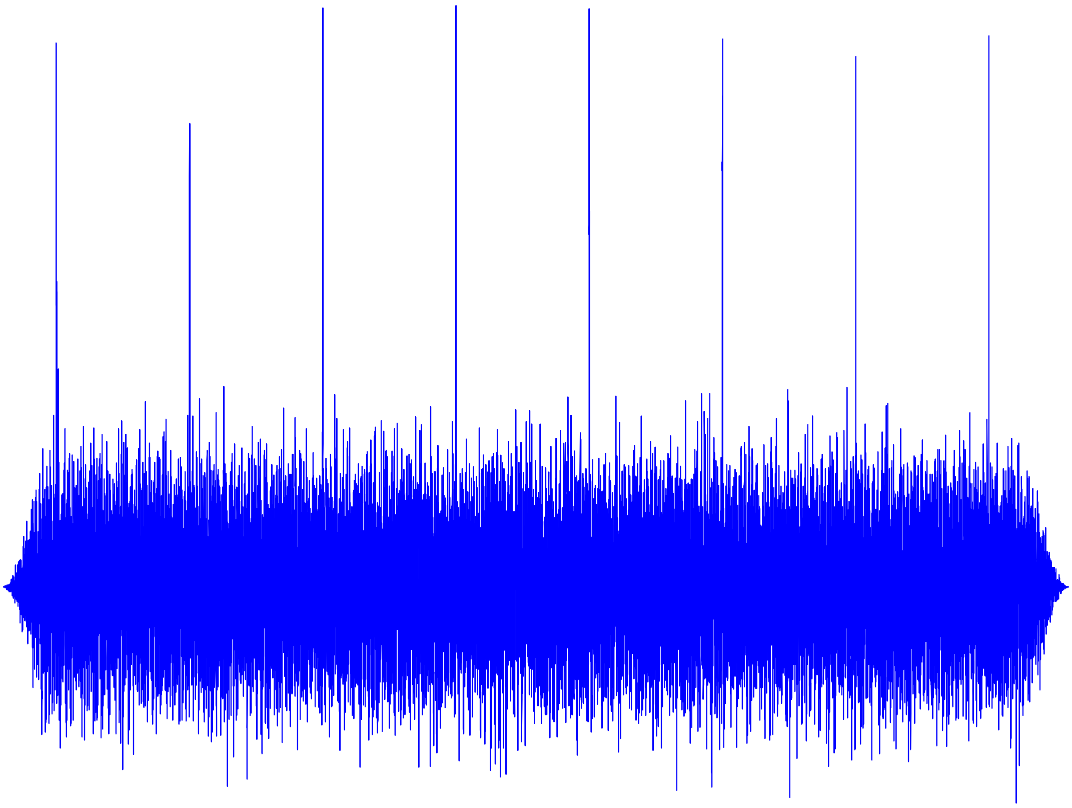
AALBORG UNIVERSITY

DETECTION AND RATINGS OF AUTOMATIVE IMPULSIVE EVENTS



Master Thesis

Martin Kynde

Spring 2012

AALBORG UNIVERSITY

Institute of Electronic Systems
Frederik Bajers Vej 7B5
9220 Aalborg Ø
Phone (+45) 9940 9940
Fax (+45) 9815 4008
http://www.sict.aau.dk

Project title:
Detection and ratings of automative impulsive events

Semester theme:
Master Thesis

Project period:
Master Program in Acoustics
4th. semester, Spring 2012

Project group:
12gr1061

Group member
Martin Kynde

Supervisor:
Christian Sejer Pedersen

Secondary supervisor
Woo-Keun Song, Brüel & Kjær

No. of printed copies: 5

No. of pages: 53

No. of appendix pages: 27

Total no. of pages: 87

Completed:
May 31, 2012

Abstract:

The detection and ratings of impulses can be used in wide fields within the industry. An engineer working with detecting impulses and reducing unwanted sounds in consumer products, most prominent job is to prevent all annoying sounds as impulses to be perceived by the costumer. Sound quality is a part of the total product quality.

In the area of NVH engineering they are often dealt with detecting and eliminate Buzz, Squeak and Rattle (BSR) events in vehicles.

Many attempts has been done in order to automatic detect BSR events and often the detection threshold is depending on an absolute level determined by a subjective evaluation or set manually by the authors.

In this project an algorithm independent of absolute level has been chosen to calculate the amount of impulsiveness of an acoustic signal.

Synthetically generated samples consisting of impulses mixed with noise were used as stimuli to the proposed algorithm. Concurrently the same set of samples has been evaluated in a subjective experiment. Existing literature does not describe a similar study of impulses with the mentioned parameters, conducted in a subjective experiment. That was the motivation to explore the described scenario.

The goal of the project was to reveal if any correlation existed between the objective and subjective evaluation of the samples.

No significant correlation was found between the two. Future work suggest to tune the settings in the algorithm to evaluate if an improvement can be achieved that match the subjective results with a higher degree of correlation.

Contents

Reading guide

ABSTRACT

The detection and ratings of impulses can be used in wide fields within the industry. An engineer working with detecting impulses and reducing unwanted sounds in consumer products, most prominent job is to prevent all annoying sounds as impulses to be perceived by the costumer. Sound quality is a part of the total product quality.

In the area of NVH engineering they are often dealt with detecting and eliminate Buzz, Squeak and Rattle (BSR) events in vehicles.

Many attempts has been done in order to automatic detect BSR events and often the detection threshold is depending on an absolute level determined by a subjective evaluation or set manually by the authors.

In this project an algorithm independent of absolute level has been chosen to calculate the amount of impulsiveness of an acoustic signal.

Synthetically generated samples consisting of impulses mixed with noise were used as stimuli to the proposed algorithm. Concurrently the same set of samples has been evaluated in a subjective experiment. Existing literature does not describe a similar study of impulses with the mentioned parameters, conducted in a subjective experiment. That was the motivation to explore the described scenario.

The goal of the project was to reveal if any correlation existed between the objective and subjective evaluation of the samples.

No significant correlation was found between the two. Future work suggest to tune the settings in the algorithm to evaluate if an improvement can be achieved that match the subjective results with a higher degree of correlation.

The report is structured in the following way.

CHAPTER 1 **Introduction and problem analysis**
Contains Introduction and problem analysis. An overview of the importance of detecting impulses in the industry is described. Existing methods for detecting impulses is listed and the main problem of this project is stated.

CHAPTER 2 **System overview**
Contains a block-diagram of the main features of the project and links it to the corresponding chapters.

CHAPTER 3 Impulsive Sounds
Contains a description of the different parameters involved in designing the samples to the objective and subjective evaluation and the technical aspects of achieving a flat frequency response and the correct levels concerning the subjective experiment.

CHAPTER 4 Methods for evaluation the samples
Contains a description of the subjective and objective parameters used to evaluate the samples.

CHAPTER 5 Experiment
Contains information about the psychoacoustic experiment conducted in order to obtain subjective ratings of the samples.

CHAPTER 6 Results
Contains results from; The subjective experiment, the objective calculation of impulsiveness and the correlation between them.

CHAPTER 7 Discussion
Contains discussion about the issues that were encountered when analysing the subjective results and the correlation between objective parameters with subjective scales.

CHAPTER 8 Conclusion
Contains a summary of what has been done in the project and gives an answer of if the main goal has been fulfilled.

Appendices:

APPENDIX A Samples
Contains a discussion about the chosen settings for each parameter in the design of samples. The number of samples are constrained by time limitations stated in the experiment chapter. The effective testing time is estimated.

APPENDIX B Experiment formalities
The first part of the appendix contains a Danish Instruction list and a translated English version. The second part contains answers from the subjects collected in the breaks of the experiment.

APPENDIX C Additional results
Contains additional results from the subjective analyse. Important conclusions are drawn in terms of annoyance and impulsiveness.

APPENDIX D Measurement of BeyerDynamic DT990 Pro headphones + Ear transfer functions
Contains a measurement journal of the transfer function measurement of BeyerDynamic DT990 Pro used for the subjective experiment.

APPENDIX E Enclosed DVD contents
Contains a list of material included on the enclosed DVD.

INTRODUCTION

CHAPTER 1

Introduction and problem analysis

The detection and ratings of impulses can be used in wide fields within the industry. An engineer working with detecting impulses and reducing unwanted sounds in consumer products, most prominent job is to prevent all annoying sounds as impulses to be perceived by the costumer. Sound quality is a part of the total product quality (Plunt, 2006).

In the area of NVH engineering vehicle noise can be divided into two groups: The constant type and the more transient like or come and go type. The constant type are caused by engine, road and tire and wind noise. These type of noise should be the first to be eliminated because they are often more annoying and discomforting to the customers. When the constant type of noise is brought to a more acceptable level the transient like or come and go type of noise like Buzz, Squeak and Rattle (BSR) events becomes more prominent and needs to be eliminated (Trapp and Chen, 2008).

In this project, the focus is only on BSR events and not on the constant type of vehicle noise.

Customer perception of BSR events is measured by Things Gone Wrong (TGW), warranty claims and JD power surveys. A market analyse back from 1983 reported that squeaks and rattles were the third most important customer apprehensiveness in cars after 3 months of ownership (Kavarana and Rediers, 2001).

The absence of BSR leads to higher positive feedback from customers of the vehicles build quality. If any BSR is heard by the customer the vehicle is perceived as low quality and will project a negative image to the vehicle manufacturer.

In 2001, S&R issues are estimated to be around 10% of the total TGW costs from manufactures warranty bills (Kavarana and Rediers, 2001). 50% of the total BSR vehicle interior problems are caused by instrument panels, doors and seats where instrument panels (IP) is being the main source (Shin and Cheong, 2010).

BSR events is caused by vibrations due to a combination of the vehicle interaction with the road and to structural deficiencies, incompatible material pairs or poor geometric control(Kavarana and Rediers, 2001). In details squeaks are originating from the elastic deformation of the contact surfaces storing energy, which is released when the static friction exceeds the kinetic friction (Shin and Cheong, 2010). This causes audible squeaks in the frequency region from 200-10000 [Hz]. Rattle is caused by the phenomena when there is motion between to surfaces with short loss of contact. The frequency region of audible rattles are in the range from 200-2000 [Hz]. Higher frequency rattles are perceived as Buzz (Kavarana and Rediers, 2001).

One way of detecting BSR events that is used in industry is to use Road Simulators where the entire vehicle can be placed inside, where on-road driving conditions can be applied. Recordings of test-drives in real-world environments of acceleration and/or strain can then be loaded into the Road-Simulator to test different driving conditions. Since BSR results depends on climatic conditions such as humidity and temperature, these factors can often be controlled too. There have been attempts to detect BSR in terms of A-Weighted sound pressure level but in case of doubts, a human evaluator has generally been used (Cerrato, 2009). In case of a human evaluator several steps needs to be carried out. A typical scenario can look like this:

step 1 Do I hear any BSR events?

step 2 If I hear some, how bad are they?

step 3 How often will this BSR occur?

step 4 Where is it coming from?

step 5 Let´s see if it goes away?
Here the engineer start touching different parts to localise the annoying BSR event.

The question is, which of these steps can be automated? The first 3 steps can be automated using objective evaluation methods. Step can 4 be carried out using microphone arrays combined with cameras to localise the BSR event using techniques such as spherical beamforming (Veen, 1988) and near-field acoustical holography (Hayek, 2009, p.1130). Step 5 will still remain to the expertise and skills of the engineer.

Many attempts has been conducted to automatically detect BSR events from acoustical recordings. In (Feng and Hobelsberger, 1999) they outline that the temporal dynamics and relative low level, are two major difficulties for the analysis of S&R events. Any analysis method that should be considered for detecting S&R events should address those aspects. One psychoacoustic method that is particularly well suited for S&R events is time-varying, or in-stationary, loudness analysis. In this work they suggest that rattle can be detected using N2 percentile levels of the signals loudness histogram

Another approach is seen in (Chandrika and Kim, 2010). Here they describe that previous studies has shown that the wavelet-transform is ideal to characterise highly transient events. For detecting S&R events they use i.a. Wavelet transform & Loudness transformation combined with subjective results to tune the detection level of the algorithm. The algorithm performed well with a minimum of false alarms.

In the two presented studies the detection threshold was determined by a fixed level.

In (Blommer et al., 2005) they developed and impulsive algorithm which correlates well with subjective results and the model was nearly independent of loudness which makes it useful to detect engine such as ticking and knocking.

In (Song and Saito, 2011) an impulsiveness metric was developed. Results revealed that the proposed impulsiveness mapping was more effective compared to traditional SPL and intensity to detect impulsive and rattling sources.

In this project the focus is on detecting impulses in a predefined environment. The environment consists of synthetically constructed impulses where parameters such as level, repetition frequency and impulse length is under control. These impulses are mixed with noise with varying S/N in order to introduce an artificial background noise. It is not intended to mimic all the attributes of BSR-events but to some extend. These samples are analysed in a objective and subjective manner.

At first it was decided to base the objective algorithm to some extend on the work described in (Chandrika and Kim, 2010) (Mentioned previously), but further investigations leaded to another approach, which did not use a fixed detection threshold. Instead the objective evaluation was decided to be based on the Relative Approach algorithm proposed by (Sottek and Genuit, 2005). This model has been chosen since it is independent of absolute level and is able to reject a large part of the pseudo- stationary energy in the signal. From the output of the Relative Approach an impulsive metric proposed by (Sottek et al.) is used to convert it to a single number presenting the amount of Impulsiveness.

Two rating scales are chosen for the subjective evaluation; Impulsivness and Annoyance. The first scale to address the correlation between objective an subjective impulsiveness. The latter to address the correlation between Impulsiveness and Annoyance in order to estimate how annoying a given impulsive sound would be.

Existing literature does not describe a similar study of impulses with the mentioned parameters, conducted in a subjective experiment. That was the motivation to explore the described scenario.

The goal of this project is to reveal if any correlation exist between the objective and subjective evaluation of the samples.

METHODS

CHAPTER 2

System Overview

In figure 2.1 the complete system is shown.

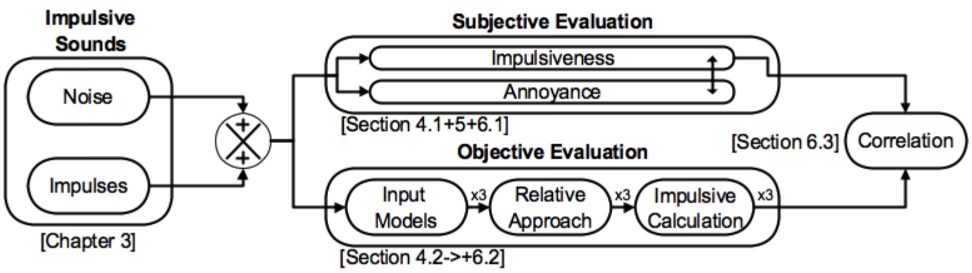


Figure 2.1: *Block diagram of the complete system*

The first two blocks from left to right generates the signal for analyse, where impulses and noise are mixed together. A deeper insight of the two blocks is described in chapter 3.

The following upper block is the subjective evaluation of the samples. Two scales are used for the analyse; Impulsiveness and Annoyance. A description of the used scales is described in section 4.1. A description of the subjects, the equipment and procedure for conducting the experiment is given in chapter 5 and the results are listed in section 6.1 where the correlation between the two scales is evaluated.

