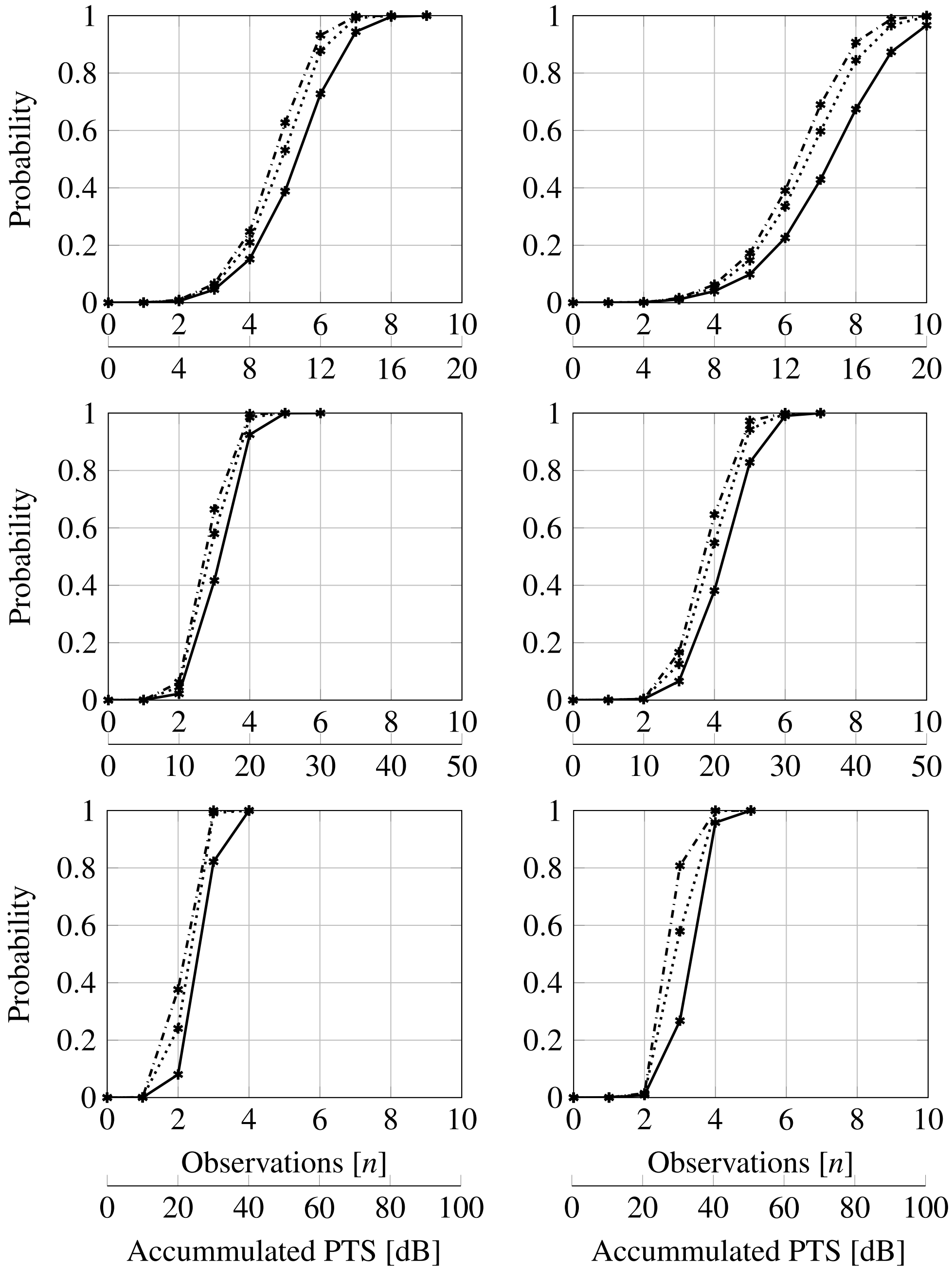

Figure 3.16: Probability of detecting a relative large progressive threshold shift. Upper: 2 dB/obs. Middle: 5 dB/obs. Lower: 10 dB/obs. Left: $\sigma = 3$ dB. Right: $\sigma = 5$ dB. The hearing threshold shift are introduced at three different points; observation 20 (——), 50 (·····), and 200 (-·-·). The upper x-axes on all plots are the number of observations, and the lower x-axes are the accumulated hearing threshold shift.

3.3.7 Real-world Data

Currently, the oil company Statoil ASA is employing the hearing monitoring regime on two offshore oil and gas installations. Figure 3.17 shows an example of a time series of hearing level measurements combined with the corresponding process control taken from one of the users. The figures display the three frequencies being tested (3 kHz, 4 kHz, and 6 kHz) for both ears. The plots show that the user has been measuring hearing for approximately three years, and that a total of 54 measurements have been performed during this period.

The lower plot in all six frames shows the CUSUM Q value from the statistical process control presented above, using the values $k = 0.5$, and $h = 6.1$. In three of the plots, a vertical stippled line can be observed. This shows the point at which the CUSUM Q value, S_i^+, crosses the control limit, indicating that the process is 'out of control'.

By calculating the difference between the mean value of the last five observations prior to the processes giving a signal, and the mean value of the observations after these five, it is possible to estimate the hearing threshold shifts detected. Table 3.4 shows a summary of the details of the process control.

Table 3.4: Estimation of hearing threshold shifts at test frequencies giving an 'out of control' signal. The 'Warning Signal' denotes the observation that triggers the control charts to provide the signal.

Ear	Freq.	Warning Signal	Post HL-value	Pre HL-value	Difference
Right	6 kHz	35	23.9 dB	21.0 dB	2.9 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

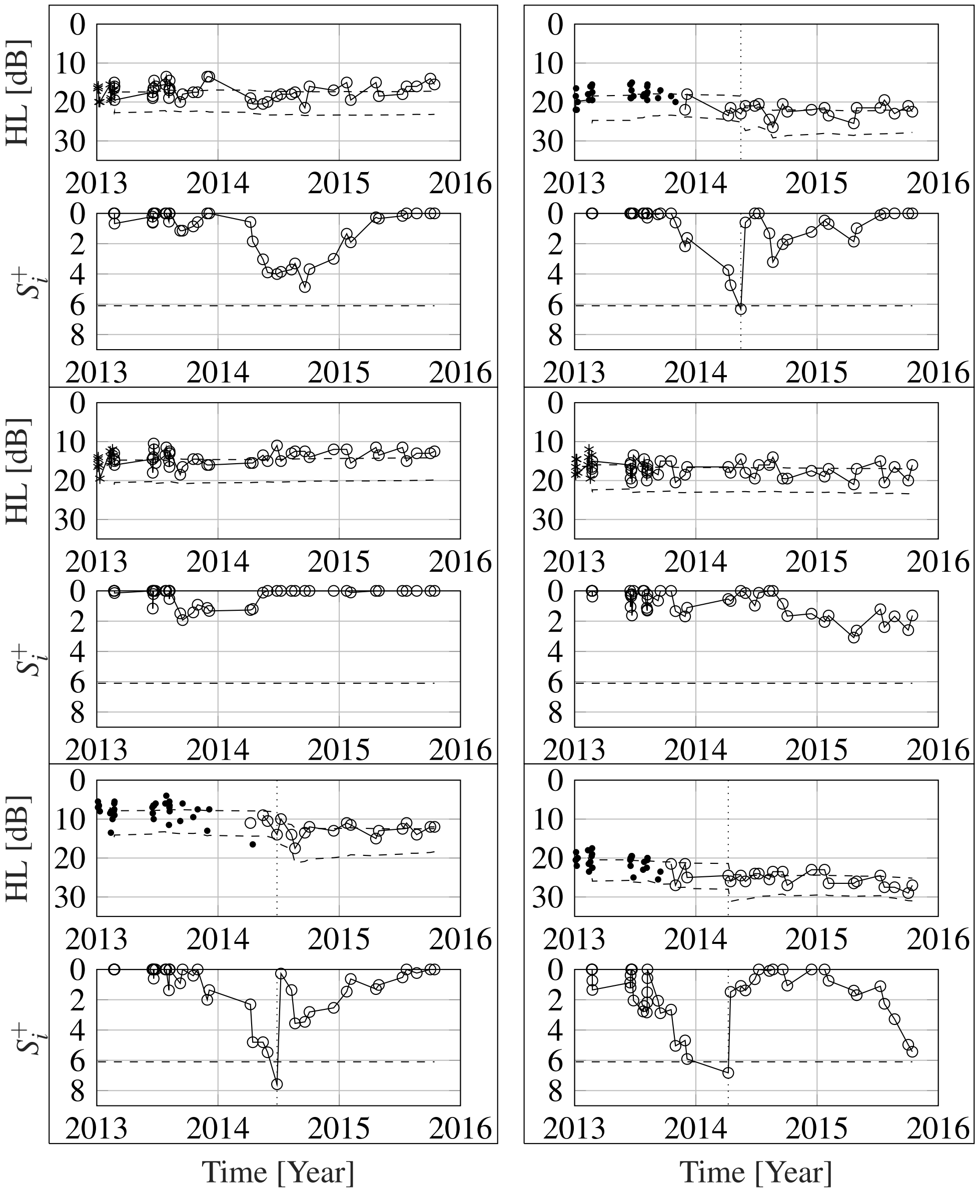

Figure 3.17: Plots of hearing level (HL) measurements (X chart), and statistical process control performed on these measurements (CUSUM Q chart), taken from a single individual performing regular measurements for almost three years. Plots on the left correspond to measurements of the left ear, while those on the right correspond to the right ear. The upper, middle and lower frames correspond to measurements at 3 kHz, 4 kHz, and 6 kHz, respectively.

3.4 Discussion

The process control proposed can be tuned to detect shifts in the hearing threshold that exceed a specified level, such as 5 dB. For individuals who exhibit consistent hearing test responses, i.e. with low variability, it may be possible to detect even smaller threshold shifts. The problem is that shifts smaller than 5 dB can be difficult for an audiologist to verify. Even if it is possible to use a smaller step size (1 or 2 dB) than the 5 dB level normally used in an audiometric test, the standard deviation for such tests is still approximately 5 dB (Jerlvall and Arlinger 1986). If statistical process control of hearing proves accurate in terms of detecting hearing threshold shifts, it may be possible to implement counteractions based on the outcomes displayed in control charts. However, more practical experience is required before a more certain conclusion can be reached.

For the process control to function as intended, many hearing measurements must be performed. This means that hearing measurements must be made readily available to the test subjects. This might require moving the testing process out of the OHSP offices and onto new platforms. Since the process control does not require calibrated input, it is possible to employ computer-based tests using off-the-shelf sound cards and headphones, or app-based hearing tests using smartphones or tablets. As long as the same equipment is used, and the background noise is under control, the method will detect changes in hearing threshold. Calibrated measurements can be recorded in the traditional way by the OHSP and used in the interpretation of results from the process control.

From Figure 3.9 it is possible to conclude that for the SS situation the run length distributions are far from normally distributed, or even symmetric. This means that the *ARL* values does not give a good description of how the control chart performs for this type of shift, for a large population. The run lengths for the RS, seen in Figure 3.10, are on the other hand symmetrically distributed. The *ARL* value is then the same as the most probable outcome for the different scenarios.

Another observation for the RS situation is that even if the run length increases as the gradient gets smaller (see Figure 3.10), the average accumulated hearing loss decreases (see Figure 3.11). 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 performing more observations after the onset of the ramp shift also would affect the estimation of the mean value, this benefit was not obvious.

The simulations of presbycusis showed that if one measures hearing more often than ten times a year, starting at the age of 18, it will be flagged as ‘out of control’ after about 10 – 15 years. This must be taken into account when considering a detected threshold shift. If the process control is initiated when the subject is older, the situation may change because the threshold shift gradient increases with age. This factor was not explored in more detail as part of the simulations.

It has also been shown that unless counteraction measures are implemented, outliers can have a detrimental effect on process control (see Figure 3.13). Thus, in the absence of counteractions, even as few as five large outliers among the first 200 observations (i.e. a 2.5 % level of erroneous data points) will render process control unable to detect hearing threshold shifts smaller than 10 – 12 dB. Use of the winsorising approach, as described in this paper, enables all large outliers to be detected, thus facilitating the performance of the control chart to be uncompromised.

The comparison between the method presented in this paper and the current regime showed that large shifts will be detected efficiently. Of particular importance is the fact that the reliability of the new method quickly approaches 100 % once more than 5 – 10 observations have been performed. This improves the sensitivity of the hearing monitoring. It also shows that it is not necessary to perform a large number of hearing measurements in order to outperform the current regime. However, fewer measurements will reduce the ability to detect small hearing threshold shifts.

If the hearing threshold shift progresses over time, as often is the case, this will affect the performance of the proposed method. Most challenging is large progression rates (i.e. rates over 5 dB/observation). The simulations revealed that if the rate of the shift is 10 dB/observation the hearing threshold shift is most difficult to detect with the CUSUM Q chart. The reason for this is that these large data points will be treated as outliers and not acted upon immediately. Such large progression will, however, most likely be detected by the X chart. When this happens it is advisable to encourage the test subject to perform more frequent tests, thus reducing the progression rate. It was also shown that a rate of less than 2 dB/observation is needed to reliably ($P > 0.9$) detect a 15 dB hearing threshold shift. This means that if the hearing monitoring system should be beneficial one must perform hearing measurements so frequent that the rate becomes smaller than 2 dB/observation.

The international standard, ISO 1999 (2013), states that the estimated progression of NIHL over time is greatest during the first ten years of exposure. Figure 1.2 shows the time development for four different sound exposure levels and clearly shows that the HL gradient is largest the first ten years. This anticipated progression of hearing must thus be taken into consideration when a control chart is constructed. Even the steepest slope, from 100 dB sound exposure, has a progression rate of less than 3 dB/Year, hence it should be possible to reliably detect such threshold shift.

Using the process control on real-world data demonstrated that small (≈3 dB) hearing threshold shifts can also be detected. However, detection is 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 an initial check that the signal is not a false alarm, and an investigation into whether the shift may be the result of natural causes, such as a common cold. If there is reason to believe that the threshold shift is noise-induced, appropriate counteractions must be considered to reduce the risk of further negative progression. Possible counteractions will be presented in Ch. 4.

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

3.5 Conclusion

This chapter has elucidated a potential monitoring scheme that can be used to detect small hearing threshold shifts. Monte Carlo simulations have demonstrated that it is possible to detect small step shifts in HL, and that the onset point is of some importance. Early onset entails that hearing monitoring has less time to provide an estimate of the actual values of the mean and standard deviation, and this is reflected in a decrease in the schemes ability to detect changes.

The detection of ramp shifts is not sensitive to onset time as similar levels of performance are observed for all the onset points simulated. Also, small ramp shift gradients are easier to detect if the total cumulative shift is used as a criterion for comparison. This means that when it comes to performing hearing tests – “the more the merrier”. It was also found that presbyacusis will be detected eventually, but that several years may pass before this happens.

Finally, the importance of outlier detection and counteractions has been demonstrated. Without high quality input data to the control charts, the monitoring scheme will be unreliable. This paper presents possible rules for detection and counteraction, and these were shown to perform well for large outliers.

Real-world data taken from an offshore installation shows that it is also possible to persuade individuals to perform frequent hearing measurements, provided that the test is made readily available. This will enable the proposed hearing monitoring regime to be used as an early warning indicator, and individually-tailored counteractions can be implemented if a hearing threshold shift commences. Such early warning will enable better protection for all individuals that are subject to loud sound exposure, including those who are more susceptible to high sound levels.