The lower block is the objective evaluation. It is divided into 3 parts; Input Models, Relative Approach and Impulsive calculations. The Input Models contains 3 different models to analyse the samples; Time-varying Zwicker Loudness, Time-varying Moore Loudness and Fractional Octave Filtering. The input models are described in section 4.2.3. The Relative Approach algorithm analyses each output from each Input Model. The Relative Approach algorithm is described in section 4.2.1. The last block in the objective evaluation contains an algorithm to calculate the amount of impulsiveness from the each output of the Relative Approach. The Impulsive Calculation is described in section 4.2.2. The results from the objective analyse is given in section 6.2.

The last block in the chain contains the correlation between the subjective and objective evaluation of the samples. The correlation is described in 6.3.

CHAPTER 3

Impulsive Sounds

In this chapter a description of how the samples has been designed to the subjective and objective analyse is explained. First a description about the impulses then about the noise which is mixed together with impulses following with an explanation of the chosen parameters. At last the post-processing of the samples is described.

3.1 Impulses

In order to construct the impulses, hanning windows has been chosen for this purpose. Hanning is normally used as a window function in e.g. a FFT-analysis (Oppenheim et al., 2008, p. 565) but in this project they are used as impulses. A hanning-window has a shape of a raised cosine function, and can be described with the following equation

$$w(n) = 0.5\left(1 - \cos\left(2\pi \frac{n}{N}\right)\right), \quad 0 \leq n \leq N \tag{3.1}$$

where the window length is $L = N + 1$

In figure 3.1 the time and frequency domain of a hanning-window of $L = 200$ is shown

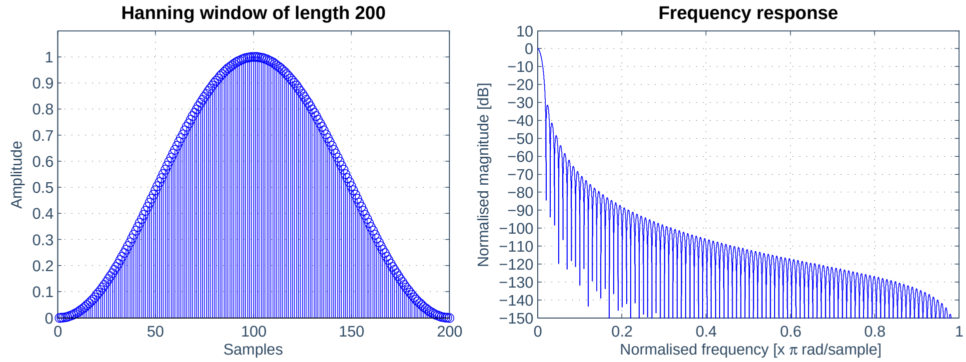


Figure 3.1: *Time and frequency domain representation of a hanning-window of* $L = 200$

Hanning-impulses has been chosen because of the smooth fade in and out caused by the nature of the raised cosine function, which avoids any sharp transients. By varying the length, different frequency responses can be obtained, which changes the pitch of the impulsive sound (International, 2012).

3.2 Noise

In order to mask the impulsive sounds by some degree, pink noise has been chosen for this purpose because the energy in each octave is the same compared to white noise where the energy is the same for each frequency. This causes white noise to sound more high-frequent than pink-noise. Since the perception of impulsiveness is mainly caused be frequencies above 500 Hz, white noise will mask the impulses by a higher degree (Song and Saito, 2011).

The pink noise has been high-pass-filtered at 100 Hz to reduce the amount of low-frequency fluctuations which is illustrated in figure 3.2.

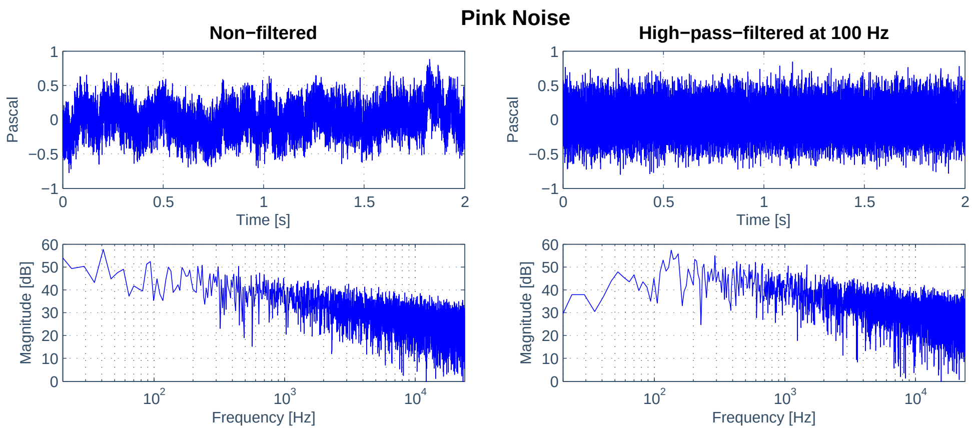


Figure 3.2: *Non-filtered and high-pass-filtered pink noise. Pink noise rolls off with 10dB/decade.*

3.3 Parameters

To achieve a diversity among the samples, several parameters have been chosen. The parameters and the corresponding nomenclature are listed as follows

- Hanning impulse length (H_L) [ms]
- Hanning impulse peak level (H_P) [dBSPL]
- Hanning repetition frequency (H_R) [Hz]

- S/N level ($H_{S/N}$) [dB]

The effect of Hanning impulse length has been described previously in the chapter, that it changes the pitch of the impulsive sound. In NVH-analyse a particular BSR event has sometimes a duration below 10 ms (Shin and Cheong, 2010).

H_P has been chosen instead of RMS level, since the peak level can be fixed, where RMS level depends on the impulse length.

H_R determines how many impulses is given within a second. By varying this parameter from e.g. 2 Hz → e.g. 300 Hz quite different experiences of the sample can be obtained. This is caused by two types of psychoacoustic phenomena; Fluctuation strength and Roughness. Fluctuation strength has its maximum at a modulation frequency of 4 Hz, with the reference sound of a 1 kHz tone at 60 dB SPL, which produces a fluctuation strength = 1 vacil (Bray, 2007). At around 20 Hz, there is a smooth transition between Fluctuation strength and Roughness. The sensation of Roughness has its maximum at 70 Hz, with the same reference tone used for Fluctuation strength, which produces a Roughness = 1 asper. The term is used for modulation-frequencies between 15-300 Hz (Fastl and Zwicker, 2006, p. 247-264). In figure 3.3 an illustration of the term fluctuation strength is shown. ΔL is the temporal-masking depth of the temporal-masking pat-

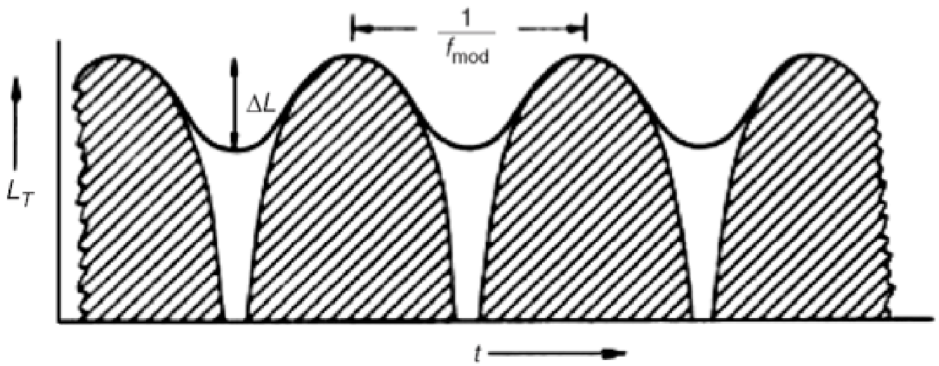


Figure 3.3: *Model of fluctuation strength: Temporal masking pattern of sinusoidal with amplitude modulated masking depth* ΔL *(International, 2012)*

tern. The fluctuation strength can be described by the following equation (Fastl and Zwicker, 2006, p. 254).

$$F \sim \frac{\Delta L}{(f_{mod}/4Hz) + (4Hz/f_{mod})} \tag{3.2}$$

From this equation it is clear that f_{mod} = 4 Hz plays an important role in determining the fluctuation strength. When f_{mod} increases beyond 4 Hz the ear exhibits integrative features as post-masking and below 4 Hz short-term-memory become an important factor.

Roughness is like fluctuation strength a sensation and can be described by the following equation (Fastl and Zwicker, 2006, p. 262).

$$R \sim f_{mod} \Delta L \tag{3.3}$$

When f_{mod} is small, ΔL is big, the product remains small. At medium frequencies around 70 Hz, the product reaches its maximum. When f_{mod} reaches 250 Hz, ΔL is very small and consequently the product turns out to be almost 0 and the sensation of Roughness disappears. Both Fluctuation strength and Roughness have been derived from subjective experiments.

The last parameter, $\mathrm{H_{S/N}}$, determines the level difference between the hanning peak level and the RMS-value of the high-pass-filtered pink noise. In appendix A the different settings for each parameter used for the listening experiment is discussed.

All samples where created using a sampling frequency of 48 kHz.

3.4 Post-processing of samples

In order to avoid any sudden transients a smooth fade in and out, of 100 ms each, has been applied to each sample. This attribute is shown in figure 3.4

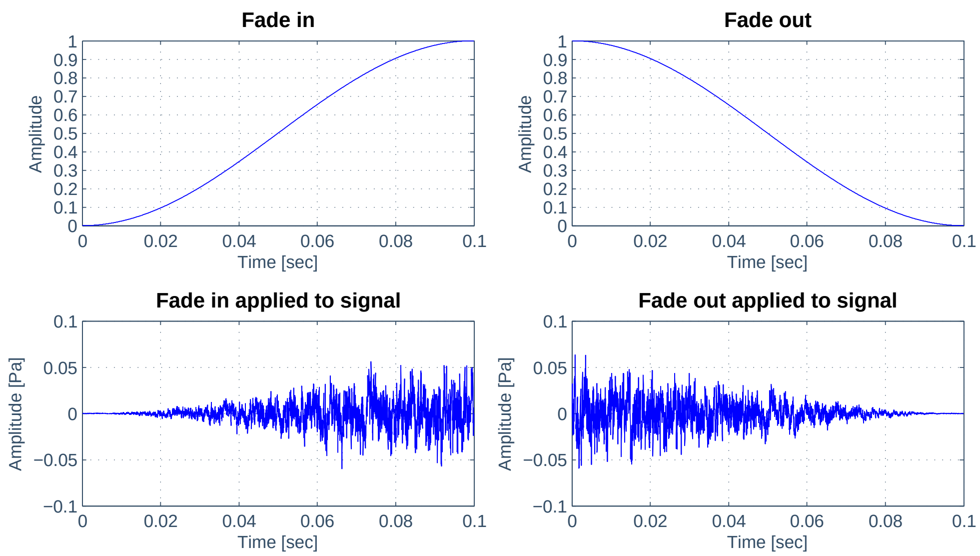


Figure 3.4: *Fade-in and out functions and fade-in and out applied to signal*

3.4.1 Inverse filtering

In section 5.2 BeyerDynamics DT-990 Pro (2x250Ω) has been used to playback the samples to the subjects. To ensure an approximately flat-frequency response at the listener, inverse-filters of the transfer function from BeyerDynamics → Ear has been utilised by using minimum-phase theory from an approach given by (Hawksford, 1997)[1]. In appendix D the measurement procedure for measuring the transfer function from BeyerDynamics → Ear is described. The measured transfer function is shown in figure 3.5. In order to avoid boosting low frequencies from the measured transfer function when doing the inverse, frequencies below 20 Hz will not be compensated and a cubic-spline-interpolation has been applied from 20-50 Hz to ensure a smooth transition to 0 dB. As described in appendix D human variations across subjects above 7 kHz is highly individual and a therefore not compensated (Blauert et al., 2005, p. 231). Again the same smooth transition is applied, now from 5-7 kHz and the inverse filter is shown in figure 3.6.

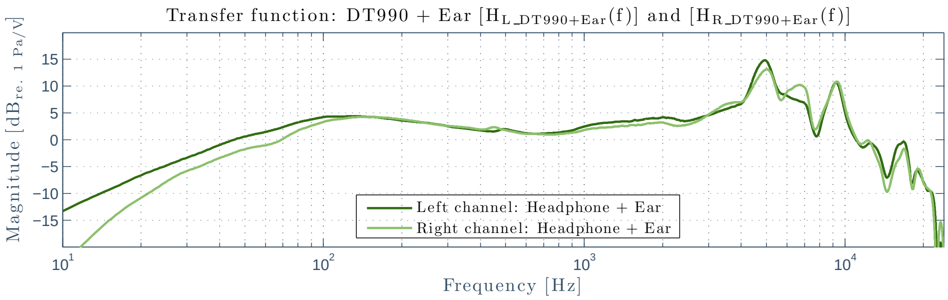


Figure 3.5: *Left and right channel transfer functions of $H_{DT990} + H_{Ear}$*

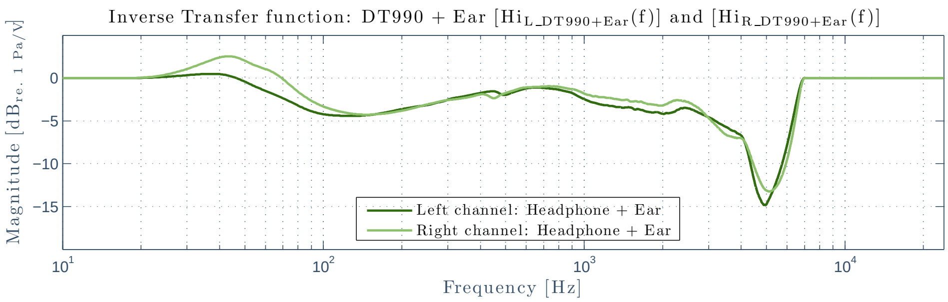


Figure 3.6: *Left and right channel inverse transfer functions of $H_{DT990} + H_{Ear}$*

A convolution between the original and the inverse filter gives the resulting response which is shown in figure 3.7. As seen the resulting transfer functions are almost completely flat in the frequency range from 50 Hz → 5 kHz.

[1]The code can be found on the enclosed DVD `/Code/Minimum_Phase_Filter/mpf2go.m`

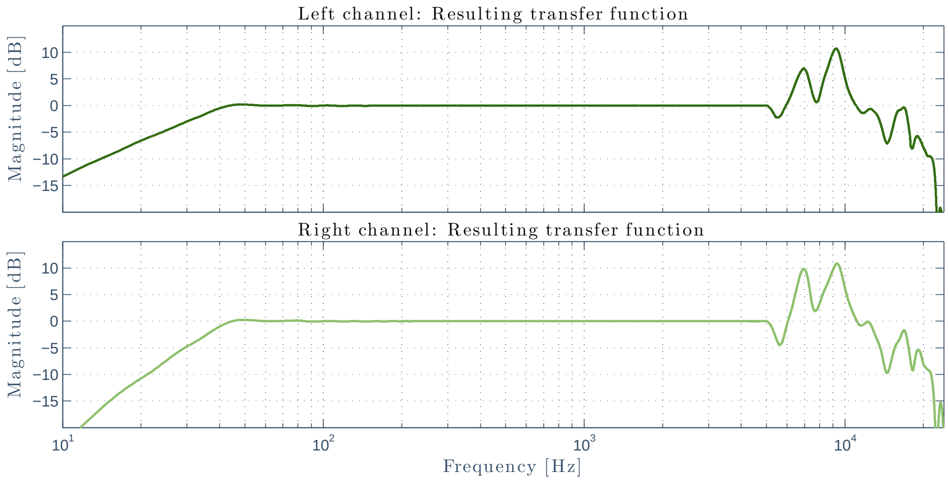


Figure 3.7: *Left and right channel resulting transfer functions of H_{DT990} + H_{Ear}*

3.4.2 Level adjustment

To ensure that the levels of the samples equals the presented levels to the subjects the output has been scaled to full-fill the requirement. The experimental setup is the same as shown in figure D.1 in appendix D

First the sensitivity of the build-in microphones of Valdemar (Christensen et al., 2000) was recorded in the computer as a digital RMS-value and denoted as $\mathrm{Mic_{scalefactor}}$. The received level is then described by the following equation

$$SPL_{in} = 94 + 20\log_{10}\left(\frac{D_{inRMS}}{Mic_{scalefactor}}\right) \tag{3.4}$$

A 1kHz test-tone of 74 dB SPL RMS was then send out from the computer, and the corresponding $\mathrm{D_{inRMS}}$ was recorded, and the $\mathrm{SPL_{in}}$ was calculated. If the measured $\mathrm{SPL_{in}}$ was not equal the test-tone, the channel was scaled according to the level difference.

CHAPTER 4

Methods for evaluating the samples

In order to evaluate the samples described in chapter 3 two methods are used. The first method is a subjective evaluation where the subjects has to evaluate the samples in an interactive manner. The second method is utilised by an objective algorithm written in Matlab which analyses the samples in several steps.

4.1 Subjective evaluation

In order to analyse the samples in a subjective way, it has been decided to use 2 different attributes to describe them. The first attribute has been chosen to be comparable with the way the objective algorithm (section 4.2) delivers its result. This attribute is denoted as "Impulsiveness" and describes how impulsive an acoustic signal is interpreted. "Impulsiveness" is not a familiar term for a non-acoustician and needs to be explained for each subject. An explanation of how the "Impulsiveness" attribute is conveyed to the subjects is given in 5.3.

The second attribute is "Annoyance" and has been chosen in order to find a connection between "Impulsiveness" and "Annoyance". If a sample is rated as "Extremely impulsive" and "Not at all Annoying" at the same time, it may not be a problem if the sample is highly impulsive.

In (Davies et al., 2009) different category scales are used to evaluate a soundscape. In this article they uses category scaling with verbal descriptors at both ends of the scale without verbal labels in between. The most "negative" descriptor is placed at the left anchor point and the most "positive" to the right anchor point. This way of subdivision (most "negative" to the left and most "positive to the right") is the same as seen in commonly employed rating scales (Bech and Zacharov, 2006, p. 74) and is therefore used in this project.

Instead of having a fixed number of categories within in the scale where the subjects can mark, it has been decided to use infinite infinite number of categories or namely a continuous scale[1] with no endpoints to reduce

[1] Because of the digital discretisation in the computer, it would not be completely con-

"lumping" at the end points.

A visualisation of the two scales is shown in figure 4.1. The scales ranges from -0.1→1.1. Only one scale was used at the time to avoid any artificial correlation between them.

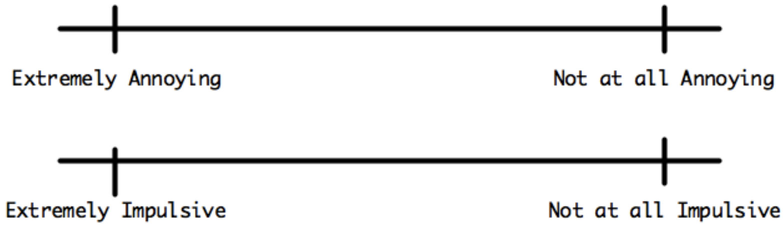


Figure 4.1: *Annoyance and Impulsiveness scale*

4.2 Objective evaluation

In figure 4.2 the Block-diagram of the objective algorithm is shown. First a stimuli is presented and then processed through 4 consecutive steps. The output is a scalar between 0 and 1. First the Relative Approach algorithm is described in section 4.2.1, then the Calculation of Impulsiveness and normalisation in section 4.2.2 following with the Input models in section 4.2.3. At the end in the chapter the different steps in the objective evaluation is tested in section 4.2.4.

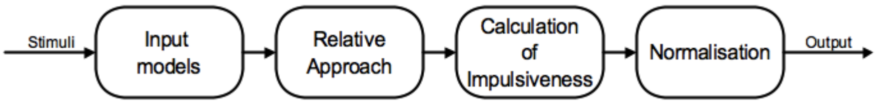


Figure 4.2: *Block-diagram of the objective algorithm*

4.2.1 Relative Approach

Introduction

Human hearing is able to detect slightly differences between two samples in an A/B-comparison test. In everyday life the human continuous evaluates the acoustic quality of sounds in the absence of the reference sound. An acoustic event is rated as annoying in terms of the time and frequency content and will remain annoying even if the event is attenuated (Genuit, 1996). Humans are not able to take absolute level into account but are assumed to create is own running reference anchor point based on the acoustic event of the tonal and temporal information moment-by-moment (Sottek et al., 2005).

tinuous.

The relative approach was developed to mimic the largely level independent aspects of human hearing described above (Genuit and Bray, 2008). It extracts the patterns of the signal while rejecting a large part of the pseudo-stationary energy (Sottek et al., 2005). Without a pattern the relative approach will be zero (Genuit and Bray, 2008). In the next section the algorithm is explained.

Algorithm description

In figure 4.3 a block-diagram of the relative approach algorithm is shown. The left part of the block-diagram is for analysis of tonal components and the right part for transient signals. In this project the focus is on detecting impulsive events which leads to $\lambda_1 = 0$. Input too the algorithm can be any type of filtered signal from e.g. fast-fourier-transform (FFT), fractional-octave-band-filters, wavelet-transform etc. or from hearing models. One type of hearing model is proposed by (Sottek, 1993). Two other types are the time-varying Moore loudness and time-varying Zwicker loudness (DIN, 2007) and (Glasberg and Moore, 2002). In this project three types of inputs will be used and explained further in this section.

The chosen inputs are:

- Fractional-octave-band-filters
- Time-varying Moore loudness
- Time-varying Zwicker loudness

Regression vs. time for each frequency band according to (ETSI, 2011, p. 21), the linear regression is performed from the past 200 [ms] for each frequency band (figure 4.4a).

The smoothing operation vs. frequency according to (ETSI, 2011, p. 21), is performed by a linear regression from the 16 neighboring frequency band for each time-slot (figure 4.4b).

The non-linear transformation according to Hearing Model of Sottek ((Sottek, 1993)) is only applied for input signals in [Pa] or sound pressure level (SPL), not on the hearing models described earlier. The compression turns the sound pressure into perceived loudness, and the term "compressed pressure" in compressed Pascal [cPa] is used as the physical unit(ETSI, 2011, p. 22). The compression is performed in the way shown in figure 4.5.

After the Non-linear transformation[2] the vs. time g_2 is subtracted from the vs. frequency f_2 and the absolute value is taken. At last sub-threshold values are set to zero.

The output of the Relative Approach Analysis is a matrix for time and frequency patterns $\mathbf{RA}(\mathbf{t},\mathbf{f})$.

[2] If the input to the relative approach is not a hearing model.

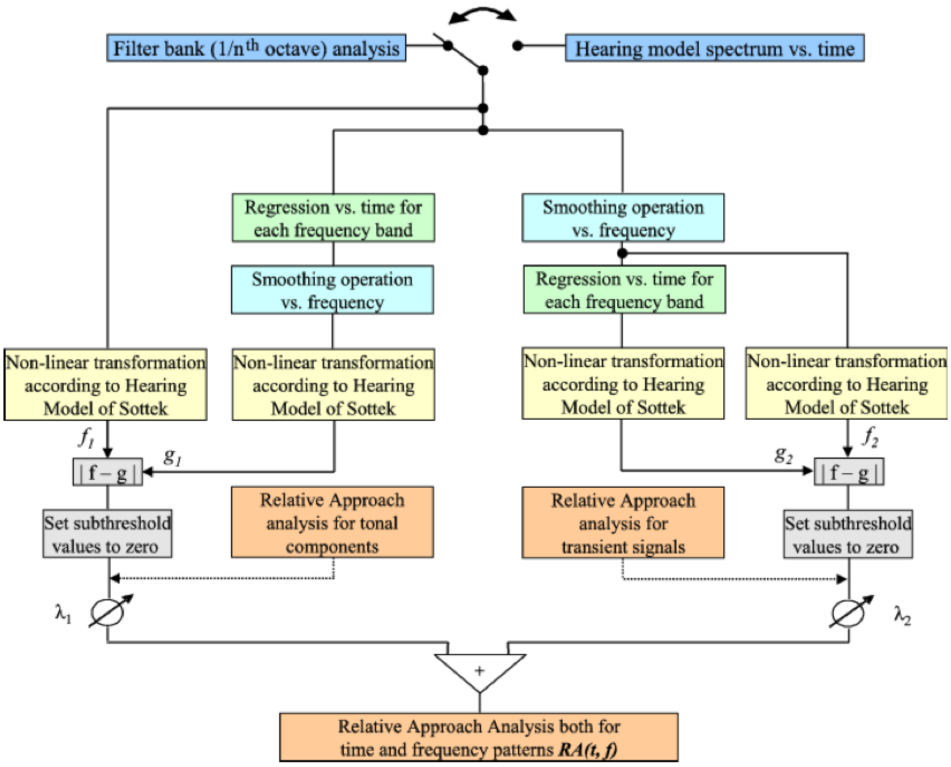


Figure 4.3: *Block-diagram of the Relative Approach Algorithm (ETSI, 2011, p. 20).*

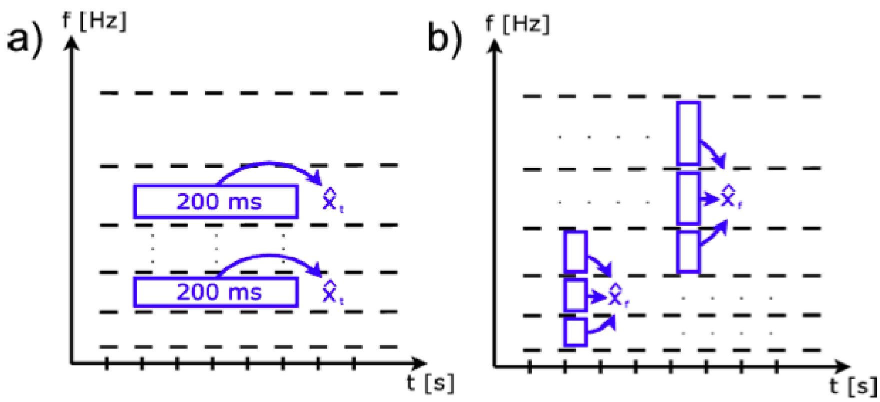


Figure 4.4: *Linear Regression used in the Relative Approach algorithm: a) Vs. time of the past 200 [ms] for each frequency band, b) Vs. frequency from the 16 neighboring frequency band for each time-slot (ETSI, 2011, p. 21).*

4.2.2 Calculation of impulsiveness

To extract a single number of the level of impulsiveness from the output of the relative approach analysis $\mathbf{RA(t,f)}$, an approach described in (Sottek et al.) has been used. The equation to calculate the impulsiveness is shown

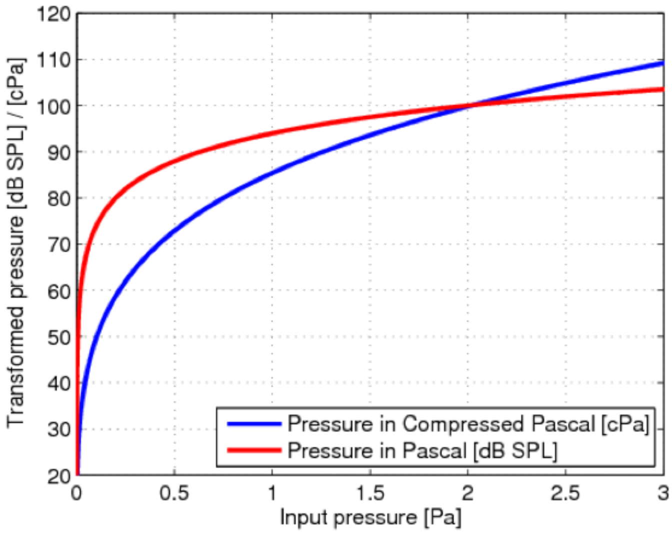


Figure 4.5: *Non-linear transformation according to Hearing Model of Sottek (ETSI, 2011, p. 22).*

in equation 4.1.

$$I = \left(\sum_{j=f>500}^{N} k_j \left(\frac{\overline{(RA_{(:,j)} - \overline{RA_{(:,j)}})^n}}{\overline{RA_{(:,j)}}^m} \right)^p \right)^q, \text{ where} \tag{4.1}$$

j is the index of each frequency-band up to the total of N bands. Only frequencies above 500 Hz are included since the perception of impulsiveness is mainly caused be frequencies above 500 Hz as described in section 3.2. k_j is a frequency-band dependent weighting factor. The appropriate factors are not available in the paper. When writing $RA_{(:,j)}$ it means that for each j, all time-slots within that j is taken into account. The bar "–" is the mean value. n should by rule be $n = 2$, but $n = 4$ can also used. $n = 2$ is used in this project. With $n - m = 0,25$ the results from the hearing model will be set into a non-linear behavior (Sottek et al.) & (Sottek, 1993). This gives $m = 1,75$. The last two exponents p & q should be set to appropriate values between 0 and 1, with 1 included. These two factors adds compression to either each j or the summed I. p & q are not included in the equation described in (Sottek et al.) but are values used in this project to adjust the calculated impulsiveness. At the end the Calculated Impulsiveness is normalised to give values between 0 and 1.

4.2.3 Input models to the Relative Approach algorithm

Fractional octave filters

In section 4.2.1 three input models to the relative approach was chosen. One of the models where fractional filters. The reason why to choose this model compared to e.g. FFT is that time-domain filter-banks is generally more suitable to detect squeak and rattle events (Feng and Hobelsberger, 1999). The reason is that when doing a FFT there is always a trade-off between having a good time or frequency resolution which can be seen in the formula below

$$T = \left(\frac{N}{F_s}\right), \text{ where} \tag{4.2}$$

T is the associated duration of the FFT-analyse, N is the block-size in samples and F_s is the sampling frequency. If e.g. $N = 4096$ and $F_s = 48\text{kHz}$ the frequency resolution F_{res} will be:

$$F_{res} = \left(\frac{48000}{4096}\right) \approx 12 \qquad [\text{Hz}] \tag{4.3}$$

The corresponding time-resolution will be:

$$T = \left(\frac{4096}{48000}\right) \approx 85 \qquad [\text{ms}] \tag{4.4}$$

If the BSR event has a much shorter duration, than this time-resolution, it will be smeared out and drown onto the background noise

Fractional octave-filter (FOF) has the advantage that the frequency resolution of the constant percentage bandwidth will be the same independent of the chosen window-length of analyse but there are some precautions that the designer should take care of. A problem that can arise is if the filters are designed to have a cut-off frequencies close to 0 or $Fs/2$, which can cause inaccurate results. A way to solve this problem for the high-frequency-filters is to move the highest cut-off-frequencies down in frequency or use a higher sampling frequency. For the low frequency part, downsampling can be applied. A proper use of downsampling allows for re-using the same filter-coefficients used for e.g the highest octave, to all descending octaves.