Chapter 4

Implementing the New Method

The strategy presented in Ch. 3 uses frequent hearing measurements and statistical process control to detect permanent threshold shifts that are smaller than 5 dB. This renders possible a completely new regime when it comes to early warning indicators, and can become an additional barrier against NIHL.

There are, however, several choices that have to be made before such process control can be put into action. The objective of this section is to give an overview of different topics that must be taken into account when implementing such a monitoring scheme. These topics have been divided into pre-, mid-, and post-topics, reflecting when they occur in the process control, in addition to some other aspects that must be deliberated. The pre-topics must be considered before the control chart is constructed, the mid-topics are relevant during the monitoring, and the post-topics are pertinent when an 'out of control' signal has been detected.

Figure 4.1 shows an example of what such a control chart might look like. The data have been taken from real hearing test data and show the HL from 51 measurements at 6 kHz from the right ear taken during a period of approximately half a year. No permanent hearing damage seems to have occurred during this period since none of the observations exceed the upper control limits (UCL).

The next section will discuss the different topics that are relevant for the hearing monitoring. Table 4.1 shows a summary of the topics and can be used as a check list when implementing the monitoring scheme.

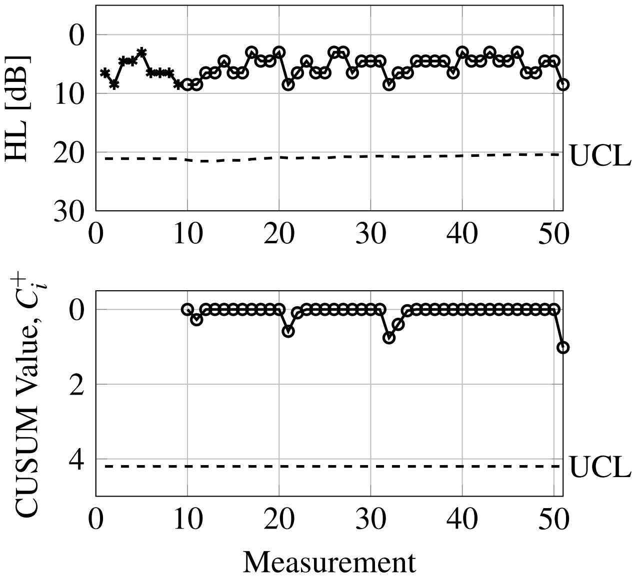

Figure 4.1: Example of a self-starting control chart. The dashed lines are the upper control limit (UCL) for the control charts. Upper: Hearing level (HL) measurements. Lower: CUSUM Q chart of the HL values. The nine first observations (marked with * in the upper plot) are used as initial input for the control chart. Therefore the CUSUM chart does not have values preceding measurement number 10. Notice that the y-axis is flipped to comply with audiograms.

Table 4.1: Summary of the different topics that must be considered before a hearing monitoring can be implemented.

Phase	**Topic**	**Description/Question to be answered**
Pre-topics	Test rate	How often will the test be performed?
	False alarm rate	What is an acceptable false alarm rate?
	Monitoring sensitivity	How small threshold shifts should be detected?
	Information and training	Persons who are going to be monitored must be given adequate information and training.
Mid-topics	Reliable data	It must be ensured that the input data to the hearing monitoring are reliable. Outlier detection and counteractions are important to make the hearing monitoring robust against large, unwanted data points (noise).
	Motivation	Different strategies to motivate the test persons to perform regular hearing tests must be considered.
	Visualizing data	What information should be presented to the test person during the monitoring?
	Misuse	Counteractions against misuse (e.g. addiction-like behaviour) of the hearing tests should be considered
Post-topics	Out-of-control alarms	Who should be notified when an out-of-control signal is given?
	Type of warning	How should the warning signal be given to avoid that the test person is unnecessarily worried?
	Follow-up strategy	What kind of follow-up and counteractions should be considered when an out-of-control signal is given?

4.1 Pre-Topics

First of all it is important to decide where the hearing tests should be performed and to consider what equipment should be used. A solution is to have dedicated test rooms where workers can go to perform an automatic hearing test. The background noise level must be taken into account when considering such dedicated rooms. The requirements will depend on the sound attenuating properties of the equipment used for the hearing test. Earlier, in Table 2.3, the attenuating properties of the very common Sennheiser HDA 200 headphones will be presented. The attenuation is around 30 dB to 40 dB, depending on the frequency. This means that the room does not necessarily have to be extremely quiet if a pair of good headphones are used. Background noise level monitoring could also be applied in such a room to detect episodes with higher background noise, such that hearing tests performed during increased noise can be marked as more dubious. By using pre-determined background noise limits as exclusion criteria, these points can either be excluded directly from the control chart, or be repeated to verify the response given.

Another possibility is to use mobile personal devices to perform the hearing tests. Smart-phones, tablets, dedicated devices, or HPDs can be possible test platforms. The advantage of such solutions is that they make the hearing test much more available. During the Next Step research project the hearing test has been embedded into the hearing protection device (HPD), making it easy to perform whenever the user wishes. Such mobile solutions will, however, give more stringent demands on the attenuating properties of the equipment, and background noise level monitoring should also be considered. For the HPD used in this project, the sound attenuating properties are verified by a fitting test after the person has inserted the earplugs. In addition, the sound level both outside and inside the earplug is measured continuously during use, so it is possible to measure the real ear attenuation. This means that it is also possible to measure if the background noise level is too high during a hearing test and a warning can be given, preferably before the test starts, if the background noise level is too high. It is also, as mentioned above, possible to prompt the test subject to repeat the test if needed.

Next, one must decide which hearing test to use. Even if it is possible to use any kind of hearing threshold measurement as input to the process control, it will be beneficial to use an efficient method. Additionally it will be important that the same method is used for all measurements, or that any differences between methods are quantified and adjusted for. Otherwise one might end up detecting methodological differences instead of hearing threshold shifts. The NEWT method is an example of such an efficient test where the hearing threshold is reliably found after only six sound stimuli per frequency (Vinay, Svensson et al. 2015).

Before the statistical process control can start one must also make two main decisions; what false alarm rate can be tolerated and how small shifts in hearing threshold should be detected. These choices are contradictory, but different adjustments can be applied to find a good compromise as discussed in Sec. 3.2.1.

The first choice is linked to how often hearing measurements are performed. The false alarm rate is given as the number of observations between each false alarm. If the number is 50 this means that, on average, the control chart will give a false alarm once every 50 observation. How often this will be, with regards to time, depends on how often measurements are performed. If, for instance, daily measurements are performed, there will be false alarms approximately every second month, while if you perform only one measurement per month, approximately four years will go between each false alarm. It is therefore necessary to decide how often hearing measurements should be performed before the control limit can be set.

Regarding the rate of hearing tests it would be advisable to divide the workers into different groups, depending on their sound exposure. Those with high sound exposure should test their hearing more often than those working in more quiet areas. It will also be possible to monitor those who seem to be more susceptible to noise more closely. These individuals will be detected by the process control itself, thus the test rate might have to be adjusted during the monitoring. The test rate is shown as Δt in Figure 1.3 on page 8.

The size of the threshold shifts that can be detected will, at least if the proposed self-starting control chart is used, be dependent on the variability of the individual. Since the self-starting control charts normalize the measurements and use mutual control limits, a small variability will lead to an increased ability to detect small changes. It is, however, possible to adjust the sensitivity to some extent by changing the allowance value, as mentioned in Sec. 3.2.1. Often an allowance value, k, of 0.5 or 0.75 is used, which means that measurements that are more than 0.5 or 0.75 standard deviations from the estimated expectancy value will contribute to the CUSUM value. The standard deviation of the 6 kHz measurements presented in Figure 4.1 is 1.7 dB, which means that changes of only 0.9 dB ($k = 0.5$) or 1.3 dB ($k = 0.75$) will accumulate and eventually be detected. Assuming a standard deviation of 5 dB means that it is possible to detect hearing threshold shifts smaller than 2.5 dB ($k = 0.5$) or 3.8 dB ($k = 0.75$).

How quickly the change will be detected depends mainly on the false alarm rate. If one wants to keep the false alarm rate low, then one has to accept that the sensitivity to small hearing changes will be reduced. The estimates above do, however, indicate that small changes (< 5 dB), can be detected in the general population, and that much smaller (around 1 dB) can be detected in extreme cases.

Even if it is possible to start the self-starting process control already from the second observation, it is advisable to use a few measurements as initial start-up values as shown in Figure 4.1. The reason is that outliers during this phase will have a detrimental effect on the estimation of the parameters, and thus will change the expected behaviour of the control chart. One should also try to get representative measurements from the process during this phase with expected value and natural variation. This can, for instance, be done during a training session where the system and hearing test are presented for the workers. Such training sessions are also important to make sure that the workers are introduced to the monitoring regime and that they understand what they should do, how they should do it, and not least, why they should do it. If they can understand the importance it is likely that they will start spending time on this.

4.2 Mid-Topics

When the process control has started, some additional issues need to be considered. First of all it is important that the data that goes into the process control is reliable and representative for the process. The term ‘rubbish in gives rubbish out’ is definitely valid for SPC as well, and good strategies to make sure that the input data is of high quality are important.

Initially it must be ensured that the input from the hearing test is valid. As one example, the hearing test that is embedded in the HPD used in the research project Next Step (see Sec. 2.4) flags if a test was not completed, if a person did not respond, or if a person responded to all test signals (i.e. the test person heard the faintest test tones and the hearing test cannot estimate the HL). Either way, the test results cannot be trusted and should not be used in the SPC. Similar approaches can be implemented in other systems as well, and should be considered. One might also open up for the possibility to let the test subject exclude measurements from the SPC themselves. If something happens during the test, e.g. someone interrupts the test, an increased background sound level, or technical problems, occurred, it can be beneficial to exclude these measurements from the SPC. It is, however, somewhat problematic to let test subjects exclude measurements themselves. This can lead to exclusion of correctly performed measurements that the test subject is not satisfied with. Such ‘cherry-picking’ would deteriorate the reliability of the SPC. Another possible solution would be to let the test subject mark the suspect measurements and leave the exclusion to the occupational health service provider (OHSP) if needed.

Even if a thorough procedure to ensure high quality input data is implemented, outliers will occur. The winsorizing approach presented in Sec. 3.2.3 replaces suspect Q values in the CUSUM Q chart by the first ‘non-suspect’ value. The proposed control chart uses a winsorizing value of 3. This means that any observation that is more than 3 standard deviations away from the estimated mean value will be treated as an outlier and given the value 3. This way of treating suspect data points will ensure that large shifts that are real are detected quickly. Values outside the limit will also be excluded from the data used to find the estimates of the mean value and variance. Such exclusion will prevent large outliers to bias the estimates, which would otherwise seriously deteriorate the performance of the control chart. As mentioned in Sec. 3.2.3, other methods exist, but these are not discussed further here.

Secondly, it will be important to ensure that the workers perform regular hearing measurements. This is typically a question of motivation. Within the field of motivational factors one often use two basic distinctions; intrinsic and extrinsic. Intrinsic motivation refers to doing something because it is inherently interesting or enjoyable, while extrinsic refers to doing something because it leads to a favourable outcome, an external reward (Ryan and Deci 2000). Since few people experience a hearing test as interesting or fun by itself, i.e. intrinsic motivating, it is important that all the workers understand the reason why the measurements must be performed. If such education is done thoroughly the workers will realize that performing measurements can be in their interest since it can prevent future hearing loss, and that they can become intrinsically motivated by this knowledge. It is, however, most likely necessary to also use extrinsic motivational factors such as rewards, especially in the beginning of the process control, so that the workers can get the hearing measurements ‘under the skin’. This way of facilitating intrinsic motivation is a known technique that can be useful to get a person interested in doing a task, but it must also be done with caution since external rewards can undermine an existing intrinsic motivation (Deci et al. 2001).

Thirdly, it must be decided what information should be presented to the test persons when they view their data. This is related to the previous topic since a good visual presentation also could increase the motivation to perform measurements. In a world with more and more biosensors and smart-phone health apps, the amount of information about our health is exploding. An EU funded project recently found that 44 % of the people using health apps want health information about themselves, 33 % want this to support a better lifestyle, and 46 % want to track and monitor their symptoms in order to benchmark their progress (PatientView 2015). This supports a regular monitoring of the hearing, and people at risk, i.e. those working in noisy environments, could be motivated to perform the necessary hearing measurements.

A possible concern about the hearing monitoring is that some workers might become overly focused on their hearing. Since hearing tests must be readily available, it is possible to become 'addicted' to tests, something that would affect the productivity of the worker and possibly lead to a nocebo effect[a]. Such 'abuse' can, however, be detected easily since the measurements can be seen in the person's measurement history and manual or automatic detection of very frequent measurements can be used.

4.3 Post-Topics

The post topics consider the issues connected to the detection of a hearing threshold shift. First of all it must be decided who should be notified when a hearing threshold shift is detected. This could depend on the size of the change. If a large shift has been detected the probability is large that this is not a permanent threshold shift (PTS), but rather a faulty measurement or a temporary threshold shift (TTS). It would therefore be recommended that the test person is informed about the large shift and that he/she is encouraged to perform a re-test to see if it the shift remains. If the change still is measured at the second test, there is still a probability that it is a TTS, and the person can be advised to perform new measurements more frequently the next few days, preferably when the hearing is recovered. Such a feature can also be automated, i.e. the test program would notify the worker when new hearing

---

[a] The nocebo effect is when a negative expectation of a phenomenon causes it to have a more negative effect than it otherwise would. A nocebo effect causes the perception that the phenomenon will have a negative outcome to actively influence the result. (Wikipedia 2017)

tests should be performed.

Occupational health service providers (OHSP) should also be notified when large shifts occur. If these large changes happen regularly, it can be an indicator that the worker does not perform the test correctly, or that the person has frequent TTS. Both situations are important, but must be treated in two different ways. The first can be corrected by personal communication with the worker and/or additional training, while the latter must be treated as a possible noise exposure issue. Even if it has proven difficult to find a relationship between TTS and PTS, they definitely share a common cause; loud sound exposures (Melnick 1991). Previous studies also indicate that TTS can be more problematic than first expected. Permanent neural damages have been reported in animals, affecting supra-threshold sound levels, even if the hearing thresholds have returned to normal (Kujawa and Liberman 2009; Plack et al. 2014). This strengthens the importance of acting upon frequent TTS and to implement relevant counteractions for these as well.

Alarm signals that indicate small changes in hearing threshold, i.e. those detected by the CUSUM Q chart, should probably not be sent to the workers, only to the OHSP. The reason for this is that the interpretation of the results can be more difficult, and that these alarms should always be taken seriously.

When an alarm signal has been given a multi-step procedure must be started.

First of all it must be decided if this looks like a false alarm or not. Looking at the other test frequencies and the time course of the CUSUM values can give valuable information. If similar negative trends can be seen in some of the other frequencies, this will support that this is an actual damage. However, if none of the results for the other frequencies looks suspicious, it will be advisable to let the process control continue for a few more measurements. The worker should then be urged to perform more frequent hearing tests in the following period to get more information quickly. It is also important to reset the CUSUM Q chart after the signal such that new warning signals can be triggered.

If it is likely that there is an actual change in hearing, the next step is to perform an anamnesis. This will find out whether the change has a natural explanation, e.g. presbyacusis or a common cold, or if there are other causes, e.g. high noise exposures. If the latter is concluded, individual counteractions must be considered. For natural explanations it might be sufficient to ‘re-start’ the control chart, meaning that the CUSUM value is set to zero. This must, however, be done with caution to prevent the hearing monitoring to become unnecessarily insensitive.