From (ETSI, 2011, p. 22) some design-rules are mentioned about designing FOF to adopt the Relative Approach for speech analysis. The variables $\lambda_1 = 0$ and $\lambda_2 = 1$ was chosen for this analysis (see figure 4.3). Thus the model is suited for detecting transient events. $1/12^{th}$ FOF according to (ANSI, 2004) is used in the described model with a time resolution of $\Delta t = 6.66$ [ms] (150 blocks per second). The frequency range from 15 Hz to 24 kHz is divided into 128 frequency bands Δf_m which corresponds to a $1/12^{th}$ octave resolution.

In this project 120 $1/12^{th}$ FOF from 20 Hz to 20 kHz is used. With $\Delta t = 6.66$ [ms]. As described in 3 the samples are generated with $F_s = 48$ kHz. This yields (equation 4.5) 320 samples per block.

$$N_{samples} = 48000\left(\frac{6.66}{1000}\right) \approx 320 \qquad [\text{samples}] \tag{4.5}$$

120 $1/12^{th}$ FOF from 20 Hz to 20 kHz spans over 10 octaves. They are successfully implemented by first creating the filter-coefficients for the 2 highest octaves (5-10 kHz, 10-20 kHz), then apply 4 times downsampling and

re-use the filter-coefficients for the next 2 descending octaves, then apply 4 times downsampling and so on. For each filter a 2^{nd} butterworth characteristic has been used. The FOF response is shown in figure 4.6

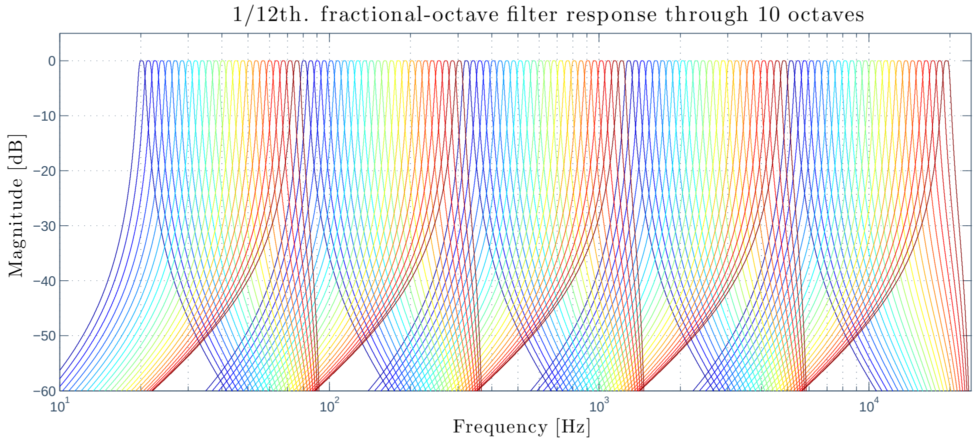


Figure 4.6: *1/12th. fractional-octave filter response through 10 octaves.*

Loudness models

The two other input-models chosen to the RA-algorithm are; Time-varying Zwicker Loudness and Time-varying Moore Loudness. Both models tries to mimic how human perceive loudness. They depend on the sound level, frequency and duration. They have been chosen instead of the hearing model proposed by Sottek (Sottek, 1993), because of time-limitations and online-toolboxes for calculating Loudness already existed which will be explained at the end of this section.

One attempt to mimic loudness was introduced in 1936 by Stevens in the unit of Sone. The scale is obtained by asking subjects to put a number proportional to the stimuli loudness presented for several stimuli tones at different levels (Genesis, 2009). The level of 40 dB of a 1-kHz tone was used to obtain a reference sensation, i.e. 1 Sone. A tone that is perceived twice as loud would have the value of 2 Sone. This is only true for sines with levels of 40 dB SPL or above. For more complex sounds and levels below 40 dB SPL this relationship does not hold (Fastl and Zwicker, 2006, p. 206).

Another attempt to mimic loudness was introduced in the twenties by Barkhausen, where a shortening of his name, Bark, was used as a physical unit for the critical-band rate. The loudness of a sound was determined by comparing it to a 1kHz tone in a plane wave at frontal incidents that was similar loud as this sound, and the physical unit is phon. It can be measured for any sound but best known for frequencies of pure tones. Several experiments has shown that this holds for durations above 500 [ms]. In figure 4.7 the equal loudness contours is shown. The threshold seen in the figure corresponds to the threshold at quiet to 3 dB at 1 kHz. This equal-loudness contour is indicated by 3 phon (Fastl and Zwicker, 2006, p. 204).

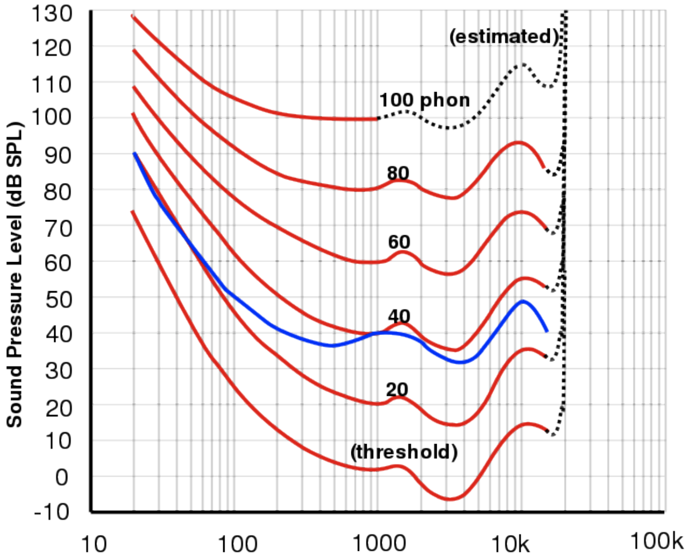


Figure 4.7: *Equal-loudness contours (red) (from ISO 226:2003 revision). Original ISO standard shown (blue) for 40-phons (Mus, 2012)*

The equal-loudness contour at 40 phon is used to measure the physical quantity dB_A which takes into account the variations of the ear-sensitivity across frequency and was developed to measure loudness. An important notice is that dB_A does not take into account physiological phenomena such as frequency masking or the filter bank of human ears, and are thus insufficient to correctly estimate loudness (Genesis, 2009).

On the basis of this new methods was needed to estimate loudness. The first model was introduced by Zwicker in 1958 designed for stationary sounds. In 1996 Moore´s model was published. Concerning time-varying sound, where temporal aspects is taken into account, two new models where developed. First by Zwicker and Fastl in 1999 and then by Glasberg and Moore in 2002.

The steps for archiving the time-varying loudness of Zwicker and Moore will not be explained in details here, since the models are quite comprehensive but a brief overview will be given. The reader are referred to (DIN, 2007) and (Glasberg and Moore, 2002) for a deeper insight in the models.

Loudness for time-varying sounds

In figure 4.8 the different stages for calculating the Loudness for time varying sounds is shown.

Both Zwicker and Moore model does corrections to take into account the transfer function of outer/middle ear.

Zwicker converts the stimulus into 1/3 Octave-filters in time windows of 2 [ms]. Moore uses 6 parallel sliding FFT windows, with window sizes from 2 [ms] → 64 [ms] with steps of power two. This procedure is done for each 1 [ms]. Afterwords the excitation in the critical band filters is estimated.

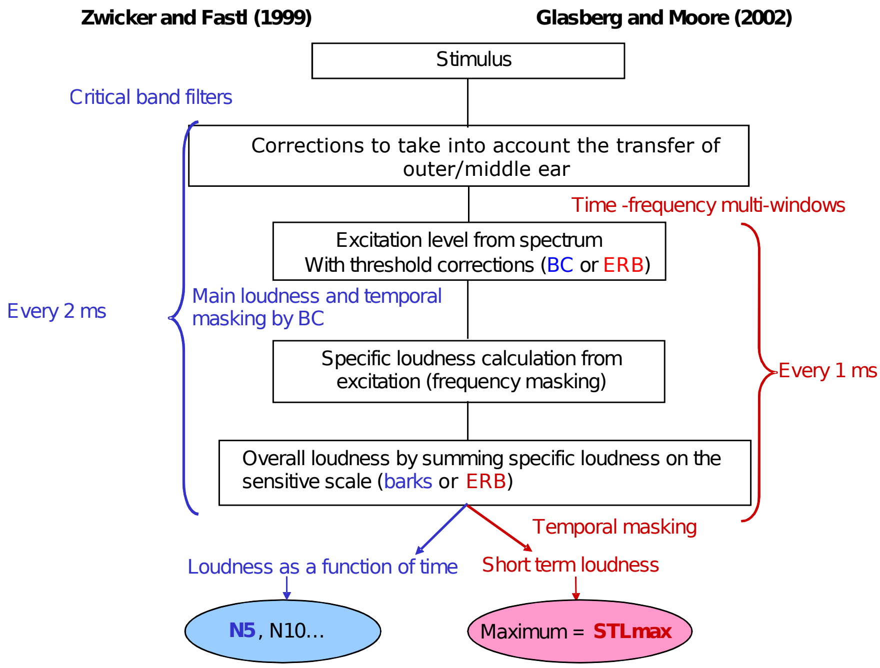


Figure 4.8: *Different stages of the calculation of Loudness for time-varying sounds (Genesis, 2009)*

The critical-band filters have been derived from several subjective experiments and describes the human filter sensitivity along the basilar membrane. The bark scale is used to represent them and there exist 24 barks. Each critical-band are overlapping and increases in band-width with frequency except from bark 2→4 where the band-width is equal. Moore uses another name to model the auditory filters, namely the Equivalent Rectangular Bandwidth (ERB) bands, which are close related to the critical bands (Genesis, 2009).

Next the spectral masking is calculated and the overall specific loudness is summed. Zwicker uses table-look-up where Moore uses an analytical formula.

Zwicker models the temporal masking with an electrical circuit with a time constant of 100 [ms]. It does only take post-masking into account, not pre-masking. Moore models the temporal masking by an automatic gain control (AGC) with an attack and release-time constant and does account for both post and pre-masking. Two different settings is used; Short-Time-Loudness (STL) and Long-Term-Loudness (LTL) each with different attack and release time constants. LTL is calculated from the STL (Glasberg and Moore, 2002) and STL is calculated from the Instantaneous Loudness.

In this project the Genesis Loudness toolbox has been used to evaluate loudness of time-varying Zwicker and Moore (Genesis, 2012).

4.2.4 Test of algorithms

In this section the Relative Approach Algorithm, Calculation of Impulsiveness, Fractional-octave-filters and Loudness models are tested, to see if their respective outputs looks like. Three different samples from the list described in appendix A has been used to test the algorithms. The 3 test samples are:

- Sample 8: $H_P = 60$, $H_{S/N} = 10$, $H_R = 10$ and $H_L = 1$
- Sample 72: $H_P = 80$, $H_{S/N} = 20$, $H_R = 20$ and $H_L = 10$
- Sample 79: $H_P = 80$, $H_{S/N} = 100$, $H_R = 4$ and $H_L = 0.1$

First the 3 input-models to the Relative approach are tested, then the Relative Approach Algorithm and at last the Calculation of Impulsiveness.

Test: Input models

In figure 4.9 [3] their respective output is seen of the 3 samples. In sample 8, no seems not to be able to detect any pulses because of the high background noise. In sample 72, Zwicker seems to be superior but to the two other models also detects the structure of the pulse-train. In sample 79, all models are able to detect the impulses which is expected because of the S/N-value.

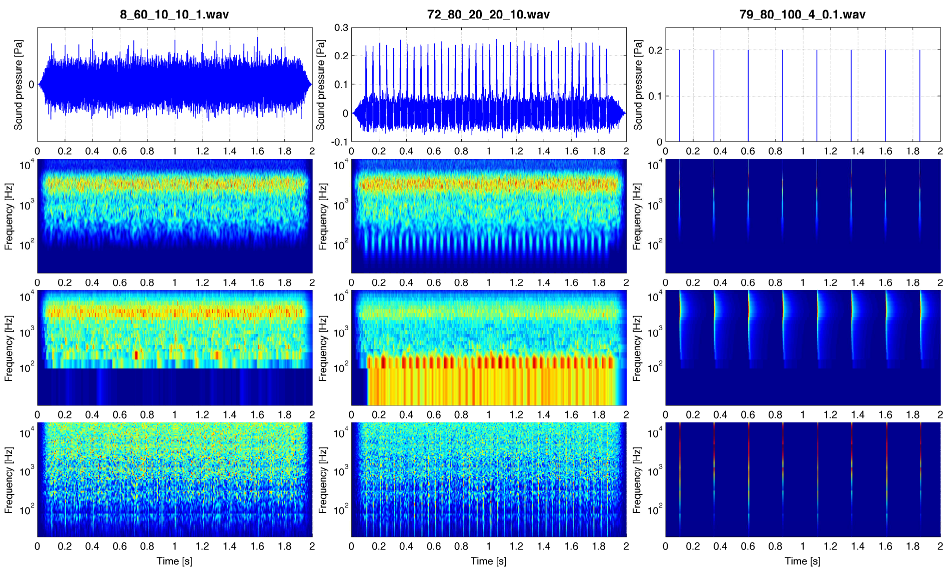


Figure 4.9: *Frequency-output through sample 8, 72 and 79 of Moore (2nd row), Zwicker (3rd row) and Octave (4th row). The first row shows the time-domain plot of the sample of interest.*

[3]The figure can be found on the enclosed DVD `/Plots/Misc/Multi_Freq_8__72_79.pdf`

Test: Relative Approach

These outputs from the input models are feed into the Relative approach algorithm and the output is shown in figure 4.10 [4]. The artifacts seen at least in the Moore and Zwicker situation are caused by the settling time of the Linear-regression. A sub-threshold of 30% of the maximum value in the Relative Approach output has been used remove noise. This maximum value is found in the part between the artifacts. In sample 8 only Octave seems to detect a minor part of the pulses. Zwicker also slightly detects some of the pattern. In sample 72 all models are able to detect the structure but a lot of noise is present. In the last sample all models detects the impulses as expected.