It is also important that the warning messages are delivered in an appropriate way, especially since false alarms inevitably will occur. Otherwise the test person (worker) can be worried without any reason. To prevent unnecessary concern it would also be advisable to emphasize that the workers should contact the OHSP if they feel the need to discuss the results or other things about the hearing monitoring.

4.4 Other Aspects

The hearing monitoring also brings some ethical concerns. Similar issues, related to genetic testing, have been reviewed in the last decades (Grandjean and Sorsa 1996; Serra et al. 2007; Fisher and Harrington McCarthy 2013). The concern is that by finding persons at higher risk for certain diseases or those who are more susceptible for certain exposure, we will make a society with ‘selection of the fittest’. There are, however, some major differences between genetic testing and hearing monitoring. One of the ‘problems’ with genetic tests is that they are predictive. This means the test can be used to find persons who have an increased risk of getting a certain disease, before it has occurred. For the hearing monitoring the situation is different. When a warning signal is given from the process control, the hearing is already damaged. This means that the process control is not predictive, but unprophetic or reactive. This further means that the genetic test can be used prior to an employment, where the employer uses such genetic screening to exclude the ‘weakest’ job applicants. The hearing monitoring is, on the other hand, a process that is performed during the employment, thus it cannot be used directly to exclude job applicants. If a worker who already has used the hearing monitoring is applying for a new job, it is, however, possible to get information about the worker’s susceptibility to noise by looking at historical data. Similar aspects have also been discussed for genetic test results, and it has been suggested that the worker is the owner of the results and should decide how these are used, and that it must be allowed to withhold information likely to prove detrimental to one’s self-interest (Grandjean and Sorsa 1996). However, knowledge about the workers susceptibility to noise can also be used constructively to improve the safety of the working staff.

In genetic testing there is also an issue whether or not people want to know if they are at higher risk for certain diseases. Such information can affect the life

Table 4.2: Definitions of preventive actions. Adapted from Institute for Work and Health (2015).

Name	**Description**
Primary prevention	Aims to prevent disease or injury before it ever occurs. This is done by preventing exposures to hazards that cause disease or injury, altering unhealthy or unsafe behaviours that can lead to disease or injury, and increasing resistance to disease or injury should exposure occur.
Secondary prevention	Aims to reduce the impact of a disease or injury that has already occurred. This is done by detecting and treating disease or injury as soon as possible to halt or slow its progress, encouraging personal strategies to prevent reinjury or recurrence, and implementing programs to return people to their original health and function to prevent long-term problems.
Tertiary prevention	Aims to soften the impact of an ongoing illness or injury that has lasting effects. This is done by helping people manage long-term, often-complex health problems and injuries (e.g. chronic diseases, permanent impairments) in order to improve as much as possible their ability to function, their quality of life and their life expectancy.

of not only the tested person, but also the family who might inherit the same genetic trait. A person's knowledge about her/his susceptibility to noise will possibly affect the life of the affected, but the interventions are relatively easy to carry out, e.g. using hearing protection device. Even if there are indications that noise susceptibility can be hereditary (Sliwinska-Kowalska and Pawelczyk 2013), it is unlikely that knowledge about a family member's NIHL will affect the life of the descendants.

Actions to prevent occupational damage is often divided into primary, secondary and tertiary prevention. The definitions from the Institute for Work and Health (2015) are given in Table 4.2. It must be emphasized that the hearing monitoring proposed in this paper is a secondary preventive action. In other words, it is not intended to replace any of the existing primary preventive barriers against NIHL. The current primary prevention regime in Norway is made of three actions, presented chronologically below;

1. Reduction/removal of the noise/sound source
2. Adjustment of the workplace (e.g. sound absorption in the room or reduced work time in the noisy area)
3. Using personal safety equipment (e.g. hearing protection devices)

These steps are important, and especially the first point is effective for all exposed workers. In the Norwegian act it is, however, stated that if a hearing damage is detected, this shows that the primary preventive actions are not effective enough, and that a new risk assessment should be carried out (The Norwegian Labour Inspection Authority 2013). A problem with the current regime is that the detection of hearing damage requires large hearing threshold shifts, i.e. more than 15 dB, for the specialists to conclude that a damage has occurred (see Sec. 3.1.2 for a discussion). By implementing the proposed hearing monitoring regime it is possible to add another criterion when to perform a new risk assessment. This can make the secondary preventive action much more effective.

Another concern can be mentioned as an extension of this topic. One might end up with a situation where the only available counteraction for a noise damage employee is reduced work time or relocation. An identical problem is already mentioned in the Norwegian guideline where it is stated that relocation of an employee can be a possible counteraction if a hearing damage is detected (The Norwegian Labour Inspection Authority 2013). Furthermore, this leads to an ethical question; should an employer be allowed to adjust the work time for each individual, based on the noise susceptibility? The utmost consequence is that workers can be discriminated on basis of their hearing. There is not an easy answer to this question, but it is important that these questions are considered before the monitoring starts. The proposed process control will then be a secondary preventive action to monitor the hearing such that the risk of further negative development is minimized.

4.5 Conclusion

In this chapter different aspects around a hearing monitoring scheme have been discussed, both technical choices that have to be made and considerations on a more superior level. The hearing monitoring presented can improve the preventive actions against noise-induced hearing loss by giving early warnings when a damage is in progress. Further this can be used to implement counteractions on an individual basis, possibly stopping the negative development of the hearing. This can be especially important for those persons that are more susceptible to noise who are disregarded by the current noise legislation with common noise limits. It is accepted that these individuals are at high risk of developing hearing loss, even if the noise limits are met.

Chapter 5

Real-World Data Examples

This chapter presents data from automated hearing measurements carried out on an offshore installation. The data will demonstrate several important aspects of the proposed process control. Both hearing measurements and exposure are presented and an example of how these data can be utilized together is presented as well.

During the Next Step research project 18 workers performed one or more measurement with the hearing test embedded in the hearing protection device. These workers were recruited from ‘high noise areas’ on the offshore installations, most of them working at the helideck where they dispatched helicopters. The measurements were carried out during a period of twenty months, and the participants were asked to perform the test regularly, but with no specific requirements. As a result, the participants did very varying numbers of tests, as described below. In each test, 3 frequencies were tested (3 kHz, 4 kHz, and 6 kHz), for both ears, and the test took approx. 2 min to finish. The testing was presumably done in different background noise situations, but this cannot be verified since no noise measurements were done during the tests.

5.1 Standard deviation

As discussed earlier in this thesis, the standard deviation in the hearing tests performed is an important value when it comes to hearing monitoring. Earlier it was decided to estimate the individual standard deviation in the control charts instead of using a common value. To assess this decision a statistical analysis of the measured standard deviations was performed.

It was chosen to exclude workers with less than 5 hearing measurements from the analysis. This reduced the data set to 11 workers. The hearing levels (HL) were given as input to a Levene's variance test using frequency (3 kHz, 4 kHz, and 6 kHz), side (right and left) and person ID as group factors. Extreme values, i.e. HL measurements at 60 dB or -20 dB, were excluded since they do not represent credible hearing values (see Sec. 2.4 for details).

Table 5.1 shows the results from a homogeneity of variance test. Since the hearing level measurements showed a non-normal distribution (not shown) a Levene's test statistic was used to analyse the data (Levene 1960). The hearing measurements for each frequency and ear have been normalized, i.e. adjusted to have a zero mean value. This is done because it is assumed that the standard deviations are independent of the hearing level, and the analysis will give a better impression of the variability of the measurements across individuals.

As can be seen in the table, neither frequency nor side have any significant effect. Person ID, however, is a highly significant factor, meaning that there are individual differences in variability of the hearing measurements. This supports the decision on using the individual estimates in the hearing monitoring.

It is also possible to see that the standard deviations are in the range 2.8 dB – 12.0 dB, median value 6.1 dB (see Table 5.1 and Figure 5.1). This is somewhat higher than previously proposed, but one should notice that no systematic training was given to the test subjects.

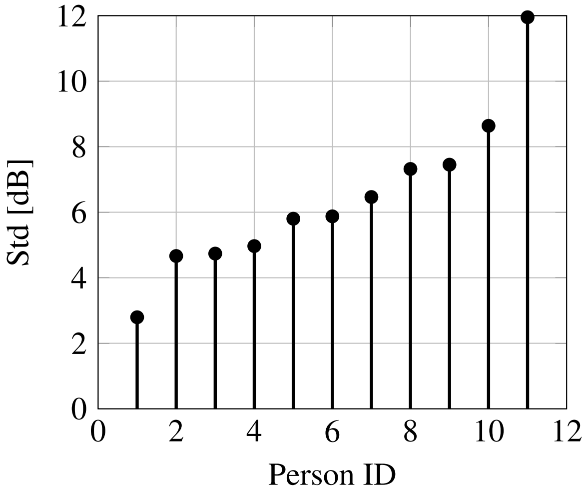

Figure 5.1: Distribution of standard deviation estimates for each individual.

Table 5.1: Levene's variance test of the hearing measurements from 11 workers on an offshore installation. The input data were all the HL measurements.

	Group	Count	Mean	Std Dev
Frequency	**3 kHz**	347	0	6.21085
	4 kHz	349	0	6.49157
	6 kHz	345	0	6.44306
	Pooled	1041	0	6.3831
	Levene's statistic (absolute):		0.476	
	Degrees of freedom:		$2, 1038$	
	p-value:		0.621	
Side	**Left**	522	0	6.35211
	Right	519	0	6.40798
	Pooled	1041	0	6.38002
	Levene's statistic (absolute):		0.363	
	Degrees of freedom:		$1, 1039$	
	p-value:		0.547	
Person ID	**1**	98	0	2.7945
	2	54	0	4.664
	3	174	0	4.7375
	4	79	0	4.9666
	5	114	0	5.8017
	6	180	0	6.4642
	7	54	0	5.8748
	8	72	0	7.3207
	9	73	0	7.4516
	10	96	0	8.638
	11	47	0	11.952
	Pooled	1041	0	6.4078
	Levene's statistic (absolute):		8.385	
	Degrees of freedom:		$10, 1030$	
	p-value:		<0.0001	

5.2 Single Person Example

The proposed hearing monitoring regime was applied to all the N workers with more than 10 hearing tests in a postprocessing simulation. For one of the*N* workers, an ‘out of control’ signal resulted, once. This section will show a case study of this individual, elucidating several aspects around the hearing monitoring. One should be aware that this is a single result from one person, and that care must be taken when drawing conclusions.

In Figure 5.2 the initial value of the persons HL can be seen. The HL is estimated by calculating the mean value of the 10 first hearing measurements. The person is a male and he was approximately 50 years old during the test period, thus his hearing was well within what is considered ‘normal’ for his age (see Figure 3.7). Since the HL is not >25 dB for any frequency, and since the average (across frequencies) is not >20 dB, the person is not considered hearing impaired according to The Norwegian Labour Inspection Authority (2013).

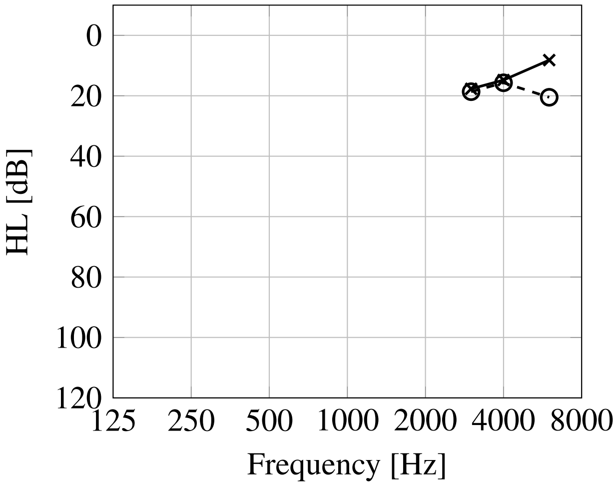

Figure 5.2: Audiogram for the person in the example. The HL for the three test frequencies is estimated using the mean of the 10 first observations in the process control.

5.2.1 Complete Time Series

The example person has performed hearing tests regularly for approximately three years, conducting a total of 54 measurements. Figure 5.3 shows the time series for the three frequencies at both ears. The person performed 34 measurements in 2013, 11 in 2014, and 9 in 2015. No systematic effort was made to try to increase the number of measurements during the time period.

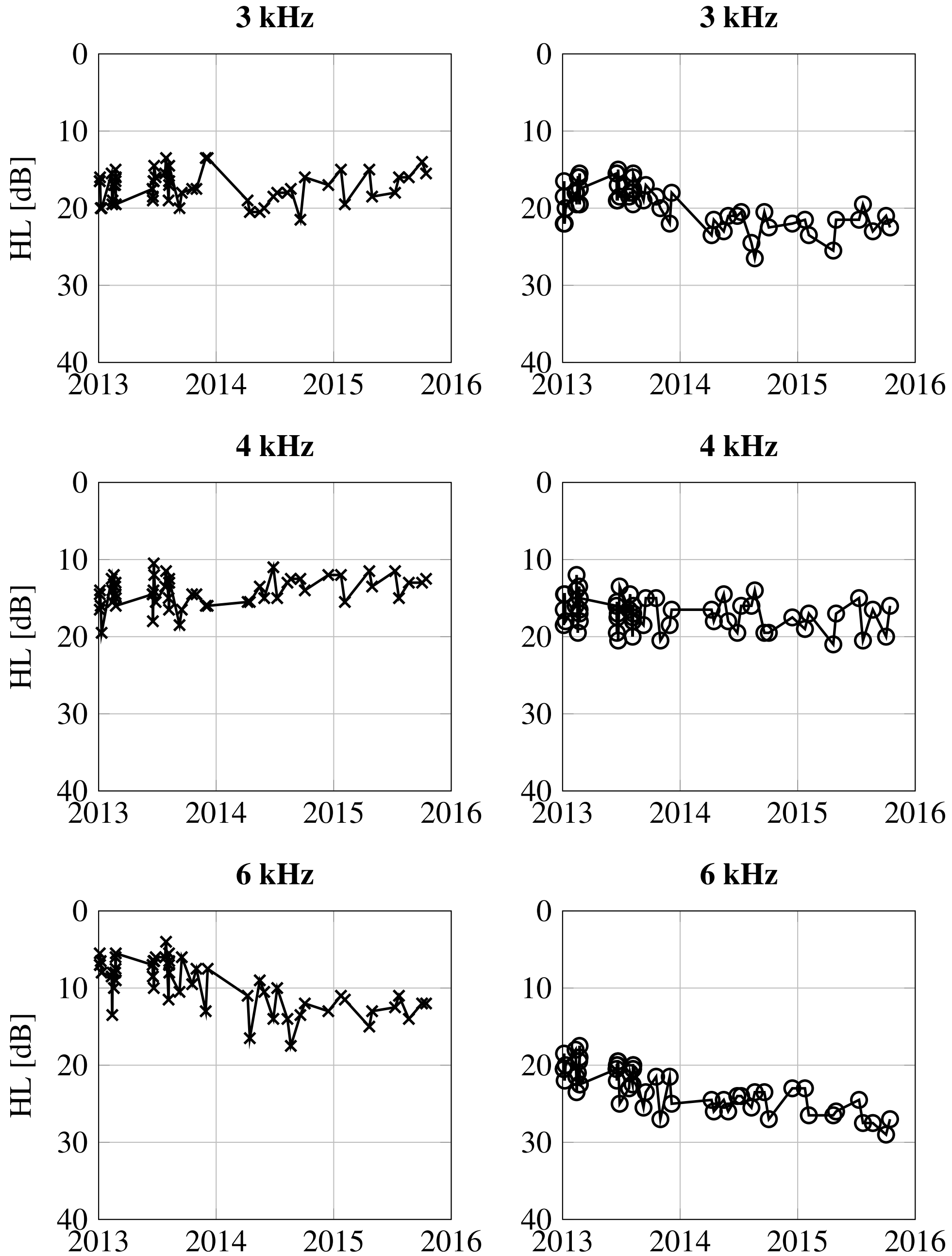

Figure 5.3: Illustration of the time series of the hearing level (HL) measurements for the three test frequencies for the left (left panel) and right (right panel) ear.