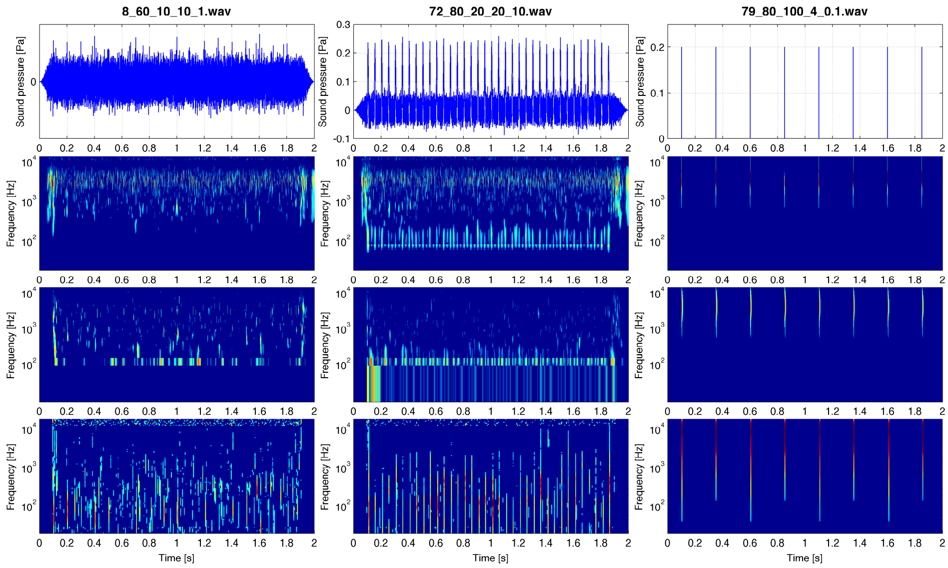


Figure 4.10: *Relative Approach output through sample 8, 72 and 79 of Moore (2nd row), Zwicker (3rd row) and Octave (4th row). The first row shows the time-domain plot of the sample of interest.*

Test: Calculation of impulsiveness

In table 4.1 the calculated impulsiveness of the Relative Approach output of all models trough the 3 test samples is shown. As seen all models exhibit some kind of consistency within the same sample. Sample 79 reaches the highest average rating of the tested samples. In section 6.2 the impulsiveness through all samples is shown.

[4] The figure can be found on the enclosed DVD `/Plots/Misc/Multi_RAout_8__72_79.pdf`

	Calculated Impulsiveness		
Model	Sample: 8	Sample: 72	Sample: 79
Moore	0.7987	0.7529	0.9244
Zwicker	0.8625	0.8625	0.8571
Octave	0.8386	0.8544	0.9009

Table 4.1: *Calculated compressed and normalised impulsiveness values, when $p > q$ in equation 4.1, of sample 8, 72 and 79.*

CHAPTER 5

Experiment

This chapter contains information about the psychoacoustic experiment conducted in order to obtain subjective ratings of the samples. Its sections include:

Subjects
Their age, gender and study/work

Experimental setup
Experimental setup and the equipment used for the experiment.

Experimental procedure
Design of the experiment, GUI design and subjects formalities. Pilot test, the duration and comments from subjects.

5.1 Subjects

The subjects that participated in the experiment were 14 young people[1], in the age between 21-30, all of them are students, their field of study ranging from electronic engineers, business, psychology, acoustics, psychology & design psychology. 3 and 11 were female and male respectively. The nationalities were mixed but most were Danes.

5.2 Experimental setup

For conducting the experiment 4 rooms on Aalborg University were used: Waiting room, Listening Cabin B, Control room A and Audiometry. The waiting room were used to give instructions and for coffee breaks. Listening Cabin B for Audiometry and the Control room A and Audiometry were used for conducting the listening experiment. The equipment used are listed below.

[1]More were called in, but 3 did not pass the audiometry which was mandatory to enter the listening test.

Equipment

Listening Cabin B

- Madsen Electronics Orbiter 922 Version 2 (Headphones included) (AAU nr.: 33968)

Figure 5.1: *Listening Cabin B: Setup used for Audiometry*

Control room A

- Rotel Six channel power amplifier RB-976MKII, with fixed gain-setting of 0dB (AAU nr.: 33978)
- Edirol USB Audio Capture UA-25EX (AAU nr.: 78377)
- 20 dB Attenuator box (AAU custom made)
- Fluke 37 Multimeter (AAU nr.: 08285)
- Linux (Ubunto) computer, with Matlab R2010b, PortAudio and Playrec preinstalled (AAU nr.: 61158)
- Fujitsu Siemens 19" CRT Monitor (AAU nr.: 53100)
- Mouse & Keyboard
- Bouyer PC 1153 Master Intercom (3 channel, 7 buttons) (AAU nr.: 2156-02)
- ATEN USB KVM Extender Local CE700AL (AAU nr.: 2153-6)

Audiometry

- BeyerDynamics DT-990 Pro (2x250Ω) (AAU nr.: 2036-77)
- Samsung SyncMaster 152N 15" LCD Monitor (AAU nr.: 60872)
- Mouse
- Bouyer PC 1101 Slave Intercom (2 buttons) (AAU nr.: 2156-02)
- ATEN USB KVM Extender Remote CE700AR (AAU nr.: 2153-6)

Figure 5.2: *Control room A: Setup used for controlling the listening test*

Figure 5.3: *Audiometry: Setup used for listening test*

5.3 Experimental procedure

When designing the experiment it was decided that a session for each subject should not last more than 1 hour, including introduction and breaks, since they were not paid for their participation. This made restrictions on how many samples it were possible to present during the experiment.

The samples were generated synthetically on the computer using Matlab to have full control of each parameter and its settings. In appendix A the different settings for each parameter used for the listening experiment is discussed and the effective testing time is estimated.

From appendix A the total number of samples was calculated to be 360, which were divided into two parts with 180 samples in each. In the first part the subject had to rate the samples with the annoyance scale shown in figure 5.4.

Figure 5.4: *Annoyance scale. GUI screenshot*

The second part with the impulsiveness scale shown in figure 5.5.

Figure 5.5: *Impulsiveness scale. GUI screenshot.*

Each part consisted of 2 sessions with 90 samples in each. A break of 2 min was placed in between. In each session the samples were randomised to avoid any other effects between subjects. To get the subject familiar with the task, a short training session of 4 samples were conducted before each part. Half of the subjects followed this pattern with the first part in the beginning followed by the second part. The other half of the subjects started with the second part and then the first part. Again to average out any order effects. A break of 5 minutes was placed between the two parts.

Prior the listening test an audiometry were conducted to ensure that the subjects did not have a pronounced hearing loss (HL) in any of the tested frequencies[2].

Before the listening test experiment, they were given a short introduction consisting of a written instruction, which can be found in appendix B.1. If they felt uncomfortable or did not want to carry out the experiment they could by any time terminate the test.

In the training session the subjects was first explained how the term is understood. In case of annoyance it was easy since the term is well known. In case of impulsiveness it required more explanations since the term is not so

[2] 250, 500, 1000, 2000, 4000, and 8000 [**Hz**]. A hearing loss of 20 dB or lower at each frequency was required to enter the listening test.

familiar. An impulsive sound was described as a sound similar to the noise produced by a pile driver at a construction site. A non-impulsive sound was described as the sound of running water. Those two examples was the first two samples in the training session[3]. The last two samples were samples similar to those in the listening test. The annoyance training session consisted of 4 samples similar to those in the listening test. The subjects were told that they did not have to think a lot about were to rate the sample since it was their spontaneous perception that was final.

Matlab was used for collecting data from the participants with an interactive graphical user interface shown previously in figure 5.4 and 5.5 [4]. The scales used were continuous analog visual scales as described in section 4.1. When the subject had rated the sample the following sample would be played.

After the two listening parts some of the subjects were asked what they felt most annoying, impulses or noise? The answers are listed in appendix B.2.

In appendix Athe effective testing time was estimated to be 24 minutes for conducting both listening parts. In practical the effective testing time was measured to be approximately 19:08 minutes, where each session had an average duration of about 4:47 minutes[5]. The audiometry had an average duration of 14:38 minutes. All in all, the whole experiment had an duration between 45-60 minutes.

In section 6.1 the results from the subjective experiment is described.

[3]Samples can be found on the enclosed DVD `Code/GUI/`

[4]GUI can be found on the enclosed DVD `Code/GUI/`

[5]The duration was measured of about half the subjects.

RESULTS

CHAPTER 6

Results

This chapter summarises the results of the psychoacoustic experiment, described in section 5. The chapter is divided into 3 main sections, which are:

Subjective results
: Results from the psychoacoustic experiment (with all the subjective ratings shown).

Objective results
: Results from the objective calculation of Impulsiveness.

Correlation
: Correlation between subjective and objective results.

6.1 Subjective results

In order to give an illustration of how the samples has been rated, three samples has been chosen where each exhibit different trends. The samples are:

- Sample 30: $H_P = 60$, $H_{S/N} = 20$, $H_R = 70$ and $H_L = 10$
- Sample 60: $H_P = 80$, $H_{S/N} = 10$, $H_R = 70$ and $H_L = 10$
- Sample 88: $H_P = 80$, $H_{S/N} = 100$, $H_R = 70$ and $H_L = 0.1$

In figure 6.1, 6.2 and 6.3 the ratings of both repetitions of sample 30, 60 and 88 is shown. The scales has been reversed in order to have a more convenient look such that the extreme value corresponds to the highest value on the scale. As seen in figure 6.1 the ratings on the annoyance scale is quite spread but there is a trend towards the "Not at all annoying" part. The Impulsive scale has been rated less scattered than the Annoyance scale and the trend is towards the "Not at all impulsive" part. When looking at how the individual subjects rated the sample across repetitions the highest consistency is seen for the Impulsiveness scale. E.g. subject number 11 rated around 2.7 in the first repetition and around 8.5 in the second repetition for the Annoyance scale.

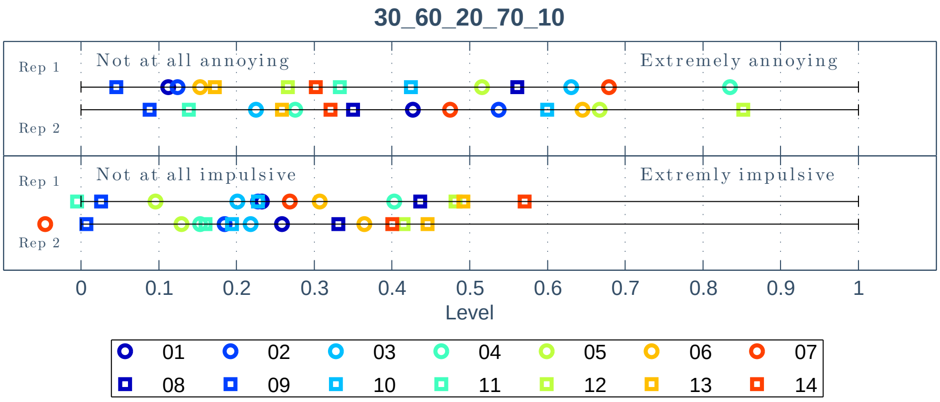


Figure 6.1: *Subjective rating of Sample 30:* $H_P = 60$, $H_{S/N} = 20$, $H_R = 70$ *and* $H_L = 10$. *Ratings at both repetitions are shown. Each circle/square corresponds to one subject.*

In figure 6.2 the data is much less scattered in the use of the Annoyance scale. The decrease of background noise compared to sample 30 has moved the ratings of Annoyance towards the upper part. Impulsiveness has too been rated more to the right part of the scale with two outliers in each repetition.

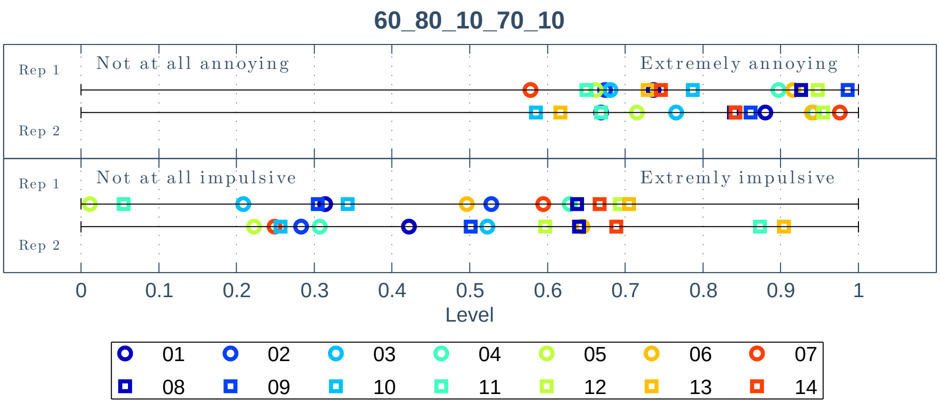


Figure 6.2: *Subjective rating of Sample 60:* $H_P = 80$, $H_{S/N} = 10$, $H_R = 70$ *and* $H_L = 10$. *Ratings at both repetitions are shown. Each circle/square corresponds to one subject.*

In figure 6.3 the sample that exhibits the highest ratings towards the "Extremely annoying" part is seen. The combination between the high $H_{S/N}$, high H_R and low H_L are perceived very annoying between subjects. The sample is too rated "Extremely impulsive" with some few outlier[1].

A more general view of all the responses through all samples is seen in the box-plots, figure 6.4 and figure 6.5. Figure 6.4 shows all $H_P = 60$ results from

[1]The ratings through all samples can be found on the enclosed DVD `/Plots/Sample_Ratings/`

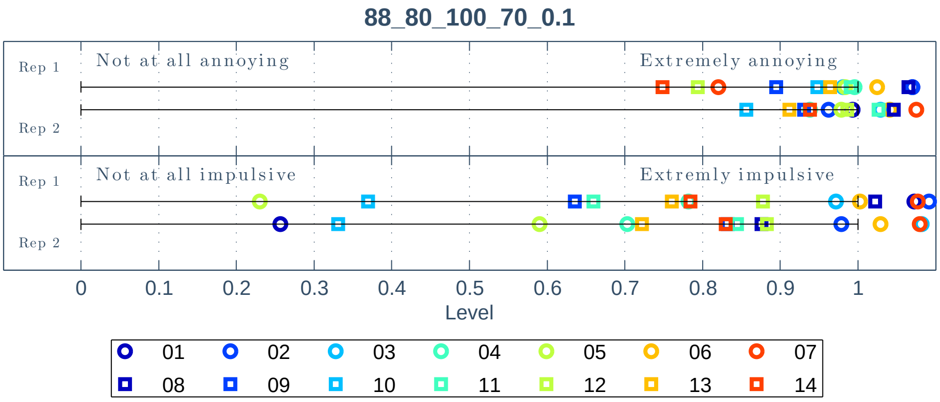


Figure 6.3: *Subjective rating of Sample 88:* $\mathrm{H_P} = 80$, $\mathrm{H_{S/N}} = 100$, $\mathrm{H_R} = 70$ *and* $\mathrm{H_L} = 0.1$. *Ratings at both repetitions are shown. Each circle/square corresponds to one subject.*

sample 1→45 and figure 6.5 shows all $\mathrm{H_P} = 80$ results from sample 46→90. After each 15 sample, the $\mathrm{H_{S/N}}$ changes from 10→20 and 20→100. In both plots, an average between each repetition has been done in order to reduce data-size, since repetition 1 and 2 both are correlated with an p-value $\ll 0.05$[2]. The last subplot in each figure is a combined plot where each central median from Annoyance and Impulsiveness are used. The boxplot-style is chosen since it shows well the spread in the data.

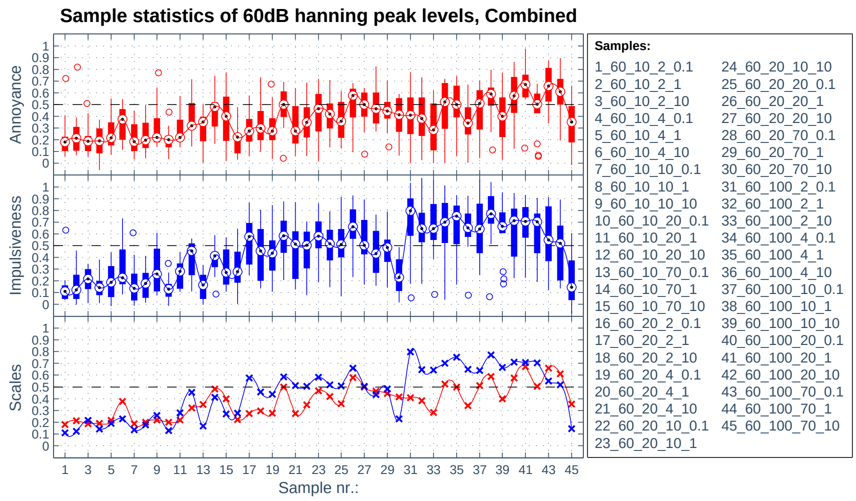


Figure 6.4: *Boxplot of all 60 dB hanning peak levels through 2 scales of the average between repetition 1 and 2. 14 subject responses for each. The last subplot is a combined scale plot where each central median from Annoyance and Impulsiveness are used. The legend to the right shows the names of the samples, where _ separates each attribute. From left to right: Sample nr.*_$\mathrm{H_P}$_$\mathrm{H_{S/N}}$_$\mathrm{H_R}$_$\mathrm{H_L}$.

[2]The figure can be found on the enclosed DVD `/Plots/Misc/RepCons_Mean.pdf`

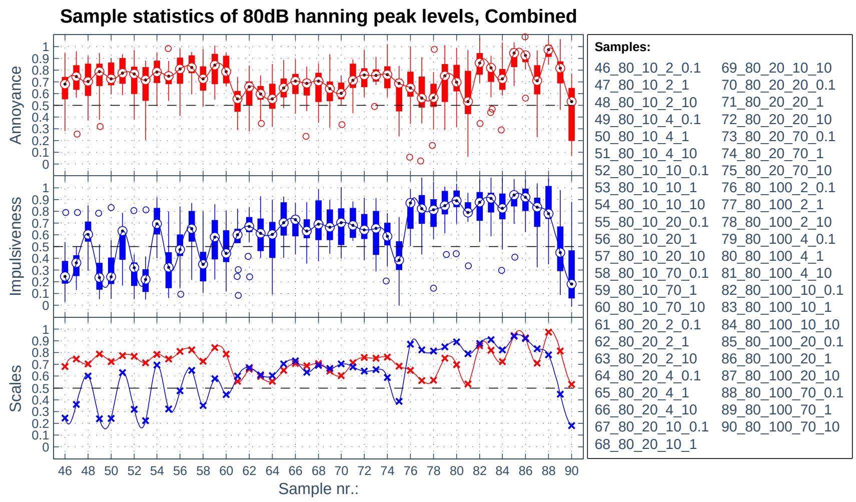


Figure 6.5: *Boxplot of all 80 dB hanning peak levels through 2 scales of the average between repetition 1 and 2. 14 subject responses for each. The last subplot is a combined scale plot where each central median from Annoyance and Impulsiveness are used. The legend to the right shows the names of the samples, where _ separates each attribute. From left to right: Sample nr.*_H_P_$H_{S/N}$_H_R_H_L.

As seen in figure 6.4 both scales seems to be correlated. The same trend is not completely seen in figure 6.5. At least not from sample 60→75. Statistics shows that the scales are correlated with a p-value $\ll 0.05$[3].

In order to test which of the parameters that causes a significance difference, a N-way analysis of variance (ANOVA) has been conducted. Subjects are considered as a random factor since they are chosen randomly from the population. 2nd-order terms are included in the analysis. Values with p-values<0.05 are considered as significant. In order to test that the assumptions associated with the ANOVA-analysis are not violated a normal probability plot of the residuals is shown in figure 6.7 (bib, 2012). As seen the residuals follows smoothly the red dashed line except in the lower and upper ends. It is thus concluded that the assumptions of the ANOVA model is not violated.

From the ANOVA-analysis seen in table 6.6 4 interesting conclusions are drawn and listed below. Each beginning with the sentence: "There is a significance difference in how the "[4]

1. SNRatio and HannLevel effects the ratings.
2. SNRatio and RepFreq effects the ratings.
3. SNRatio and ImpLen effects the ratings.
4. RepFreq and ImpLen effects the ratings.

[3]The figure can be found on the enclosed DVD `/Plots/Misc/ScaleCorr_Mean.pdf`
[4]Same notation used in the ANOVA-table is used in the list.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subjects	15.223	13	1.171	1.38	0.1967
Scales	0.109	1	0.1086	0.13	0.7233
HannLevel	38.399	1	38.3986	68.1	0
SNRatio	20.028	2	10.014	25.91	0
RepFreq	3.113	4	0.7782	10.66	0
ImpLen	0.788	2	0.3941	5.85	0.0065
Subjects*Scales	3.087	13	0.2374	11.08	0
Subjects*HannLevel	2.316	13	0.1782	8.32	0
Subjects*SNRatio	10.048	26	0.3865	18.04	0
Subjects*RepFreq	3.795	52	0.073	3.41	0
Subjects*ImpLen	1.495	26	0.0575	2.68	0
Scales*HannLevel	4.013	1	4.0133	10.46	0.0047
Scales*SNRatio	9.072	2	4.5359	211.77	0
Scales*RepFreq	3.432	4	0.8579	40.05	0
Scales*ImpLen	0.271	2	0.1356	0.38	0.6924
HannLevel*SNRatio	2.376	2	1.188	55.46	0
HannLevel*RepFreq	0.118	4	0.0296	1.38	0.2381
HannLevel*ImpLen	0.264	2	0.1319	1.94	0.1532
SNRatio*RepFreq	2.321	8	0.2901	13.55	0
SNRatio*ImpLen	6.678	4	1.6696	77.95	0
RepFreq*ImpLen	3.271	8	0.4088	19.09	0
Error	49.885	2329	0.0214		
Total	180.102	2519			

Constrained (Type III) sums of squares.

Figure 6.6: *N-way analysis of variance (ANOVA). 2nd-order terms are included. The mean of the repeated samples from the Annoyance and Impulsiveness listening test has been taken respectively. Subjects are considered as a random factor. Scales are the annoyance and impulsiveness scale used in the experiment.*

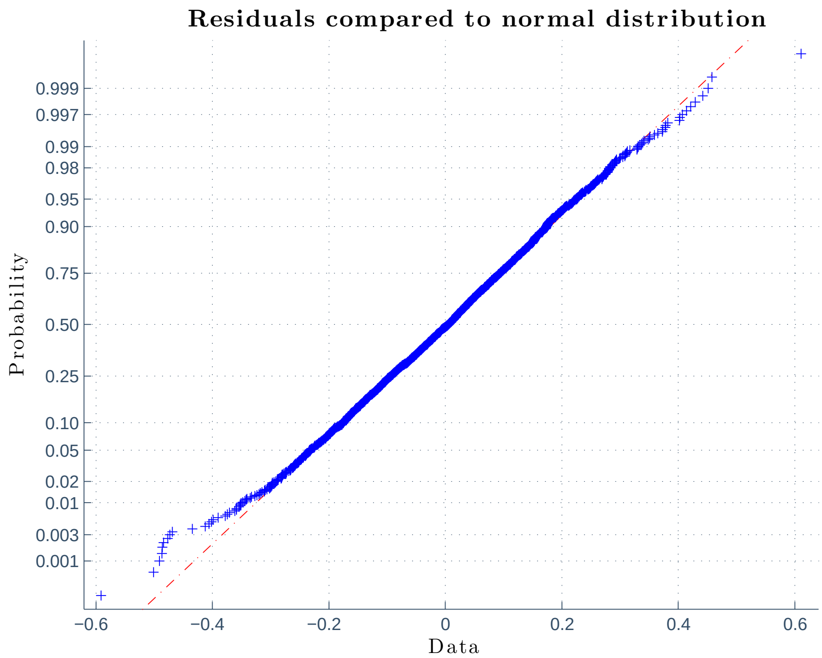


Figure 6.7: *Residuals from Anova analyse compared to normal distribution.*

As seen the $H_{S/N}$ (SNRatio) has a big influence in the variations of the results. When looking at the two box-plots (figure 6.4 and figure 6.5) this effect is visual shown. H_R (RepFreq) and H_L (ImpLen) also has influence in the variations. In appendix C different plots are used to visualise those variations. The conclusion from the analysis are:

In terms of Annoyance

Noise is a little bit more annoying than impulses when $H_P = 80$. When $H_P = 60$ the Annoyance increases with increasing $H_{S/N}$. Samples are generally perceived more annoying when increasing H_R except when $H_{S/N} = 100$ for $H_L = 10$.

In terms of Impulsiveness

The perception of Impulsiveness increases with increasing H_P and $H_{S/N}$. The variation in H_R does not vary much within the same H_P, $H_{S/N}$ and H_L except when $H_R = 70$, $H_{S/N} > 10$ and $H_L > 0.1$ where the perceived impulsiveness decreases.

6.2 Objective results

In section 4.2.2 on page 20 equation 4.1 was given to describe the amount of impulsiveness of the output from the Relative Approach. The impulsiveness through all samples is shown in figure 6.8. The calculated impulsiveness are compressed by setting $p > q$ in equation 4.1. Other values of p and q has been tested, but did not show promising results.

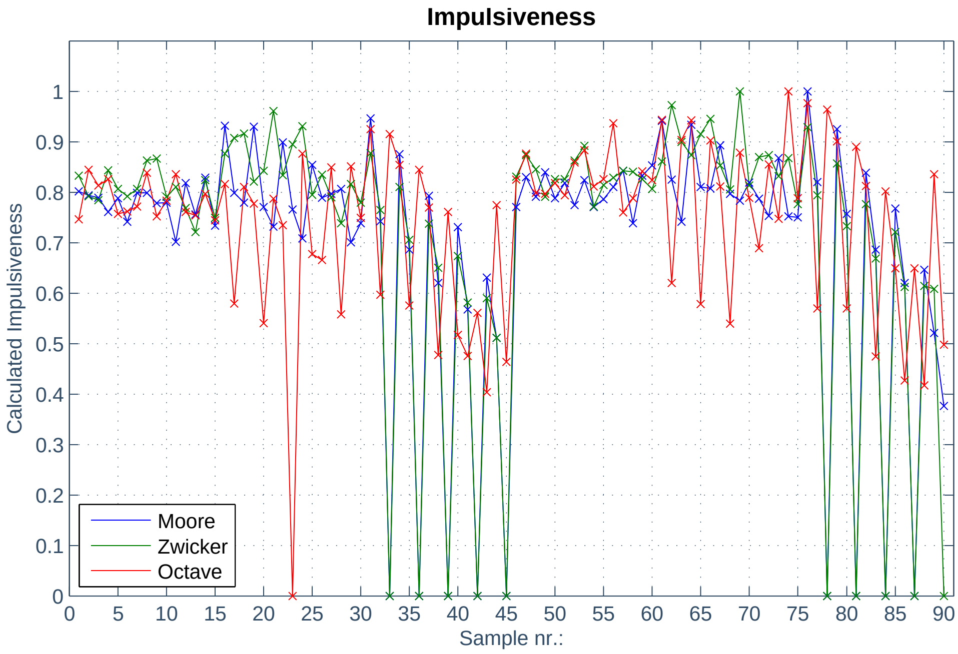


Figure 6.8: *Calculated compressed and normalised impulsiveness, when $p > q$ in equation 4.1 of the output from the Relative Approach algorithm by using time-varying Moore & Zwicker loudness and fractional-octave-filters as input trough all samples.*

As seen in the figure some consistency exist between the three input models, but the general impulsiveness are general almost constant around 0.8 from sample 1→30 and from sample 46→75 with the highest fluctuation caused by Octave in both regions. A closer look of these variations in sample 17, 20 and 23 is shown in figure 6.9 [5] where the RA-output of each model is shown.

In all the samples $H_L = 1$ and the low values of the Calculated impulsiveness of Octave is caused by that most of the frequency content is below 500 Hz and the Impulsiveness algorithm (section 4.2.2) only account for frequencies above 500 Hz.

[5] The figure can be found on the enclosed DVD `/Plots/Misc/Multi_RAout_17_20_23.pdf`

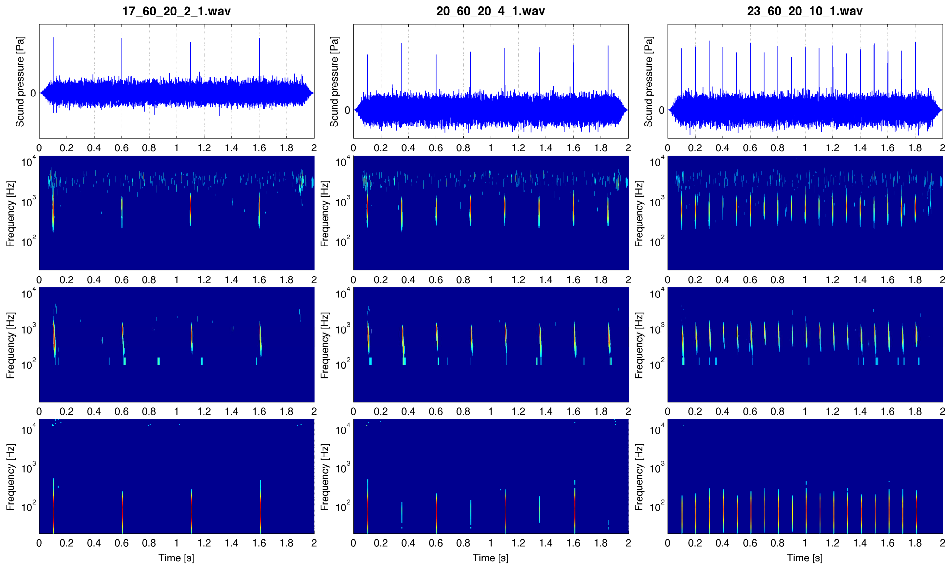


Figure 6.9: *RA-output through sample 17, 20 and 23 of Moore (2nd row), Zwicker (3rd row) and Octave (4th row). The first row shows the time-domain plot of the sample of interest.*

In the region from sample 32→45 large variations is seen in figure 6.8, where Moore[6] and Zwicker reaches a value of zero several places. In figure 6.10 [7] the RA-output of sample 33, 36 and 39 is shown. Here both Moore and Zwicker has frequency contents below 500 Hz which is the reason for the zero value

The largest calculated impulsiveness through all samples is seen in sample 69 for Zwicker, 74 for Octave and 76 for Moore. The RA-output is seen in figure 6.11[8]

6.3 Correlation

In order see if any correlation exist between the results from the subjective experiment and the objective analysis the two set of results are visualised in figure 6.12 on page 46.

No really correlation is seen in figure 6.12. In the following three correlation-plots the data from Moore vs. Subjective, Zwicker vs. Subjective and Octave vs. Subjective are shown in figure 6.13, 6.14, and 6.15 respectively. As seen non of the combinations shows a significant correlation. Octave vs. Subjective exhibits the highest negative correlation value.