Table 5.2 shows the standard deviation estimation for the data in the time series for each frequency and ear. As can be seen, the standard deviation (Std) is in the range 2.0 dB to 3.6 dB. Additionally the two-span moving range estimates of the standard deviation is shown (see Sec. 3.2.2 for details). These are more robust against outliers and show that the standard deviation might be even smaller (around 2 dB). This means that it can be possible to detect very small hearing threshold shifts.

Table 5.2: Standard deviation estimation for the person in the example. $\widetilde{mR}$: median of the two-span moving range. $\overline{mR}$: mean of the two-span moving range.

Ear	**Frequency** [kHz]	**Std** [dB]	$\mathbf{0.886\widetilde{mR}}$ [dB]	$\mathbf{1.047\overline{mR}}$ [dB]
Left	3	2.4	1.7	1.6
Left	4	2.0	1.9	2.1
Left	6	3.6	2.3	2.1
Right	3	2.7	2.1	2.1
Right	4	2.2	2.3	2.1
Right	6	3.1	1.7	2.1
Mean		2.7	2.0	2.0

5.2.2 Hearing Monitoring

Figures 5.4 to 5.9 show the results from applying the SPC on the hearing measurements shown in Figure 5.3. Only the first 35 observations are shown since this is the point where the monitoring of 6 kHz at the right ear signals an ‘out of control’ situation. After such a signal is given an anamnesis should be performed, including an assessment of whether it was a false alarm or not.

Table 5.3 gives a description of the notation used in Figures 5.4 to 5.9.

Table 5.3: Description of data presented in Figures 5.4 to 5.9.

Plot	Description
Upper	*General*: Hearing level (HL) measurements as function of time. *Asterisks*: Measurements used as start-up values for the SPC (10 observations). *Circles*: Measurements used as input to the on-going process control. *Dots* (only Figure 5.9): Measurements excluded from the process control due to a resetting when the out-of-control signal is given. See Ch. 4 for details. *Dotted line*: Running average continuously updated. *Dashed line*: 3 σ control limit (UCL).
Lower	*General*: Self-starting CUSUM control chart of the HL measurements in the upper plot. *Dots*: Self-starting CUSUM values. *Dashed line*: Upper control limit ($h = 6.1$)

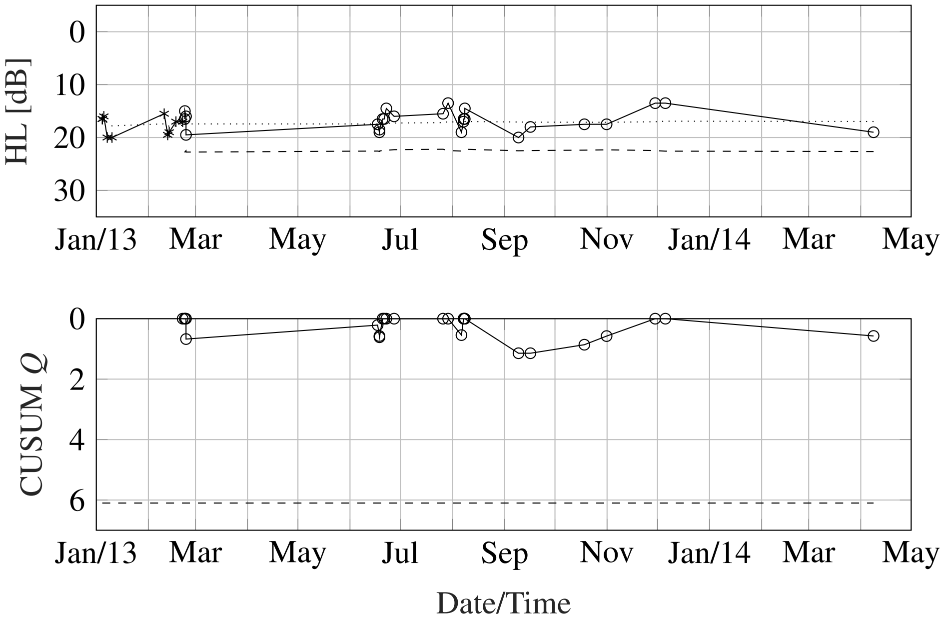

Figure 5.4: Hearing monitoring of 3 kHz at the left ear.

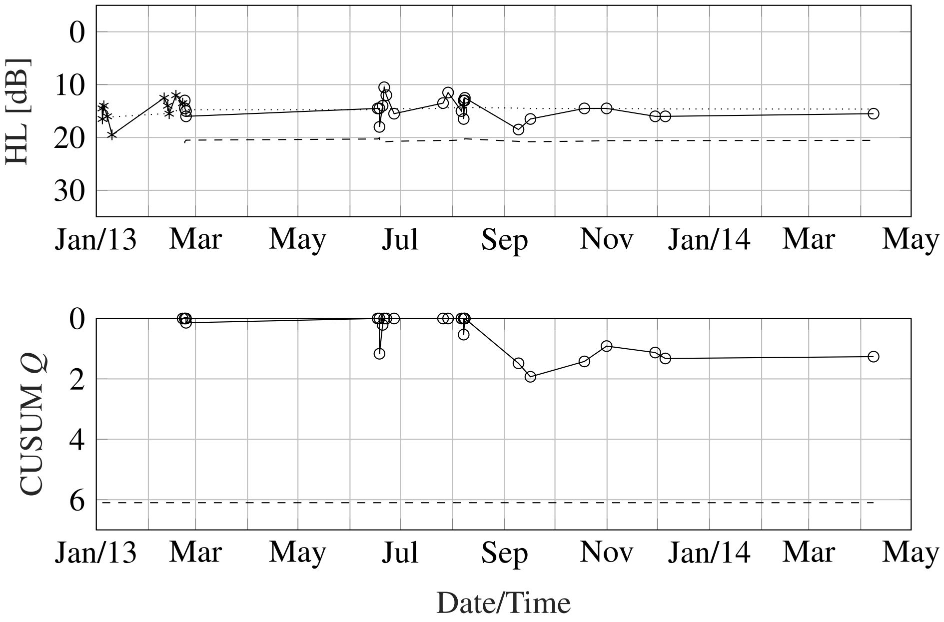

Figure 5.5: Hearing monitoring of 4 kHz at the left ear.

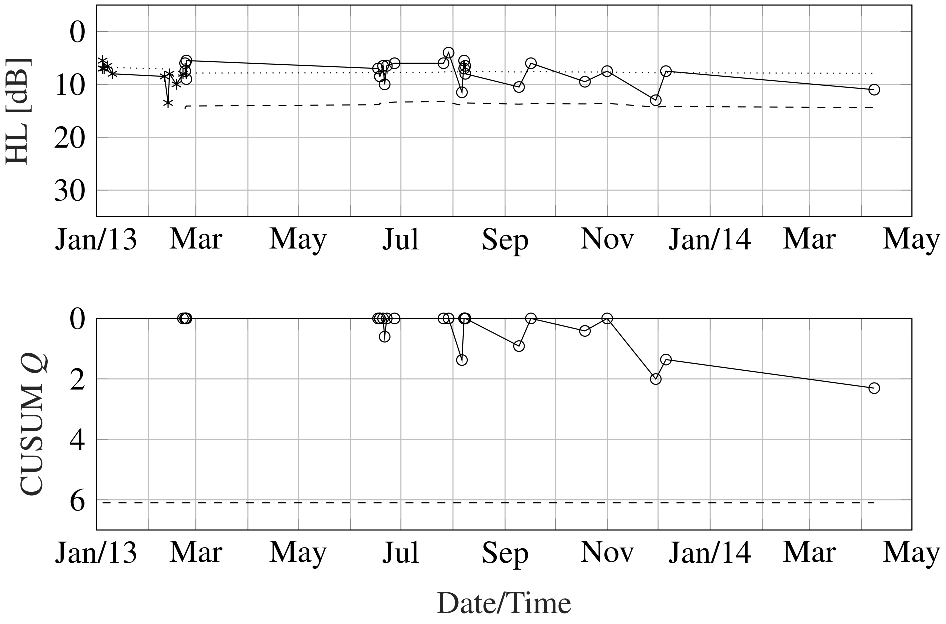

Figure 5.6: Hearing monitoring of 6 kHz at the left ear.

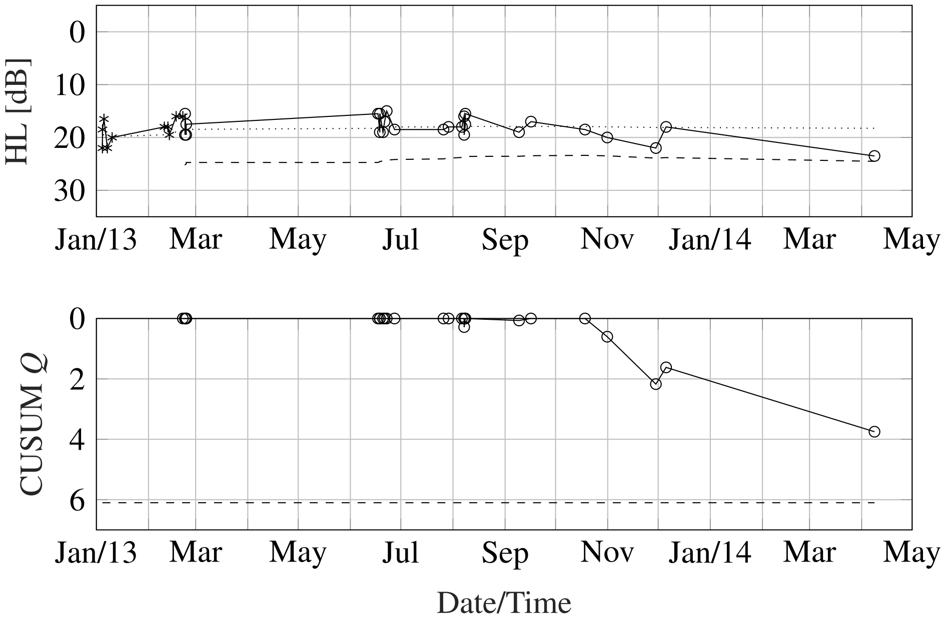

Figure 5.7: Hearing monitoring of 3 kHz at the right ear.

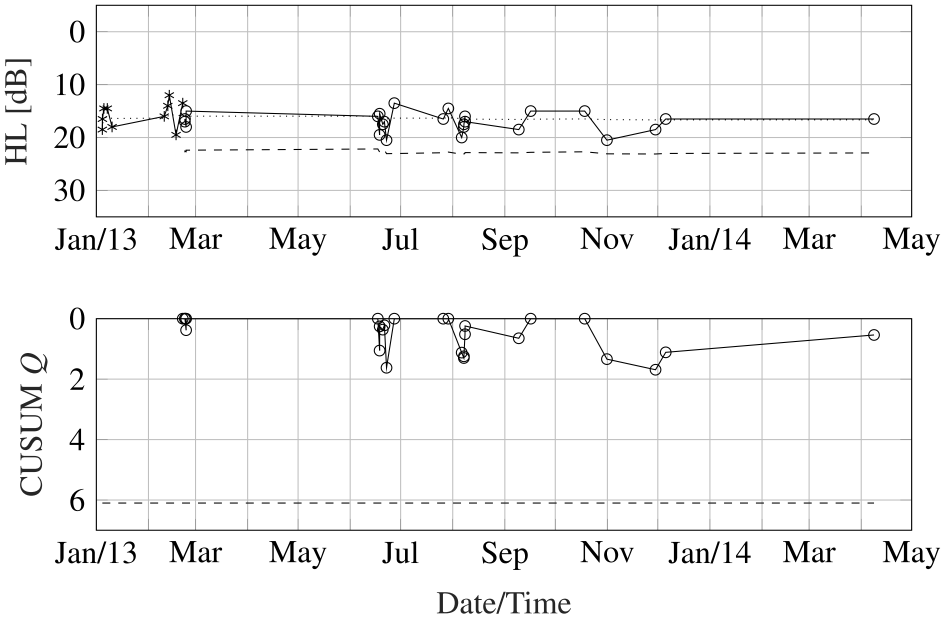

Figure 5.8: Hearing monitoring of 4 kHz at the right ear.

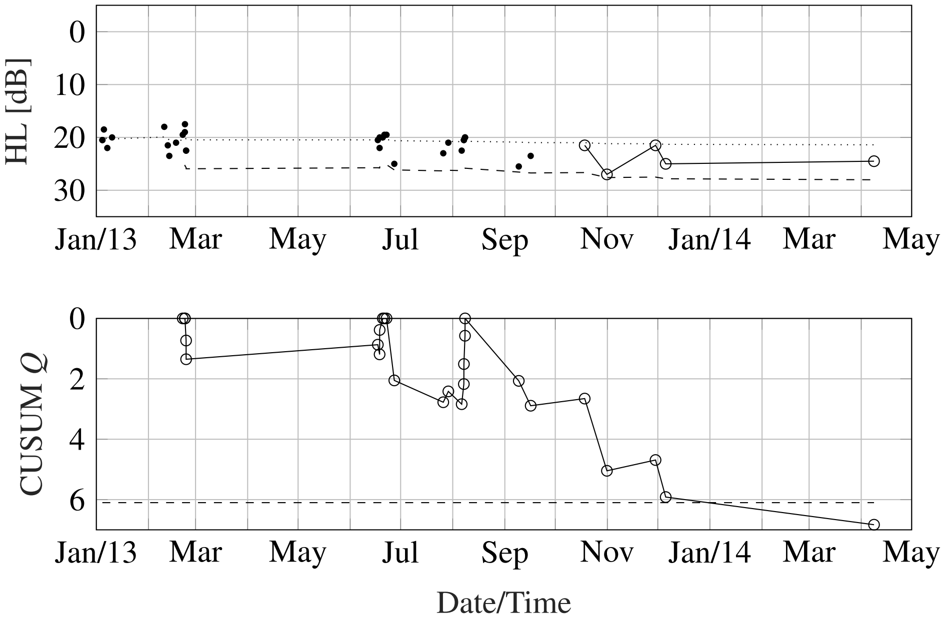

Figure 5.9: Hearing monitoring of 6 kHz at the right ear. Notice that when the ‘out of control’ signal is given, the control chart automatically resets. This is shown by the black dots without connecting lines. The new ‘baseline’ is defined by the last five measurements.

5.2.3 Anamnesis

A possible procedure to assess whether this is a false alarm or not is to inspect the results for the other test frequencies. Even if it is possible to have a damage at only one frequency and ear, an inspection of the CUSUM values for the other frequencies/ears can be performed to see if similar trends can be found there (as mentioned in Sec. 4.3).

The CUSUM Q chart makes it possible to estimate when the hearing threshold shift most likely occurred, as described in Sec. 3.2.3. Looking at the CUSUM values in Figure 5.9 it is possible to see that the last point where the value was zero was at the last measurement performed in August. Keeping this point in mind while inspecting the other control charts we can see that the CUSUM value does in fact increase for both 3, 4, and 6 kHz in the left ear (although the 6 kHz value only consist of an increase at one observation), and possibly for 4 kHz in the right ear. The CUSUM value at 3 kHz in the right ear increases, but the increase is slightly delayed. Based on this information it is reasonable to conclude that it does not seem to be a false alarm.