[6]Moore is hidden behind the Zwicker line

[7]The figure can be found on the enclosed DVD `/Plots/Misc/Multi_RAout_33_36_39.pdf`

[8]The figure can be found on the enclosed DVD `/Plots/Misc/Multi_RAout_69_74_76.pdf`

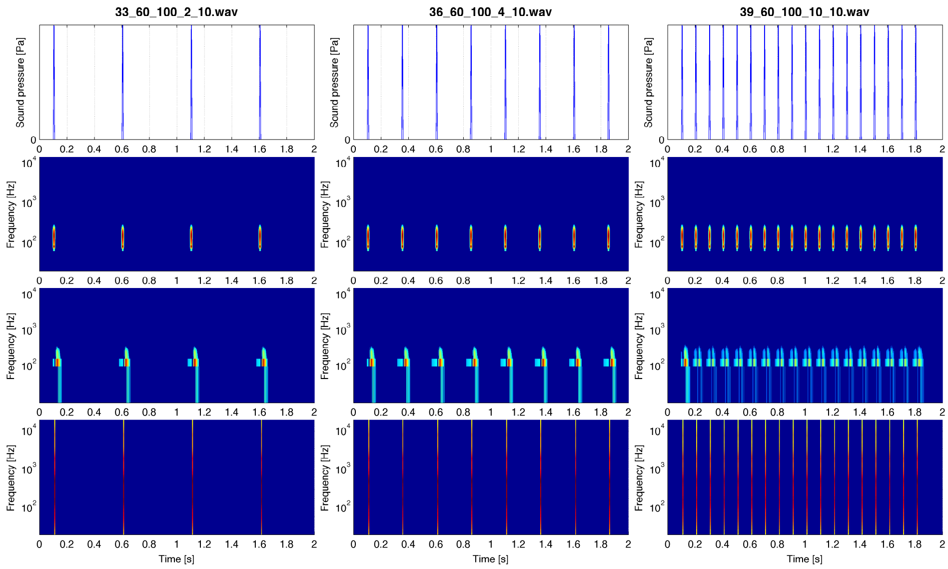


Figure 6.10: *RA-output through sample 33, 36 and 39 of Moore (2nd row), Zwicker (3rd row) and Octave (4th row). The first row shows the time-domain plot of the sample of interest.*

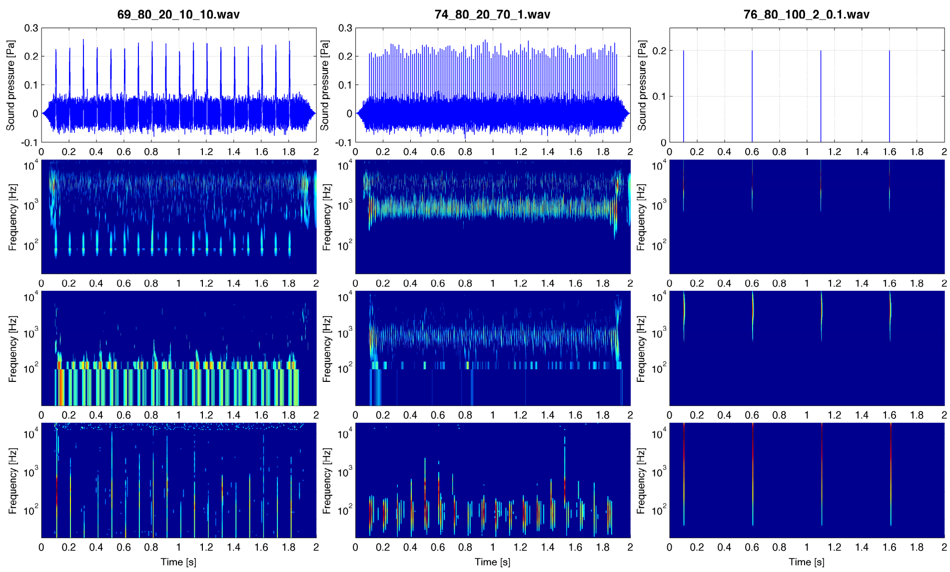


Figure 6.11: *RA-output through sample 69, 74 and 76 of Moore (2nd row), Zwicker (3rd row) and Octave (4th row). The first row shows the time-domain plot of the sample of interest.*

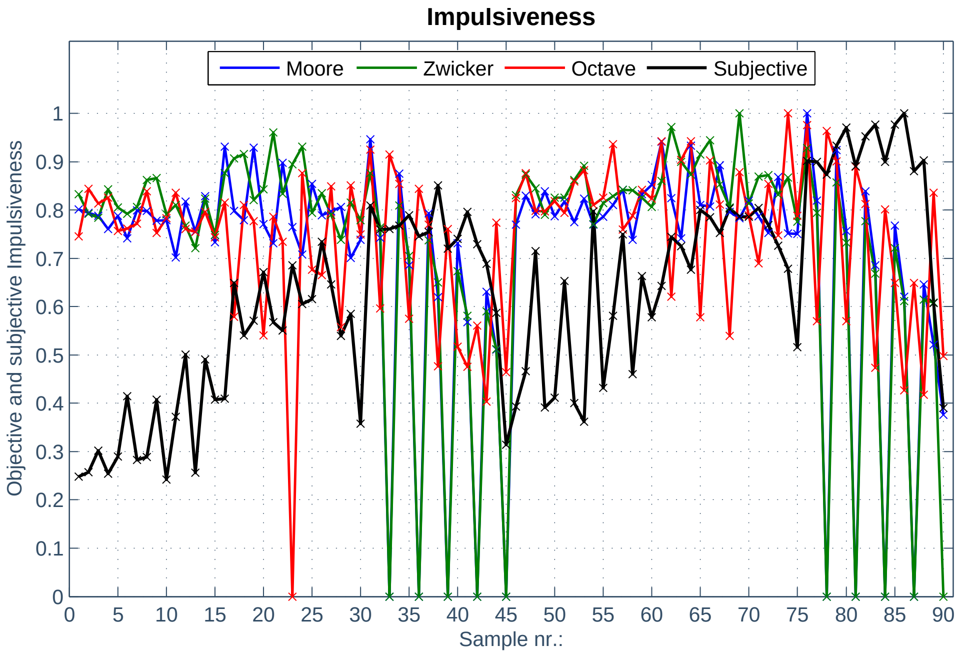


Figure 6.12: *Objective and Subjective Impulsiveness through all samples. A mean value across the subjective impulsiveness has been performed and subsequently normalised.*

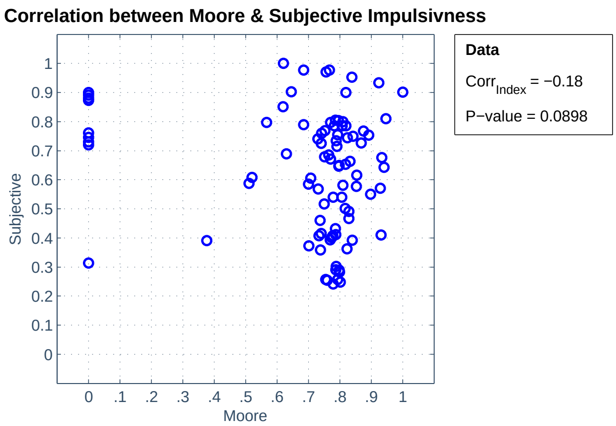


Figure 6.13: *Correlation between Moore and Subjective Impulsiveness*

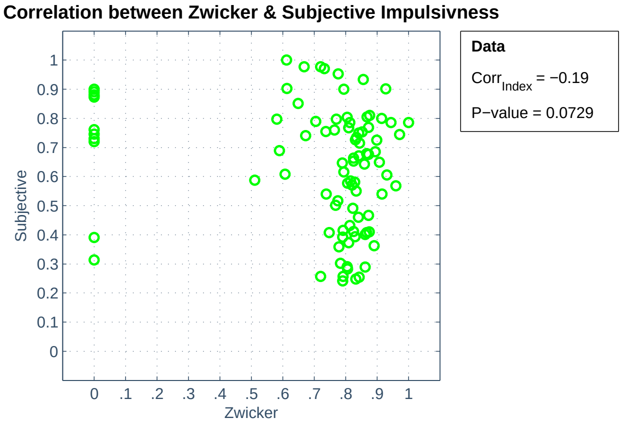


Figure 6.14: *Correlation between Zwicker and Subjective Impulsiveness*

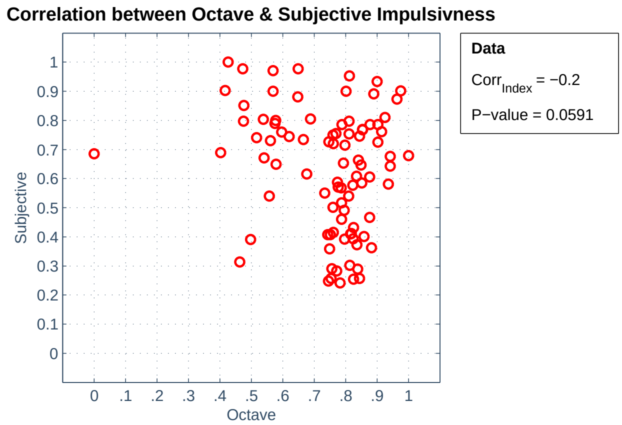


Figure 6.15: *Correlation between Octave and Subjective Impulsiveness*

DISCUSSION AND CONCLUSION

CHAPTER 7

Discussion

In this project the psychoacoustic experiment was carried out on 14 subjects. A larger group containing a more spread age range and an equal number of each gender would perhaps have given more reliable results. Some issues regarding the listening test could have been addressed in a better way.

The training session that was presented before each listening part could have consisted of the extremities among the samples to give the subjects an impression of the range of the test. That would have minored the problem of rating samples according to prior knowledge. One of the subjects addressed this problem (B.2).

When giving the subjects instruction it was not told that the ratings of each sample should be given by the impression of the overall sample, which should have been done, neither of the individual impulses noticed in the sample. The subject was then leaved to his/her own impression. It would be interesting to see, if the right information would have changed the results.

One of the subjects told that if the duration of the samples has been longer, it would have been more annoying. Another subject told that the sounds may be more annoying if he should listen to them every day. This suggest that the samples should had been longer in duration than 2 seconds in order to give the subject a better chance to rate the samples.

Regarding the noise presented in the samples, constant noise levels could have been used for both Hanning Peak Levels, instead of having the S/N connected to the Hanning Peak Levels since high levels of noise are perceived more annoying than impulses.

Concerning the Objective Algorithm, several parts could be improved. The noise level after running the RA-algorithm reaches too high levels, which causes the calculation of impulsiveness to be too high for samples concealed in noise. A solution could be to adjust the sub-threshold value in the relative approach to a higher value, as described in section4.2.4, but the trade off could be to remove some of the salient structure. Several settings was tried but because of time limitations no optimal value was found. Another solution could be to calculate the mean and maximum value of the

RA-output and compare them. If the difference is to small it could be an indication of that no structure could be detected and the calculated impulsiveness should be adjusted to match subjective ratings better.

3 other factors that could be tuned are the p,q and k_j values described in 4.2.2. The p,q values adjusts the compression of the calculation of impulsiveness and by changing these values quite different results can be achieved. k_j is the frequency weighting factor, appropriate values may change the results in a proper way.

The toolboxes used for calculating Time-varying-loudness contained some minor errors. The output of Time-varying Moore loudness contained 153 different frequencies which is not a fraction of 24 ERB, and in the function the resolution was set to 0.25 ERB. By changing this numbers to ERB=1, the number of frequencies was reduced to 39 . A newer version of the toolbox was tried, but another issue was found. The time-vector for Time-varying Moore loudness contained only time steps to about 1.6 seconds compared to the sample duration of 2 seconds. The old version was as a result chosen instead.

New input models could be tested like e.g. Discrete-Wavelet-Transform, Loudness model for impulsive events proposed by Boullet (Genesis, 2009), Matching Pursuit (Mallat, 1993) , Reassigned Spectrogram (Hainsworth and Macleod, 2001), High-Resolution Spectral Analysis (Genuit and Bray, 2008), Hearing model proposed by Sottek (Sottek, 1993) etc. to verify if they better captures the impulsive event.

7.1 Correlation between objective and subjective results

The calculated impulsiveness did not show promising results. The overall rating of impulsiveness was simply to high because that the background noise was not probably been removed from the RA-output as addressed previously. This has among other things caused the correlation with the subjective results to be poorly correlated. In (Song and Saito, 2011) the author of the article has kindly provided results generated on the same sample-set used in this project of their proposed impulsiveness algorithm. Their algorithm shows promising results as seen in figure 7.1. Their results are labeled as "External". As seen their results are closely related to the subjective impulsiveness. In figure 7.2 a correlation plot between External and Subjective Impulsiveness is shown. As seen there is a significant correlation between the External and Subjective Impulsiveness with a p-value $<$ 0.05.

Their proposed impulsiveness algorithm takes into account the dependency of impulse-repetition frequency and emphasize bark-bands between 10 and 13. The detection of impulses is performed by taking a FFT of the time-varying loudness and subsequently detecting peaks in the frequency domain. Those processing steps could be transferred to the proposed algorithm in this project in future investigations in-order to test if the adjustments improved the performance.

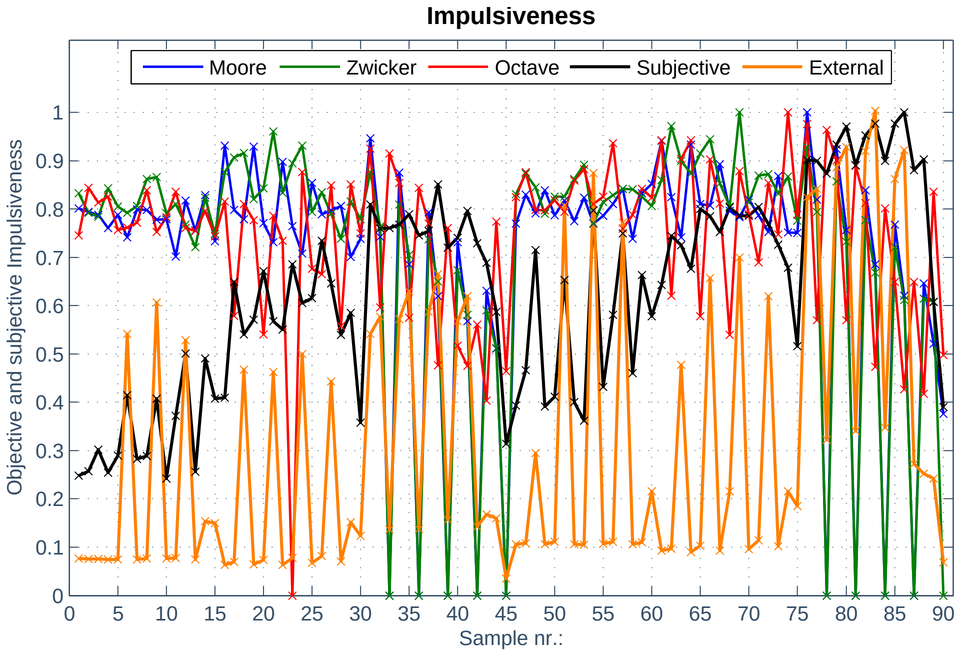


Figure 7.1: *Objective and Subjective Impulsiveness through all samples. A mean value across the subjective impulsiveness has been performed and subsequently normalised. The External are data provided by the author of (Song and Saito, 2011)*

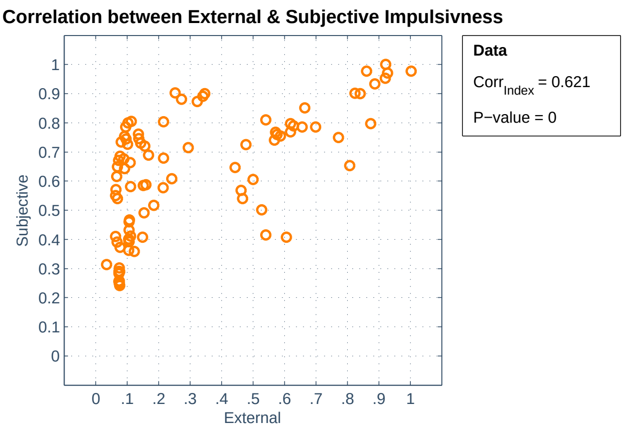


Figure 7.2: *Correlation between External and Subjective Impulsiveness*

CHAPTER 8

Conclusions

The goal of the project was to explore if any correlation existed between the objective and subjective evaluation of the samples.