Next, one might take a closer look at the HL measurements at the control chart that gave the signal (see Figure 5.10). A visual inspection of the data reveals a possible linear trend in the hearing levels. The figure therefore also shows two linear interpolations; dashed line: from measurement 10, and dash-dot: from measurement 24. The latter is chosen because this is the point where the CUSUM estimates that the process went out of control. Such estimation will, however, give a starting point that is later than the real one for a progressive shift. The reason is that the allowance value, k, will 'mask' the small shifts early in the progression.

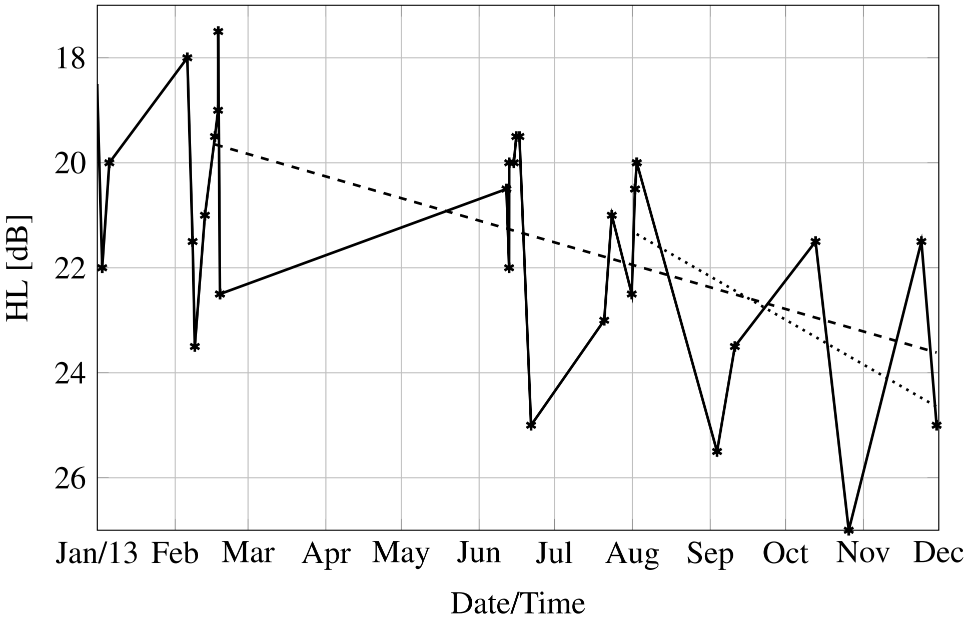

Figure 5.10: A detailed part of the hearing measurements from 6 kHz at the right ear seen in Figure 5.9. The plot also includes two linear interpolations discussed in the text.

Looking at the figure it seems like the progressive shift could have started during the first three months of the year. Linear lines were therefore fitted to the data set with different starting points, from measurement 1 to 14 (i.e. from January to March). The best fit ($R^2_{\text{adj}} = 0.2319$) was on observation 10. This fit was also better than the one from measurement 24 ($R^2_{\text{adj}} = 0.2163$).

Thereafter, it might be reasonable to assess if the change has any natural cause, e.g. if it is age- or sickness-related. Ear infections, common cold, etc. should ideally be taken into account, but since no health information were collected in the project only age will be considered here.

Since ARHL is a progressive loss it might be reasonable to assume that the shift could be age-related. The ISO-standard (ISO 7029 2000) estimating hearing levels as function of age and gender (see Figure 3.7) estimates that the hearing shift rate at 6 kHz is approximately 1.2 dB/year around the age of 50. The best fit to the example data has a rate of approximately 5.0 dB/year, hence age does not seem to be the only cause of the shift.

The final step presented here is the analysis of the sound exposure. Figure 5.11 shows both the one-minute equivalent SPLs inside the HPD and the corresponding accumulated dose. The dose is calculated using 16 h as integration window, and is given in percent of the allowed sound exposure for Statoil employees ($L_{p,\mathrm{A},12\,\mathrm{h}} = 80\,\mathrm{dB}$) which is approximately 3 dB lower than the limit given by the Norwegian Working Environment Act ($L_{p,\mathrm{A},8\,\mathrm{h}} = 85\,\mathrm{dB}$). This also means that a 200 % dose is considered acceptable by the Norwegian Working Environment Act.

As can be seen in the figure there were two episodes where the sound exposure was relatively high. In August 2012 the right ear dose was about 100 %, and in October the same year the dose was just above 200 %. Unfortunately the hearing measurements did not start before January 2013, hence it is difficult to conclude that the exposure is the reason for the hearing threshold shift. The sound dose from January 2013 has, however, been relatively low (<25 %), and should not be the reason for the shift.

One should also take a look at the one-minute equivalent levels for the sound exposure. The reason is that the SPL can be high even if the dose is low, if the duration of the exposure is short. Even if the sound dose is the prevailing risk factor for NIHL, a high SPL can also have an effect on the hearing.

Looking at the first high-dose episode one can see that the right ear has two minutes with $L_{p,\mathrm{A},1\,\mathrm{min}} \approx 105\,\mathrm{dB}$ and two minutes around 95 dB. These four minutes are the reason why the dose got so high.

The second high-dose episode has lower SPLs, but the exposure time is longer (not possible to see in the figure). The exposure time was almost four hours with approximately 88 dB equivalent SPL on the right ear, while the left ear had 10 dB lower levels.

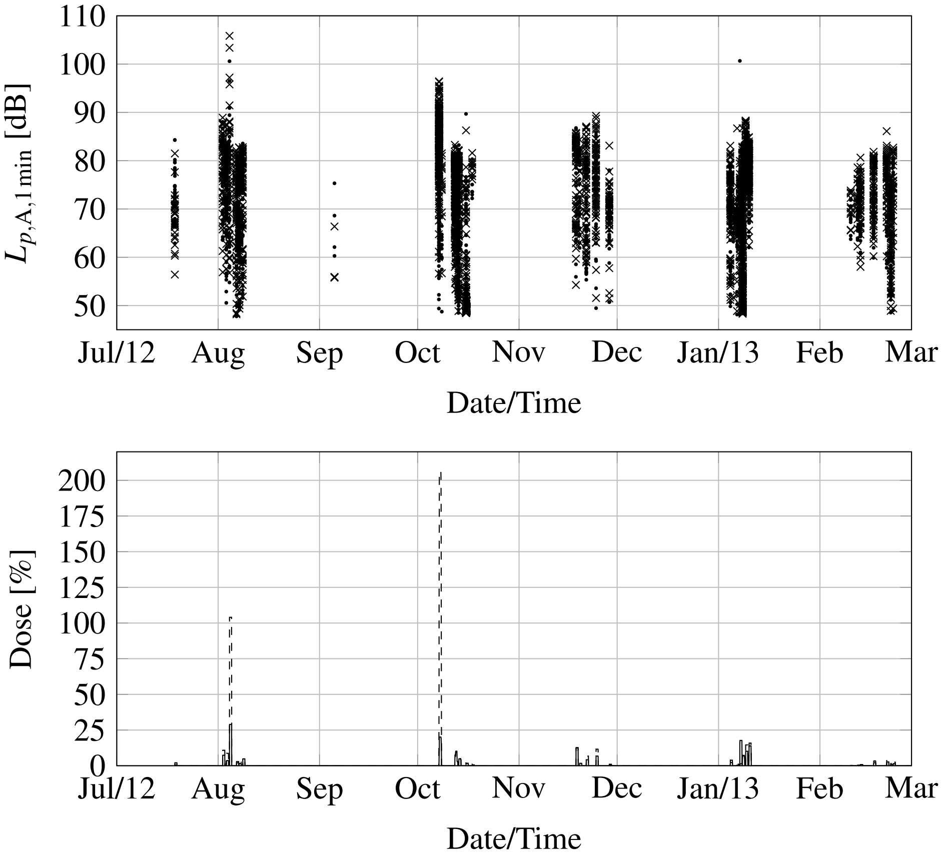

Figure 5.11: Personal sound exposure inside the earplug of the person. The sound pressure levels (SPLs) are A-weighted one minute equivalent levels, and the dose is in percentage of the allowed dose at offshore platforms ($L_{p,\mathrm{A},12\mathrm{h}} = 80\,\mathrm{dB}$). The filled circles represents measurements from the left ear, while the x's are from the right. Correspondingly the solid line in the lower plot is from the left ear, and the dashed lines are from the right.

Whether or not these two episodes are the reason for the potential hearing threshold shift is hard to say. Looking at the hearing level measurements in Figure 5.10 one might accept that there seems to be an ongoing negative development of the hearing levels during the entire period. Yamashita et al. (2005) reports that delayed production of damaging free radicals can go on for up to two weeks after a loud sound exposure, but since the high dose episodes are measured several months earlier they are probably not the cause of the hearing threshold shift. Another possibility is that the person did not wear QP when exposed to the damaging sound. It is, therefore, difficult to conclude what might have caused the hearing threshold shift based on the data collected.

Because of this lack of evidence it is also difficult to recommend any concrete counteractions. A possibility would be to start a dialogue with the employee to put focus on the hearing. Several general recommendations could be given, e.g. emphasize the importance of sufficient hearing protection both at work and during leisure activities, and recommend more frequent hearing measurements during the following time period. The latter will ensure that further negative development is detected as soon as possible, opening up for more specific counteractions.

5.2.4 Continued Hearing Monitoring

As mentioned, one recommendation that should be given after a warning signal is to keep performing hearing measurements. This way one might detect if more frequencies are ‘out of control’.

In Figure 5.12 to Figure 5.17 the process control of the entire time series is shown. One should note that 6 kHz at the left ear, and 3 kHz at the right ear both signals an out-of-control process not long after the first signal. More precisely, the three signals occur at observation 35, (right ear, 6 kHz), 37 (right ear, 3 kHz), and 39 (left ear, 6 kHz).

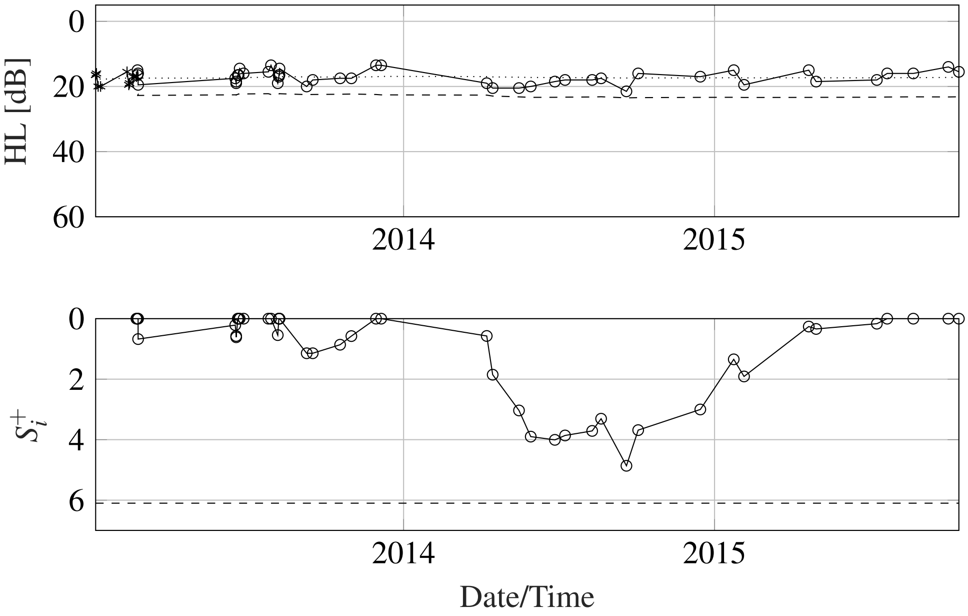

Figure 5.12: Hearing monitoring, full time series, of 3 kHz at the left ear.

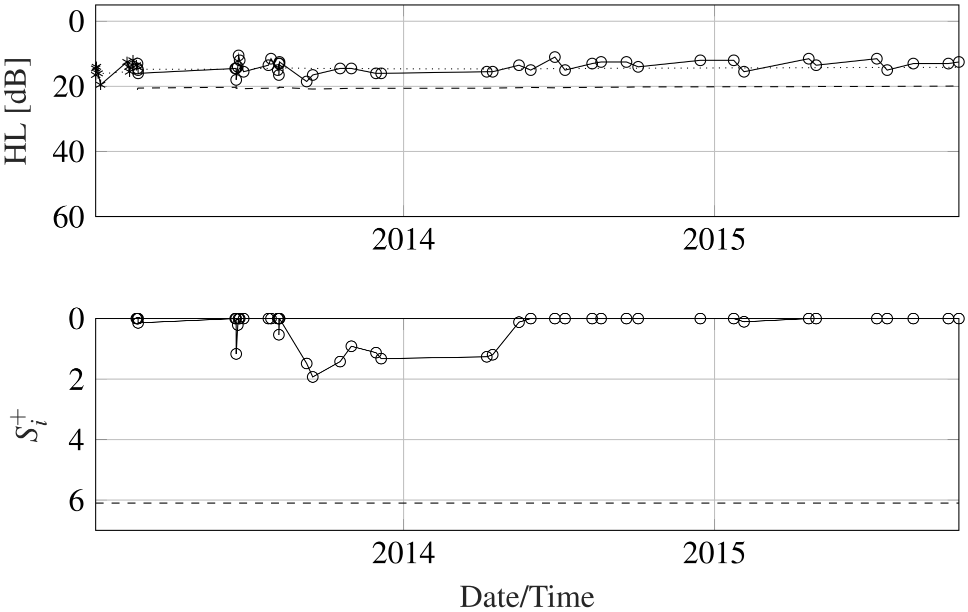

Figure 5.13: Hearing monitoring, full time series, of 4 kHz at the left ear.

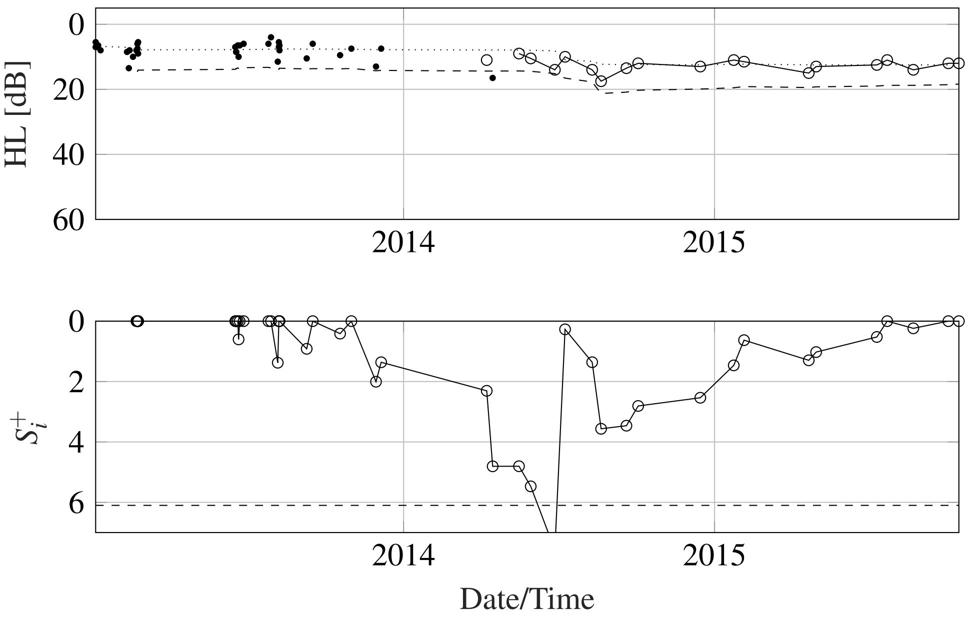

Figure 5.14: Hearing monitoring, full time series, of 6 kHz at the left ear.

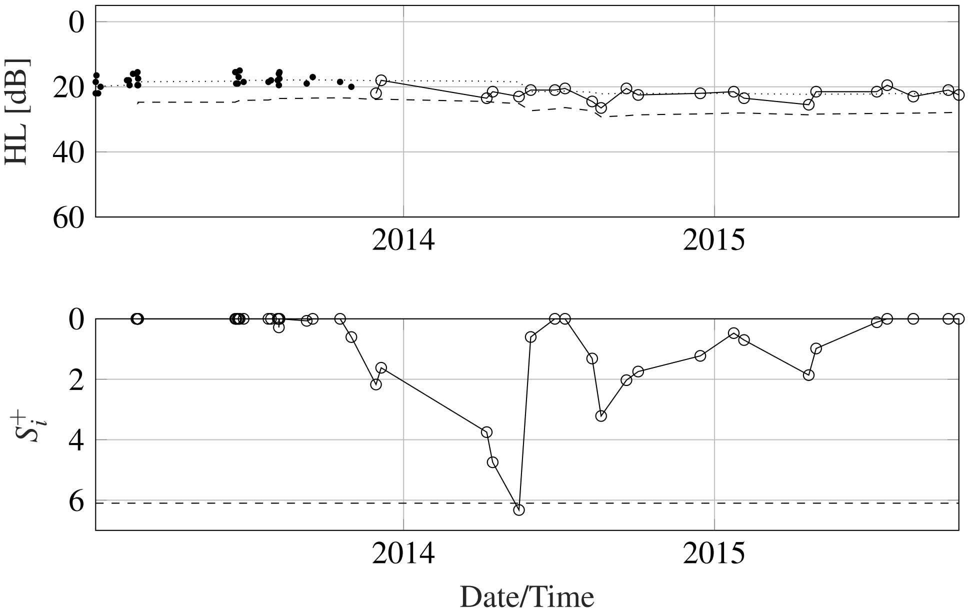

Figure 5.15: Hearing monitoring, full time series, of 3 kHz at the right ear.

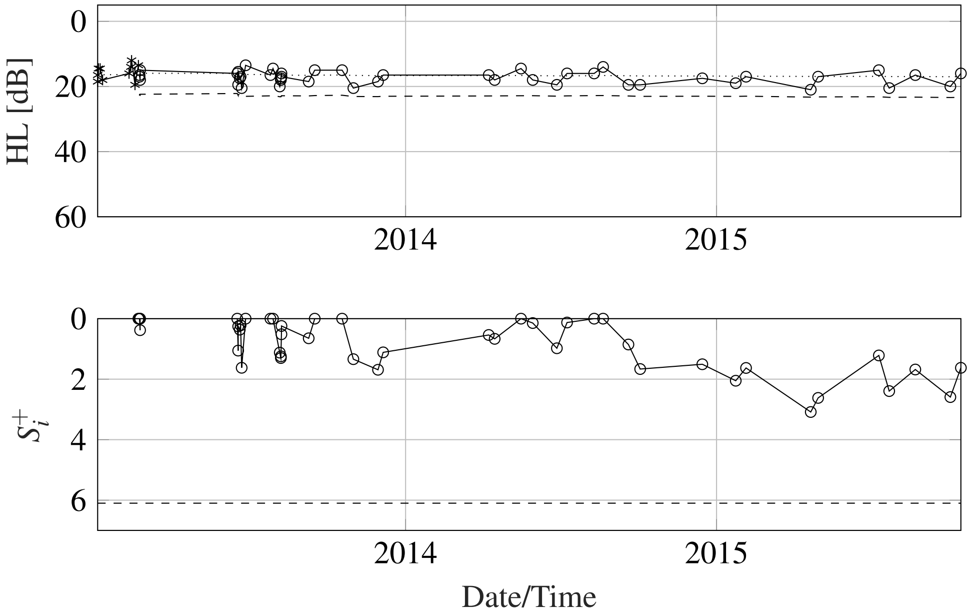

Figure 5.16: Hearing monitoring, full time series, of 4 kHz at the right ear.

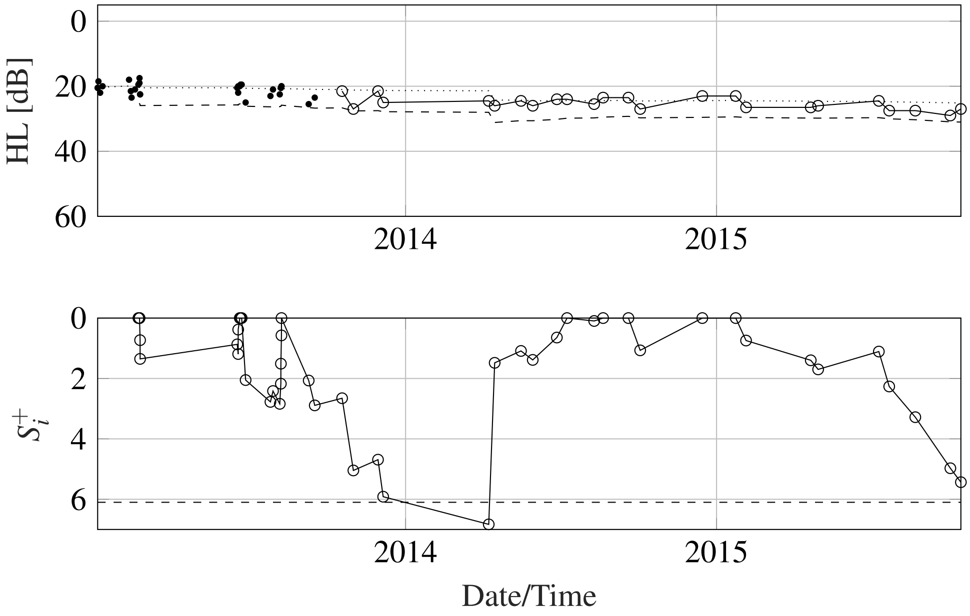

Figure 5.17: Hearing monitoring, full time series, of 6 kHz at the right ear.

Looking at the control charts one might also see that 6 kHz in the right ear seem to be approaching a new warning signal. This might further support the conclusion that the hearing is under negative development and that the worker should be advised to take extra good care of the hearing and that closer follow-up might be necessary.

The size of the shifts can also be determined by looking at the difference between the mean value of the last five measurements and the measurements before these five observations. Table 5.4 shows the results from such estimation. As can be seen around 3 dB hearing threshold shifts are detected.

Table 5.4: 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 chart signals an out-of-control situation.

Ear	**Freq**	**Warning signal**	**Post HL-value**	**Pre HL-value**	**Difference**
Right	6 kHz	35	23.9 dB	21.0 dB	2.9 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

Chapter 6

Discussion

Hearing conservation is, and will continue to be, an important topic for public health. As the age of the population increases it becomes increasingly important to reduce any hearing loss accumulated throughout life. Even if age itself most likely contributes to this age-related hearing loss, there is also a noise-induced portion of the hearing threshold shift for many individuals. Using the statistical process control regime presented in this thesis one might go from a very reactive hearing health care system, to a more proactive one. Today it is necessary to see a change in the hearing threshold of at least 15 dB between two consecutive measurements to conclude that the hearing has changed. The statistical process control approach introduced in this thesis can detect much smaller shifts (< 5 dB). This also means that it is not, strictly speaking, a proactive regime since a hearing threshold shift must be present for the process control to detect an ‘out of control’ situation, but the shift is so small that if further negative development is prevented, the person will not be considered hearing impaired. The assumption is that by detecting and stopping a noise-induced hearing impairment, the size of the age-related hearing loss will be reduced.

One should also be aware that a TTS will be considered to be ‘noise’ in the process control. ‘Noise’ means, in this context, that it is an unwanted input signal. Temporary threshold shifts are, as the name indicates, a temporary condition. If such measurements are used in the control chart the process might be flagged as being ‘out of control’, even if nothing has changed permanently. Especially temporary threshold shifts that are consistent over several measurements, something that can happen during common colds, are problematic. It is therefore important to assess all possible causes of an individual's hearing threshold shift when a warning signal is given.

As presented in Sec. 2.5, it seems difficult to predict the amount of TTS that a person acquires during exposure to loud sound. Even if there exist time functions that are fitted to experimental TTS data, the individual differences are large. This means that more harm than good can be done by trying to adjust for the TTS. Instead, if sound exposure data is available, one might use these to mark hearing measurements that are performed within 8 h – 16 h of a loud exposure. Then it is possible to see if a hearing measurement could be a result of TTS.

It can also be argued that TTS should not be adjusted for in the control chart since these shifts might be more problematic than previously thought. As mentioned earlier, Kujawa and Liberman (2009) have showed that permanent damage to the nerves might be present even if the hearing thresholds return to normal. However, others argue that TTS can be protective or that high sound exposure can toughen the hearing (Ahroon and Hamernik 1999; Henderson et al. 1999; Housley et al. 2013).

Another important point about the hearing monitoring regime presented is that it is an individual approach, meaning that the individual susceptibility to sound/noise and ototoxic chemicals is taken into account when detecting a hearing threshold shift. It must be emphasized that the hearing monitoring should not replace the existing barriers against NIHL, but be used as a complementary tool. Then it will become an extra protection especially important for the susceptible individuals that are 'overlooked' in the current regime.

For the hearing monitoring to work as intended it is important that a reliable hearing test is used as input. The monitoring is not limited to the test used in this project. Any hearing test giving consistent answers (i.e. variations in method and equipment must be controlled) will work, and the test does not even have to be calibrated. If uncalibrated measurements are used one will only detect if the hearing threshold shift has changed, not the actual hearing level. Since a person must be forwarded to authorized health personnel if a permanent damage is detected, this is not problematic. This also means that simple smart-phone or web solutions using standard audio equipment can be used as input to such hearing monitoring, opening up for new possibilities in preventing NIHL.

Since it is beneficial to have many hearing measurements as input to the control charts such simple hearing test solutions can also be of great importance. These render possible frequent testing ‘whenever’ and ‘wherever’ the test subject prefers. The quotation marks imply that there are some limitations, e.g. regarding the background noise level, but the solutions can be considerably more available than ordinary audiometry performed by OHSP. It will also be advisable to measure the background noise level during the hearing tests since they can be performed in environments with varying noise levels. If possible, the noise level should be measured underneath/inside the headphone/earplug used in the test. For the HPD used in this project this can be done since the earplug both has a microphone and transducer on the inside of the earplug.

As seen in the real-world data example it can also be a challenge to keep the workers motivated to perform hearing tests for a prolonged time. The example revealed that the worker performed 34 hearing tests the first year, 11 the next, and finally 9 the last year of the project. No systematic effort was made to try to increase the number of tests, but strategies for motivating the workers seems necessary. If the hearing test can be performed using a smart-phone, a simple strategy would be to let the device give an automatic reminder when it is time to take a new test.

The Norwegian government has expressed that they want the population to be able to take more responsibility for their own health (Ministry of Health and Care Services 2015). This can only be achieved if proper education and tools are provided to promote health and prevent disease (Resnik 2007). A hearing monitoring scheme can be utilized as a tool to raise awareness of the hearing and can be used to prevent NIHL.

Chapter 7

Conclusions

The main hypothesis for the proposed hearing monitoring system was

H1: The proposed hearing monitoring system is better than the current hearing test regime.

Since it is difficult to test this hypothesis two sub-hypotheses were expressed;

$H1_A$: The proposed hearing monitoring system can detect smaller hearing threshold shifts than the current regime.

$H1_B$: The proposed hearing monitoring system can detect threshold shifts faster than the current regime.

A challenge is that these hypotheses depend on the set-up of the process control chart. It is, however, possible to find parameters that will outperform the current regime. Because of the large intra-subject variability in normal pure tone audiometry a 15 dB hearing level shift must be seen before one can conclude that there has been an elevated threshold. With the proposed hearing monitoring system it will be possible to detect shifts of less than one standard deviation. The standard deviation used in the simulations was 5 dB, but the data from the real-world data collection both from the offshore installations and office test-runs, often show less variability ($\approx$2 dB to 3 dB). The simulation results also show that a step shift of one standard deviation can be detected for 50 % of the users within 12 observations, and that if the step size is increased to two standard deviations 95 % of the shifts will be detected before 8 observations are made.

As seen in Table 3.3, the probability of detecting a 10 dB shift in hearing threshold is less than 10 % using the current regime. For the new hearing monitoring system the probability of detecting the same shift becomes almost 100 % if more than four (3 dB standard deviation) or ten (5 dB standard deviation) measurements are made.

A goal of the Next Step research project was to reduce the number of people with noise-induced hearing loss in the petroleum industry. Even if it cannot be concluded that the goal is met during the project, it is reasonable to say that by using the hearing monitoring regime presented in this thesis it is possible to detect small hearing threshold shifts. By implementing individual counteractions after detecting a threshold shift, it is therefore possible to stop further negative development and thus preventing the individual to become hearing impaired.

7.1 System Implementation

There are several fields where the presented hearing monitoring system can be applied.

First of all, any industry with high sound exposure, where hearing monitoring is part of the systematic work to assure a safe working environment, is an obvious candidate. As discussed in this thesis, the proposed hearing monitoring scheme can replace the existing methods, while improving the individual's safety. The experience from the Next Step project, where Statoil ASA has been a participant, is that the system can be implemented in such a large company. A possible success criterion for the system is the economic profit such system can give. By replacing the expensive solution used today, where workers leave their work place and go to the occupational health service provider to perform a hearing test, the cost can be reduced significantly. If the new hearing monitoring regime can prove to be better than the current regime, the system has good prospects. Further testing in real-world scenarios is needed to verify this.

Additionally, the music industry, both artists and audio engineers, could be interested. This group of professionals relies on their hearing, hence they should be interested in protecting it. A challenge is that they do not necessary want to know the status of their hearing. This statement is supported by the results from a study conducted by Laitinen and Poulsen (2008) only 35 % of musicians answered that they had tested their hearing within the last 3 years. On the question ‘Are you worried about your hearing?’, approx. 60 % answered ‘not at all’ or ‘only a little’, and less than 15 % answered ‘quite a lot’ or ‘very much’. If the hearing monitoring scheme could be introduced at an early stage in an artist’s/audio engineer’s career, when the hearing still is uncompromised, it is possible that such a system could give the person many more years enjoying (and hearing) good music. K. R. Kähäri et al. (2001) also points out that musicians are well-acquainted with the process of detecting pure tones and that normal pure tone audiometry is not sensitive enough to detect early stages of hearing disorders. If this is true, the group is ideal for the proposed hearing monitoring since they will have a low standard deviation in their hearing test results. Hence it will be possible to detect very small hearing threshold shifts.

Finally, persons undergoing ototoxic chemotherapy can also benefit from such system. If the hearing can be closely monitored during chemotherapy treatment, it is possible that hearing loss can be prevented by adjusting the dose.

7.2 Future Work

A possible improvement of the process control is to use multivariate control charts. As mentioned in Sec. 3.4, there might be a correlation between the frequencies and/or ears that can be utilized to either detect smaller hearing threshold shifts, or, perhaps more importantly, to improve the robustness of the hearing monitoring by reducing the number of false alarms. It is, however, not obvious that it will be beneficial to use multivariate control charts since they can be harder to interpret when a warning signal is given. This must be looked further into.

When the amount of data from such hearing monitoring increases it also becomes possible to get more knowledge about the hearing. When hearing threshold shifts have been detected in many individuals it is possible to start looking at an exposure-response relationship. Even if the individual susceptibility always will play a part in NIHL it might be possible to come up with a better predictor than what is used today. The current regime, using a single value sound limit (i.e. $L_{p,\mathrm{A},8\,\mathrm{h}} = 85\,\mathrm{dB}$) and the equal energy hypothesis (The Norwegian Labour Inspection Authority 2015), is not good enough to prevent damage for everyone exposed to loud sound. A possible approach is to use machine learning as large amounts of data become available. This can give completely new ways of determining if a sound is damaging or not.

Further the data collected can be used to find normal hearing development curves, both as function of age, gender and other possible parameters. This can give valuable insight into normal hearing development.

It can also be possible to use the exposure data from the Quietpro® QP100Ex to adjust the source levels for noise calculators used to determine permitted working hours for noisy operations (e.g. http://noisecalculator.statoil.com/).

Bibliography

Ahroon, W. A. and Hamernik, R. P. (1999). 'Noise-induced hearing loss in the noise-toughened auditory system'. In: *Hear. Res.* 129.12, pp. 101–110. URL: http://www.sciencedirect.com/science/article/pii/S0378595598002275 (see pp. 35, 122).

Amin, R. W. and Ethridge, R. A. (1998). 'A Note on Individual and Moving Range Control Charts'. In: *J. Qual. Technol.* 30, pp. 70–74 (see pp. 40, 44, 45).

Beck, D. L. and Repovsch, J. L. (2013). *Observations on Speech, Speech in Noise, and Embedded Tests*. The Hearing Review [Online]. Retrieved December 20, 2016. URL: http://www.hearingreview.com/2013/07/observations-on-speech-speech-in-noise-and-embedded-tests/ (see p. 11).

Berger, J. R. and Blum, A. S. (2007). 'Brainstem Audiotory Evoked Potentials'. In: *The Clinical Neurophysiology Primer*. Ed. by A. S. Blum and S. B. Rutkove. Humana Press, pp. 475–484 (see pp. 12, 13).

Berglund, B., Lindvall, T. and Schwela, D. H., eds. (1999). *Guidelines for community noise*. Geneva, Switzerland: World Health Organization (see pp. 22, 29).

Best, M. and Neuhauser, D. (2006). 'Walter A Shewhart, 1924, and the Hawthorne factory'. In: *Qual. Saf. Health Care* 15, pp. 142–143 (see p. 39).

Bogoch, I. I., House, R. A. and Kudla, I. (2005). 'Perceptions about hearing protection and noise-induced hearing loss of attendees of rock concerts'. In: *Can. J. Public Health* 96.1, pp. 69–72 (see p. 20).

Borror, C. M., Montgomery, D. C. and Runger, G. C. (1999). 'Robustness of the EWMA control chart to non-normality'. In: *J. Qual. Technol.* 31.3, pp. 309–316. URL: http://search.proquest.com/docview/214481216?accountid=12870 (see p. 51).

Bryce, G. R., Gaudard, M. A. and Joiner, B. L. (1997). 'Estimating the standard deviation for individuals control charts'. In: *Qual. Eng.* 10.2, pp. 331–341 (see p. 44).

Chakraborti, S. (2007). 'Run length distribution and percentiles: the Shewhart chart with unknown parameters'. In: *Qual. Eng.* 19, pp. 119–127 (see p. 42).

Champ, C. W. and Woodall, W. H. (1987). 'Exact Results for Shewhart Control Charts with Supplementary Runs Rules'. In: *Technometrics* 29.4, pp. 393–399. URL: http://www.jstor.org/stable/1269449 (see p. 46).

Concha-Barrientos, M., Campbell-Lendrum, D. and Steenland, K. (2004). *Occupational noise: assessing the burden of disease from work-related hearing impairment at national and local levels*. Geneva, Switzerland: World Health Organization. (WHO Environmental Burden of Disease Series, No. 9) (see p. 1).

Crandell, C., Mills, T. L. and Gauthier, R. (2004). 'Knowledge, behaviors, and attitudes about hearing loss and hearing protection among racial/ethnically diverse young adults'. In: *J. Natl. Med. Assoc.* 96.2, pp. 176–186 (see p. 38).

Dau, T., Wegner, O., Mellert, V. and Kollmeier, B. (2000). 'Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion'. In: *J. Acoust. Soc. Am.* 107.3, pp. 1530–1540 (see p. 12).

Deci, E. L., Koestner, R. and Ryan, R. M. (2001). 'Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again'. In: *Rev. Educ. Res.* 71.1, pp. 1–27. URL: http://www.jstor.org/stable/3516064 (see p. 92).

European Commission (2003). *Directive 2003/10/EC of the European Parliament and of the Council of 6 February 2003 on the Minimum Health and Safety Requirements Regarding Exposure of Workers to the Risks Arising from Physical Agents (Noise)*. OJ L 042, 15.2.2003, p. 38 (see p. 21).

European Commission (2008). *Potential health risks of exposure to noise from personal music players and mobile phones including a music playing function*. Retrieved November 3, 2016. Directorate-General for Health & Consumers, Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR), Brussel. URL: http://ec.europa.eu/health/ph_risk/committees/04_scenihr/docs/scenihr_o_017.pdf (see p. 21).

Ferraro, J. A. (2010). 'Electrocochleography: A Review of Recording Approaches, Clinical Applications, and New Findings in Adults and Children'. In: *J. Am. Ac. Audiol.* 21.3, pp. 145–152. URL: http://openurl.ingenta.com/content/xref?genre=article&issn=1050-0545&volume=21&issue=3&spage=145 (see p. 12).

Fetoni, A. R., Picciotti, P. M., Paludetti, G. and Troiani, D. (2011). 'Pathogenesis of presbycusis in animal models: a review'. In: *Exp. Gerontol.* 46.6, pp. 413–425 (see p. 2).

Fisher, C. B. and Harrington McCarthy, E. L. (2013). 'Ethics in prevention science involving genetic testing'. In: *Prev. Sci.* 14.3, pp. 310–318 (see p. 94).

Franks, J. R. (2001). 'Hearing Measurements'. In: *Occupational exposure to noise: evaluation, prevention and control*. Ed. by B. Goelzer, C. Hansen and G. Sehrndt. Geneva, Switzerland: World Health Organization, Federal Institute for Occupational Safety and Health. Chap. 8, pp. 183–232 (see p. 38).

Ghosh, D. and Vogt, A. (2012). 'Outliers: An evaluation of methodologies'. In: *Joint Statistical Meetings*. Am. Stat. Assoc. San Diego, CA, pp. 3455–3460 (see p. 54).

Gjestland, T., Kvaløy, O., Tronstad, T. V. and Melvær, A. (2016). 'Active Hearing Protection Device Provides Unique Possibilities for Hearing Research'. In: *SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility, 11-13 April*. Stavanger, Norway: Society of Petroleum Engineers, pp. 1–7 (see p. vi).

Gjestland, T. and Tronstad, T. V. (2016). 'The efficacy of sound regulations on the listening levels of pop concerts'. In: *J. Occup. Environ. Hyg.* 14, pp. 17–22 (see p. vi).

Gjestland, T., Tronstad, T. V. and Kvaløy, O. (2016). *Mulig metode for tidlig deteksjon av hørselstap. Eksempler på bruk av prosesskontrollmetoder på hørselsdata fra SLASH (in Norwegian, Possible method for early detection of hearing loss. Examples on use of process control on hearing data from SLASH)*. A27862, ISBN 9788214061369. Trondheim, Norway: SINTEF (see p. vi).

Grandjean, P. and Sorsa, M. (1996). 'Ethical aspects of genetic predisposition to environmentally-related disease'. In: *Sci. Total Environ.* 184.1-2, pp. 37–43 (see p. 94).

Green, D. M. (1990). 'Stimulus selection in adaptive psychophysical procedures'. In: *J. Acoust. Soc. Am.* 87.6, pp. 2662–2674 (see p. 15).

Hamernik, R. P., Qiu, W. and Davis, B. (2007). 'Hearing loss from interrupted, intermittent, and time varying non-Gaussian noise exposure: The applicability of the equal energy hypothesis'. In: *J. Acoust. Soc. Am.* 122.4, pp. 2245–2254 (see p. 35).

Harris, K. C., Dubno, J. R., Keren, N. I., Ahlstrom, J. B. and Eckert, M. A. (2009). 'Speech recognition in younger and older adults: a dependency on low-level auditory cortex'. In: *J. Neurosci.* 29.19, pp. 6078–6087 (see p. 11).

Hawkins, D. M. (1980). *Identification of outliers*. 1st. Monographs on Statistics and Applied Probability. Springer Netherlands (see p. 54).

Hawkins, D. M. (1987). 'Self-starting cusum charts for location and scale'. In: *J. Roy. Statist. Soc.* 36.4, pp. 299–316 (see p. 42).

Hawkins, D. M. and Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*. Ed. by P. Green, J. F. Lawless and V. Nair. Statistics for Engineering and Physical Science. Springer-Verlag New York (see pp. 42, 47).

Henderson, D., McFadden, S. L., Liu, C. C., Hight, N. and Zheng, X. Y. (1999). 'The role of antioxidants in protection from impulse noise'. In: *Ann. N. Y. Acad. Sci.* 884, pp. 368–380 (see pp. 35, 122).

Henry, J. A., Flick, C. L., Gilbert, A., Ellingson, R. M. and Fausti, S. A. (2001). 'Reliability of hearing thresholds: computer-automated testing with ER-4B Canal Phone earphones'. In: *J. Rehabil. Res. Dev.* 38.5, pp. 567–581 (see pp. 38, 56).

Housley, G. D. et al. (2013). 'ATP-gated ion channels mediate adaptation to elevated sound levels'. In: *Proc. Natl. Acad. Sci. USA* 110.18, pp. 7494–7499. eprint: http://www.pnas.org/content/110/18/7494.full.pdf. URL: http://www.pnas.org/content/110/18/7494.abstract (see pp. 35, 122).

Hughson, G. W., Mulholland, R. E. and Cowie, H. A. (2002). *Behavioural studies of people's attitudes to wearing hearing protection and how these might be changed*. 28. Edinburgh, United Kingdom: Institute of Occupational Medicine (see p. 4).

Hughson, W. and Westlake, H. (1944). 'Manual for program outline for rehabilitation of aural casualties both military and civilian'. In: *Trans. Amer. Acad. Ophthal. Otolaryng. Suppl.* 48, pp. 1–15 (see p. 14).

Institute for Work and Health (2015). 'Primary, Secondary and Tertiary Prevention'. In: *at work* 80. (Accessed: 30 Sept 2016). URL: https://www.iwh.on.ca/system/files/at-work/at_work_80.pdf (see p. 95).

ISO 1999 (2013). *Acoustics – Estimation of noise-induced hearing loss*. ISO. Geneva, Switzerland (see pp. 4, 21, 82).

ISO 389 (1998). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones*. ISO. Geneva, Switzerland (see p. 28).

ISO 7029 (2000). *Acoustics - Statistical distribution of hearing thresholds related to age and gender*. ISO. Geneva, Switzerland (see pp. 3, 56, 61, 62, 71, 113).

ISO 8253-1 (2010). *Acoustics - Audiometric test methods. Part 1: Pure-tone air and bone conduction audiometry*. ISO. Geneva, Switzerland (see pp. 10, 37).

ISO 8253-2 (2009). *Acoustics - Audiometric test methods. Part 2: Sound field audiometry with pure-tone and narrow-band test signals*. ISO. Geneva, Switzerland (see p. 10).

ISO 8253-3 (2012). *Acoustics - Audiometric test methods. Part 3: Speech audiometry*. ISO. Geneva, Switzerland (see p. 11).

Jensen, W. A., Jones-Farmer, A., Champ, C. W. and Woodall, W. H. (2006). 'Effects of Parameter Estimation on Control Chart Properties: A Literature Review'. In: *J. Qual. Technol.* 38.4, pp. 349–364 (see pp. 42, 56).

Jerlvall, L. and Arlinger, S. (1986). 'A comparison of 2-dB and 5-dB step size in pure-tone audiometry'. In: *Scand. Audiol.* 15.1, pp. 51–56 (see pp. 38, 56, 80).

Johnson, N. L. (1961). 'A Simple Theoretical Approach to Cumulative Sum Control Charts'. In: *J. Amer. Statist. Assoc.* 56.296, pp. 835–840. URL: http://www.jstor.org/stable/2281998 (see p. 50).

Jones, L. A. (2002). 'The statistical design of EWMA control charts with estimated parameters'. In: *J. Qual. Technol.* 34.3, pp. 277–288. URL: http://search.proquest.com/docview/214492520?accountid=12870 (see p. 53).

Kähäri, K. R., Axelsson, A., Hellström, P.-A. and Zachau, G. (2001). 'Hearing Assessment of Classical Orchestral Musicians'. In: *Scand. Audiol.* 30.1, pp. 13–23. URL: http://www.tandfonline.com/doi/full/10.1080/010503901750069536 (see p. 127).

Kähäri, K., Zachau, G., Eklöf, M. and Möller, C. (2004). 'The influence of music and stress on musicians' hearing'. In: *J. Sound Vib.* 277.3, pp. 627–631. URL: http://www.sciencedirect.com/science/article/pii/S0022460X04004523 (see p. 21).

Keeler, J. (1968). 'Compatible exposure and recovery functions for temporary threshold shift - Mechanical and electrical models'. In: *J. Sound Vib.* 7.2, pp. 220–235. URL: http://www.sciencedirect.com/science/article/pii/0022460X68902691 (see p. 20).

Kemp, D. T. (2002). 'Otoacoustic emissions, their origin in cochlear function, and use'. In: *Br. Med. Bull.* 63.1, pp. 223–241 (see p. 12).

Keppler, H., Ingeborg, D., Sofie, D. and Bart, V. (2015). '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'. In: *Noise Health* 17.78, pp. 253–262. eprint: http://www.noiseandhealth.org/article.asp?issn=1463-1741;year=2015;volume=17;issue=78;spage=253;epage=262;aulast=Keppler;t=6. URL: http://www.noiseandhealth.org/article.asp?issn=1463-1741;year=2015;volume=17;issue=78;spage=253;epage=262;aulast=Keppler;t=6 (see p. 38).

Khoo, M. B. C. and Quah, S. H. (2002). 'Computing the percentage points of the run-length distributions of multivariate CUSUM control charts'. In: *Qual. Eng.* 15, pp. 299–310 (see p. 42).

Kochkin, S. (2012). 'MarkeTrak VIII: The Key Influencing Factors in Hearing Aid Purchase Intent'. In: *Hear. Rev.* 19.3, pp. 12–25 (see p. 2).

Kontsevich, L. L. and Tyler, C. W. (1999). 'Bayesian adaptive estimation of psychometric slope and threshold'. In: *Vision Res.* 39.16, pp. 2729–2737. URL: http://www.sciencedirect.com/science/article/pii/S0042698998002855 (see p. 14).

Kujawa, S. G. and Liberman, M. C. (2009). 'Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss'. In: *J. Neurosci.* 29.45, pp. 14077–14085 (see pp. 10, 19, 32, 93, 122).

Laitinen, H. and Poulsen, T. (2008). 'Questionnaire Investigation of Musicians' Use of Hearing Protectors, Self Reported Hearing Disorders, and Their Experience of Their Working Environment'. In: *Int. J. Audiol.* 47.4, pp. 160–168. URL: http://www.tandfonline.com/doi/full/10.1080/14992020801886770 (visited on 13/01/2017) (see p. 127).

Leek, M. R. (2001). 'Adaptive procedures in psychophysical research'. In: *Percept. Psychophys.* 63.8, pp. 1279–1292. URL: http://dx.doi.org/10.3758/BF03194543 (see p. 14).

Levene, H. (1960). 'Robust tests for equality of variances'. In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Ed. by I. e. a. Olkin. Stanford University Press, pp. 278–292 (see p. 100).

Li, Z. and Wang, Z. (2010). 'Adaptive CUSUM of the Q chart'. In: *Int. J. Prod. Res.* 48.5, pp. 1287–1301 (see p. 53).

Li, Z., Zhang, J. and Wang, Z. (2010). 'Self-starting control chart for simultaneously monitoring process mean and variance'. In: *Int. J. Prod. Res.* 48.15, pp. 4537–4553. eprint: http://dx.doi.org/10.1080/00207540903051692. URL: http://dx.doi.org/10.1080/00207540903051692 (see p. 53).

Lie, A., Skogstad, M., Tynes, T., Johannessen, H. A., Nordby, K.-C., Mehlum, I. S., Arneberg, L., Engdahl, B. and Tambs, K. (2013). *Støy i arbeidslivet og helse*. (In Norwegian, Occupational noise and health) (see p. 1).

Maassen, M. et al. (2001). ‘Ear damage caused by leisure noise’. In: *Noise Health* 4.13, pp. 1–16 (see p. 21).

Margolis, R. and Morgan, D. (2008). ‘Automated pure-tone audiometry: an analysis of capacity, need, and benefit.’ In: *Am. J. Audiol.* 17.2, pp. 109–113. URL: http://search.ebscohost.com/login.aspx?direct=true&db=ccm&AN=105603508&site=eds-live (see p. 13).

MathWorks (2014). *MATLAB R2014a (32-bit)*. URL: http://www.mathworks.com (see p. 17).

MathWorks (2016). *MATLAB R2016a (64-bit)*. URL: http://www.mathworks.com (see p. 60).

Melnick, W. (1991). ‘Human temporary threshold shift (TTS) and damage risk’. In: *J. Acoust. Soc. Am.* 90.1, pp. 147–154. URL: http://scitation.aip.org/content/asa/journal/jasa/90/1/10.1121/1.401308 (see pp. 19, 93).

Mercier, V., Luy, D. and Hohmann, B. (2003). ‘The sound exposure of the audience at a music festival’. In: *Noise Health* 5.19, pp. 51–8. URL: http://search.proquest.com/docview/203698724?accountid=12870 (see p. 21).

Mills, J. H., Gilbert, R. M. and Adkins, W. Y. (1979). ‘Temporary threshold shifts in humans exposed to octave bands of noise for 16 to 24 hours’. In: *J. Acoust. Soc. Am.* 65.5, pp. 1238–1248 (see p. 20).

Ministry of Health and Care Services (2015). *Folkehelsemeldingen (in Norwegian, Public health report)* (see p. 123).

Montgomery, D. C. (2013). *Statistical Quality Control*. Ed. by J. Welter. 7th. John Wiley & Sons (see pp. 40, 44, 46–48, 50, 51).

National Academies of Sciences, Engineering, and Medicine (2016). *Hearing Health Care for Adults: Priorities for Improving Access and Affordability*. Ed. by D. G. Blazer, S. Domnitz and C. T. Liverman. The National Academies Press (see p. 4).

Nielson, K., Kreisman, B. M., Pallett, S. and Kreisman, N. V. (2011). ‘Effects of Noise Attenuation Devices on Screening Distortion Product Otoacoustic Emissions in Different Levels of Background Noise’. In: *J. Educ. Audiol.* 17, pp. 53–61 (see p. 29).

NIOSH (2010). *Occupationally-Induced Hearing Loss*. DHHS. Fact sheet. URL: http://www.cdc.gov/niosh/docs/2010-136/pdfs/2010-136.pdf (see p. 1).

NIST/SEMATECH (2012). *e-Handbook of Statistical Methods*. Accessed: 21 Oct 2016. URL: http://itl.nist.gov/div898/handbook/pmc/section3/pmc323.htm (see p. 50).

Opperman, D. A., Reifman, W., Schlauch, R. and Levine, S. (2006). ‘Incidence of spontaneous hearing threshold shifts during modern concert performances’. In: *Otolaryngol Head Neck Surg* 134.4, pp. 667–673 (see p. 21).

Ordoñez, R. and Hammershøi, D. (2011). ‘Time and Frequency Characteristics of Temporary Threshold Shifts Caused by Pure Tone Exposures’. In: *Acta. Acust. united Ac.* 97.5, pp. 809–818. URL: http://openurl.ingenta.com/content/xref?genre=article&issn=1610-1928&volume=97&issue=5&spage=809 (visited on 16/01/2017) (see p. 20).

PatientView (2015). ‘What do patients and carers need in health apps - but are not getting?’ In: *EU funded GET Project* (see p. 92).

Patuzzi, R. (1998). ‘Exponential onset and recovery of temporary threshold shift after loud sound: evidence for long-term inactivation of mechano-electrical transduction channels’. In: *Hear. Res.* 125, pp. 17–38. URL: http://www.sciencedirect.com/science/article/pii/S0378595598001269 (see p. 20).

Plack, C. J., Barker, D. and Prendergast, G. (2014). ‘Perceptual consequences of "hidden" hearing loss’. In: *Trends Hear.* 18, pp. 1–11 (see p. 93).

Poole, K. (2011). *Optimum test conditions and variability of otoacoustic emission testing in individuals with normal hearing*. RR840. Health and Safety Laboratory (see p. 29).

Quesenberry, C. (1991). ‘SPC Q charts for start-up processes and short or long runs’. In: *J. Qual. Technol.* 23, pp. 213–224 (see pp. 51, 53).

Resnik, D. B. (2007). ‘Responsibility for health: personal, social, and environmental’. In: *J. Med. Ethics* 33.8, pp. 444–445 (see p. 123).

Robinson, D. W. and Sutton, G. J. (1979). ‘Age effect in hearing – a comparative analysis of published threshold data’. In: *Audiology* 18.4, pp. 320–334 (see p. 2).

Ryan, R. M. and Deci, E. L. (2000). ‘Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions’. In: *Contemp. Educ. Psychol.* 25.1, pp. 54–67. URL: http://www.sciencedirect.com/science/article/pii/S0361476X99910202 (see p. 91).

Schaette, R. and McAlpine, D. (2011). ‘Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model’. In: *J. Neurosci.* 31.38, pp. 13452–13457 (see p. 32).

Sennheiser Electronic Corporation (2016). *Sennheiser HDA 200*. Online. Accessed: 2 Nov 2016. URL: http://en-de.sennheiser.com/downloads/a135337001d00197026b38d9c278f44d.pdf (see p. 28).

Serra, C., Rodriguez, M. C., Delclos, G. L., Plana, M., Gomez Lopez, L. I. and Benavides, F. G. (2007). ‘Criteria and methods used for the assessment of fitness for work: a systematic review'. In: *Occup. Environ. Med.* 64.5, pp. 304–312 (see p. 94).

Shield, B. (2006). *Evaluation of the social and economic costs of hearing impairment*. [Accessed Sept 14 2016]. hear-it.org. URL: http://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf (see p. 2).

Sliwinska-Kowalska, M. and Pawelczyk, M. (2013). 'Contribution of genetic factors to noise-induced hearing loss: A human studies review'. In: *Mutat. Res.* 752.1, pp. 61–65 (see pp. 38, 95).

Smith-Olinde, L., Nicholson, N., Chivers, C., Highley, P. and Williams, D. K. (2006). 'Test-Retest Reliability of In Situ Unaided Thresholds in Adults'. In: *Am. J. Audiol.* 15.1, pp. 75–80. URL: http://search.proquest.com/docview/204420288?accountid=12870 (see pp. 38, 56).

Sockalingam, R., Lee Choi, J., Choi, D. and Kei, J. (2007). 'Test-retest reliability of distortion-product otoacoustic emissions in children with normal hearing: a preliminary study'. In: *Int. J. Audiol.* 46.7, pp. 351–354 (see p. 29).

Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q. and Barbour, D. L. (2015). 'Fast, Continuous Audiogram Estimation Using Machine Learning'. In: *Ear Hear* 36.6, e326–335 (see p. 14).

Spankovich, C. and Le Prell, C. G. (2013). 'Healthy diets, healthy hearing: National Health and Nutrition Examination Survey, 1999-2002'. In: *Int. J. Audiol.* 52.6, pp. 369–376 (see p. 38).

Stenklev, N. C. and Laukli, E. (2004). 'Presbyacusis-hearing thresholds and the ISO 7029'. In: *Int. J. Audiol.* 43.5, pp. 295–306 (see p. 3).

Strasser, H., Irle, H. and Legler, R. (2003). 'Temporary hearing threshold shifts and restitution after energy-equivalent exposures to industrial noise and classical music'. In: *Noise Health* 5.20, pp. 75–84 (see p. 35).

Stuart, A., Stenstrom, R., Tompkins, C. and Vandenhoff, S. (1991). 'Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults'. In: *Audiology* 30.2, pp. 82–90 (see pp. 38, 56).

Tak, S., Davis, R. R. and Calvert, G. M. (2009). 'Exposure to hazardous workplace noise and use of hearing protection devices among US workers – NHANES, 1999–2004'. In: *Am. J. Ind. Med.* 52.5, pp. 358–371 (see p. 1).

The American Speech-Language-Hearing Association (1988). *Guidelines for determining threshold level for speech*. ASHA. URL: http://www.asha.org/policy (see p. 10).

The Norwegian Directorate of Health (2011). *Musikkanlegg og helse – Veileder til arrangører og kommuner*. (in Norwegian, Music equipment and health

– Guideline for organisers and municipalities.) Guideline. The Norwegian Directorate of Health. URL: http://www.helsedirektoratet.no/publikasjoner/musikkanlegg-og-helse-veileder-til-arrangorer-og-kommuner/Publikasjoner/musikkanlegg-og-helse-veileder-til-arrangorer-og-kommuner.pdf (see p. 22).

The Norwegian Labour Inspection Authority (2013). *Veiledning om Hørselskontroll av støyeksponerte arbeidstakere*. (in Norwegian, Guideline for hearing measurements of noise-exposed employees). Guideline. URL: http://www.arbeidstilsynet.no/binfil/download2.php?tid=77943 (see pp. 37, 38, 55, 57, 72, 74, 96, 103).

The Norwegian Labour Inspection Authority (2015). *The Working Environment Act*. 104. FOR-2006-04-26-456 (see pp. 21, 127).

The Norwegian Ministry of Labour and Social Affairs (2016). 'Forskrift om utførelse av arbeid'. (in Norwegian, Regulations concerning the performance of work). In: *Lovdata* 14-11 (see p. 37).

Tronstad, T. V. (2015). 'Hearing Measurements During Two Norwegian Music Festivals'. In: *58th International Conference: Music Induced Hearing Disorders*. Aalborg, Denmark: Audio Engineering Society, pp. 1–4 (see pp. v, vi, 20, 28).

Tronstad, T. V. (2017). 'Statistical tool to detect small hearing threshold shifts'. In: *Int. J. Audiol.* 56.8, pp. 596–606 (see pp. v, vi, 39).

Tronstad, T. V. and Gelderblom, F. B. (2016). 'Sound exposure during outdoor music festivals'. In: *Noise Health* 18.83, pp. 220–228 (see pp. v, vi, 20, 23).

Tsiamyrtzis, P. and Hawkins, D. M. (2008). 'A Bayesian EWMA method to detect jumps at the start-up phase of a process'. In: *Qual. Reliab. Eng. Int.* 24.6, pp. 721–735. URL: http://dx.doi.org/10.1002/qre.952 (see p. 53).

Vinay, Henriksen, V., Kvaløy, O. and Svensson, U. P. (2014). 'Development and Calibration of a New Automated Method to Measure Air Conduction Auditory Thresholds Using an Active Earplug'. In: *Acta Acoust. united Ac.* 100.1, pp. 113–117 (see pp. 7, 15, 29).

Vinay, Svensson, U. P., Kvaløy, O. and Berg, T. (2015). 'A comparison of test-retest variability and time efficiency of auditory thresholds measured with pure tone audiometry and new early warning test'. In: *Appl. Acoust.* 90, pp. 153–159 (see pp. 15, 38, 55, 56, 89).

Vogel, I., Brug, J., van der Ploeg, C. P. B. and Raat, H. (2007). 'Young People's Exposure to Loud Music: A Summary of the Literature'. In: *Am. J. Prev. Med.* 33.2, pp. 124–133. URL: http://www.sciencedirect.com/science/article/pii/S0749379707002309 (see p. 21).

Vogel, I., Verschuure, H., Ploeg, C. P. van der, Brug, J. and Raat, H. (2010). ‘Estimating adolescent risk for hearing loss based on data from a large school-based survey’. In: *Am. J. Public Health* 100.6, pp. 1095–1100 (see p. 21).

Vos, J. (2005). *Auditory Tests for the Early Detection of Noise-Susceptible Individuals – A Literature Study*. Retrieved September 8, 2016. TNO Human Factors. URL: https://www.sto.nato.int/publications/STO%5C%20Meeting%5C%20Proceedings/RTO-MP-HFM-123/MP-HFM-123-05.pdf (see p. 38).

Wagner, W., Heppelmann, G., Vonthein, R. and Zenner, H. P. (2008). ‘Test-retest repeatability of distortion product otoacoustic emissions’. In: *Ear Hear.* 29.3, pp. 378–391 (see p. 29).

Ward, W. D., Glorig, A. and Sklar, D. L. (1958). ‘Dependence of Temporary Threshold Shift at 4 kc on Intensity and Time’. In: *J. Acoust. Soc. Am.* 30.10, pp. 944–954. URL: http://scitation.aip.org/content/asa/journal/jasa/30/10/10.1121/1.1909414 (see p. 20).

Western Electric Company (1956). *Statistical Quality Control Handbook*. Indianapolis: Western Electric (see p. 45).

Wheeler, D. J. (2010). *Individual Charts Done Right and Wrong*. (Accessed 29 Jan 2016). URL: http://www.qualitydigest.com/inside/quality-insider-column/individual-charts-done-right-and-wrong.html (see p. 44).

WHO (2015). *Deafness and hearing loss, Fact sheet No 300*. Retrieved November 18, 2016. Geneva, Switzerland: World Health Organization. URL: http://www.who.int/mediacentre/factsheets/fs300/en/ (see p. 1).

Wikipedia (2017). *Nocebo*. Retrieved September 13, 2017. URL: https://en.wikipedia.org/wiki/Nocebo (see p. 92).

Woodall, W. H. and Montgomery, D. C. (2000). ‘Using ranges to estimate variability’. In: *Qual. Eng.* 13.2, pp. 211–217 (see p. 45).

Yamashita, D., Jiang, H. Y., Le Prell, C. G., Schacht, J. and Miller, J. M. (2005). ‘Post-exposure treatment attenuates noise-induced hearing loss’. In: *Neuroscience* 134.2, pp. 633–642 (see p. 114).

Zhao, F., Manchaiah, V. K., French, D. and Price, S. M. (2010). ‘Music exposure and hearing disorders: an overview’. In: *Int. J. Audiol.* 49.1, pp. 54–64 (see p. 21).