90 different samples was generated for that purpose with 4 controllable parameters in order to test the different aspects of the nature of impulsive sounds.

From the subjective experiment some conclusions has been drawn about the ratings on the annoyance and impulsiveness scale. In terms of Annoyance: Noise is a little bit more annoying than impulses when $H_P = 80$. When $H_P = 60$ the Annoyance increases with increasing $H_{S/N}$. Samples are generally perceived more annoying when increasing H_R except when $H_{S/N} = 100$ for $H_L = 10$.

In terms of Impulsiveness: The perception of Impulsiveness increases with increasing H_P and $H_{S/N}$. The variation in H_R does not vary much within the same H_P, $H_{S/N}$ and H_L except when $H_R = 70$, $H_{S/N} > 10$ and $H_L > 0.1$ where the perceived impulsiveness decreases.

It has been found from the subjective experiment that the repeated measurements is correlated with a p-value $\ll 0.05$ and thus the subjects answers consistently. The same is valid for the correlation between impulsiveness and annoyance from the subjective experiment where the p-value $\ll 0.05$.

Three input models; Time-varying Moore Loudness, Time-varying Zwicker Loudness and Fractional Octave Filters has been used to pre-analyse the samples in the objective analyse. Their outputs was feed into the Relative Approach Algorithm combined with an Algorithm to calculate the amount of impulsiveness. The results were not promising, since samples with a high amount of background noise were rated to high compared to samples with practically no background noise. The different parameters has to be revised in order to fine-tune the algorithm to perform acceptable. The Relative Approach included with the input models required much computations power, mainly due to the long execution of the Loudness toolbox. In order to analyse 90 samples it required about 8.5 hours of calculation.

Due to the problem mentioned above the correlation between the subjective and the objective parameters was very low. Results provided by the author of (Song and Saito, 2011) showed significant correlation with the subjective results. Future investigations could include elements from this algorithm in-order to test if the adjustments improved the performance.

Appendix

APPENDIX A

Samples

In section 5.3 the experiment duration for each subject has been chosen to be 1 hour. When the time for conducting the audiometry + instruction time + training session + breaks is subtracted, it yields about 30 minutes of effective testing time. This causes limitations in how many samples is possible to present within that time. An optimal scenario would of course be to test all interesting settings for each parameter but a trade-off must be considered.

In chapter 3 the different parameters has been described, and the parameters are listed as follows.

- Hanning impulse length (H_L) [ms]
- Hanning impulse peak level (H_P) [dBSPL]
- Hanning repetition frequency (H_R) [Hz]
- S/N level ($H_{S/N}$) [dB]

Four additional settings were necessary to be included in order to calculate the expected effective experiment time. They are:

- Sample length [s]
- Repetition of each sample
- Number of scales used in the subjective experiment
- Estimated rating time.

In table A.1 each parameter, the initial settings and the chosen settings are listed.

Discussion and selection of samples

The initial settings were chosen at first, but pilot tests resulted in more optimal settings listed as chosen settings used in the experiment. To reduce the experiment time, the sample length was reduced from 5 to 2 seconds. This length is very short, but since it was the spontaneous perception of the sample that was needed, this length was used. It could be argued that this

Parameter	Initial settings	Chosen settings	Unit
Estimated rating time	2	2	[s]
Number of scales	2	2	[.]
Sample length	5	2	[s]
Repetition of each sample	2	2	[.]
Hanning impulse peak level	40, 60, 80	60, 80	[dB SPL]
Hanning repetition frequency	4, 10, 20, 50, 70	2, 4, 10, 20, 70	[Hz]
S/N level	0, 10, 20	10, 20, 100	[dB]
Hanning impulse length	0.1, 1, 10	0.1, 1, 10	[ms]
Nr. of different Samples	135	90	[.]
Total Nr. of Samples	540	360	[.]
Raw time (Without rating)	45	12	[min]
Total time (With rating)	90	24	[min]

Table A.1: *Parameter and settings for the samples. The initial settings were chosen at first, but pilot tests resulted in more optimal settings listed as chosen settings used in the experiment.*

sample length is too short to give an impression of it. This topic is further discussed in the discussion chapter 7.

The hanning impulse peak level of 40 dB was removed since it was hardly audibly. The hanning repetition frequency of 50 [Hz] was removed since the difference between 50 and 70 [Hz] was negligible. 2 [Hz] was then added to introduce an event with a lower modulation frequency than 4 Hz, which causes the maximum Fluctuation strength (Chapter 3). The extrema-value of 70 Hz was chosen to represent where the sensation of Roughness is maximum (Chapter 3).

A S/N level of 0 dB was removed since the hanning impulses were almost concealed in noise and only distinguished at a repetition frequency of 70 [Hz]. 80 dB noise also sounded too loud. A S/N level of 100 dB was added to have a situation of hanning impulses alone (In practice). As described in chapter 3 a particular BSR event has sometimes a duration below 10 ms (Shin and Cheong, 2010), but an extensive literature search did not reveal any details of a typical BSR-duration. By means of that the 3 Hanning impulse length has been chosen arbitrarily with increasing of tenfold in duration per step from 0.1→10 [ms].

The total number of samples turned out to be 360 and they were divided into two parts with 180 samples in each [1]. The total experiment time was estimated from table A.1 to be 24 minutes which includes the estimated rating time of 2 seconds. A further description of the experiment is given in section 5.3.

[1] The samples can be found on the enclosed DVD `/Samples/`

Additional discussion

An important notice when changing the Hanning impulse length and the Hanning repetition frequency is that the RMS level of the impulse-sequence alone changes significantly. In the following example the two extreme situations is used. The Hanning peak level is in both cases 80 dB SPL (Fs = 48 kHz).

Case 1

Hanning impulse length = 0.1 [ms]
Hanning repetition frequency = 2 [Hz]

Case 2

Hanning impulse length = 10 [ms]
Hanning repetition frequency = 70 [Hz]

The two cases is shown in figure A.1 and A.2 respectively.

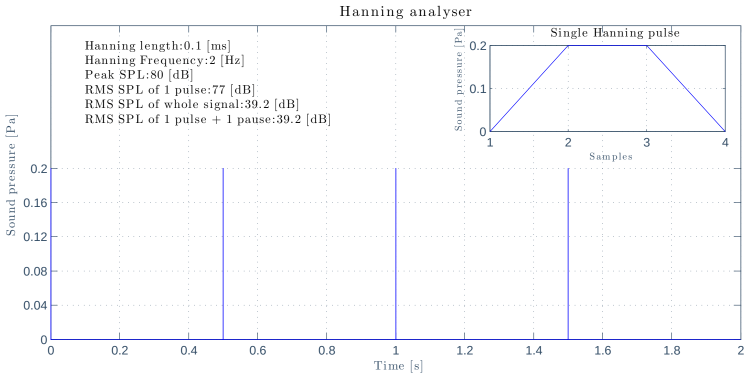


Figure A.1: *Hanning analyser of Peak level = 80 dB SPL, Impulse length = 0.1 [ms] and Repetition frequency = 2 [Hz].*

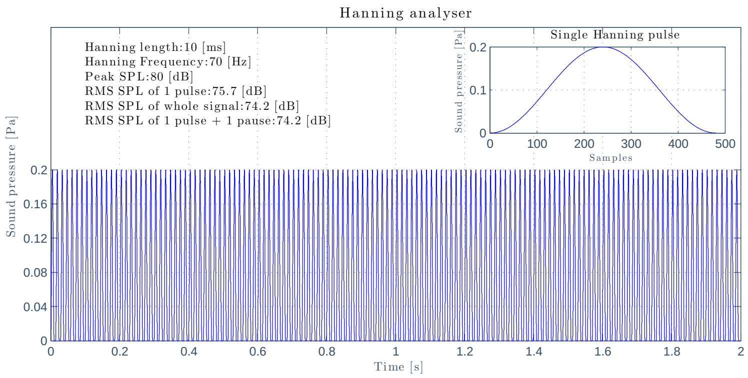


Figure A.2: *Hanning analyser of peak level = 80 dB SPL, Impulse length = 10 [ms] and Repetition frequency = 70 [Hz].*

As seen in the legend of both figures there exist a huge RMS-difference between the two cases of 35 dB. When the S/N level is low this large difference will induce that the impulses in case 1 will not be audible compared to case 2.

APPENDIX

B

Experiment formailites

This appendix contains the following experiment formalities:

APPENDIX B.1 Instruction list

APPENDIX B.2 Answers from subjects

B.1 Instruction list

This appendix contains the Danish and the translated English version of the instruction list, which the participants has to read before conducting the experiment as described in section 5.3. The lists are shown on the following 4 pages.

Kære deltager

Velkommen til akustik afdelingen på Aalborg University.

Hele forsøget kommer til at tage ca. 45-60 minutter og er opdelt i 3 deltest.

1. Høretest
2. Træningsforløb + Lyttetest del 1
3. Træningsforløb + Lyttetest del 2

Først vil vi foretage en høretest for at finde ud af om din hørelse er tilstrækkelig til at vi kan fortsætte med selve lyttetesten. I høretesten vil du blive præsenteret for forskellige toner. Din opgave er trykke på knappen hvis du kan høre tonen. Testen varer ca. 10-15 minutter.

Formaliteter (Læs før høretesten)

Lydeksponeringen i forsøget er ikke farligt for dine hørelse

Dine svar vil være fuldstændig anonym

Hvis du føler dig utilpas eller af en anden grund ikke har lyst til at foretage høretesten og/eller lyttetesten, kan du på ethvert tidspunkt ophøre forsøget.

Har du nogle spørgsmål og kan du forliges med ovenstående?

(Læs efter høretesten)

Lyttetesten varer ca. 30-40 minutter inkluderet med pauser. Den er opdelt i 2 dele.

Lyttetest del 1

I den første del skal du bedømme lydklippet med hensyn til ?annoyance? (generende, irriterende). I figuren nedenunder er skalaen vist. Hvert lydklip varer 2 sekunder og det er din spontane opfattelse du skal bedømme det efter. Efter halvdelen af lyd-klippene, vil der være 2 minutters pause og bagefter vil du blive præsenteret for de resterende lydklip.

Figure B.1: *Annoyance scale.*

Lyttetest del 2

Efter 5 minutters pause fortsætter 2 del af lyttetesten. I den denne del skal du bedømme lydklippet med hensyn til ?Impulsiveness? (impulsivitet). I figuren nedenunder er skalaen vist. Resten af proceduren er magen til den beskrevet i Lyttetest del 1.

Figure B.2: *Impulsiveness scale.*

Når all lydklippene er blevet bedømt, er forsøget færdig

Før hver lyttetest vil der være en kort træningssession så du kan blive velkendt med opgaven ?

Har du nogle spørgsmål?

Dear Participant

Welcome to the department of Acoustics, at Aalborg University.

The whole experiment has an duration of about 45-60 minutes. It Is divided into 3 parts.

1. Hearing test
2. Training session + Listening test part 1
3. Training session + Listening test part 2.

First an hearing test will be carried out to test if your hearing level is appropriate to enter the listening test. In the hearing test you will be presented to different tones one ear at the time. Your task is to push the button if you can hear the tone. The test has an duration of about 10-15 minutes.

Additional formalities (Read before hearing test)

The presented levels during the experiment are not dangerous for your hearing.

Your responses will be completely anonymous.

If you feel uncomfortable or by any other reason don't want to carry out the hearing test and/or the listening test, you can by any time terminate the test.

Do you have any question and do you agree with the above mentioned?

(Read after hearing test)

The listening test has an duration of about 40 minutes, with breaks included. It is divided into two parts.

Listening test part 1

In the first part you are asked to rate the sounds in terms of annoyance. In the figure below, the scale is shown. Each sound is very short (2 sec) and it is your spontaneous perception of the sound that should be transferred to the scale. After half of the presented sounds, a break of 2 minutes is included and afterwords the last half of the sounds will be presented.

Figure B.3: *Annoyance scale.*

Listening test part 2

After a 5 minute break, the second part of the listening test enters. In the second part you are asked to rate the sounds in terms of impulsiveness. In the figure below, the scale is shown. The rest of the procedure is similar to the one described in Listening test part 1.

Figure B.4: *Impulsiveness scale.*

When all the sounds has been rated, the experiment is over.

Prior each listening part a short training session will be given to get you familiar with the task.

Do you have any question?

B.2 Answers from subjects

This appendix contains a list of answers collected in breaks of the experiment. The answers emerged from one question, from section 5.3, and are evoked by own initiative:

"What did you felt most annoying, impulses or noise?"

The answers are listed in table B.1

Subject ID	Comments
1	Impulsiveness: Stammering vs. Fluid
2	Talks about that the impulsive scale could be a little dependent on annoying. If the duration of the samples has been longer, it would maybe have been more annoying.
3	The sounds may be more annoying if he should listen to them every day
4	White noise more annoying than impulses. Higher volume white noise = more annoying. White noise more annoying than impulses. Higher volume impulses = more impulsive.
6	Thinks that impulses are more annoying than noise, especially at the right frequency
7	White noise not so annoying. Annoying is situation dependent. Impulses more annoying when high S/N level
8	Was difficult to hear the impulses in some of the samples. She figured out that it is a repeated set. Would like a marker at the middle of each scale
9	Impulses more annoying when high S/N. Noise most annoying
10	Has difficulty understanding the impulsiveness test. Annoyance is more clear
13	Used much more time than the other subjects to rate the samples (about twice as long). He did not use the annoyance scale as much, he more used the knowledge of prior presented samples to rate new samples. The sharp sounds and high level of noise was the most annoying

Table B.1: *Answers from subjects.*

APPENDIX C

Additional results

In section 6.1 an N-way analysis of variance (ANOVA) shows that: "There is a significance difference in how the "

1. SNRatio and HannLevel effects the ratings.
2. SNRatio and RepFreq effects the ratings.
3. SNRatio and ImpLen effects the ratings.
4. RepFreq and ImpLen effects the ratings.

In this appendix a visualisation of those effects are addressed.

C.1 Ratings in terms of hanning length and repetition frequency

In figure C.1 the ratings of Annoyance in terms of hanning length and repetition frequency is shown. For $H_P = 60$ the Annoyance only increases by a minor part from $H_{S/N} = 10 \rightarrow H_{S/N} = 20$. At $H_{S/N} = 100$ the overall ratings increases for $H_L = 0.1$ and $H_L = 1$, but $H_L = 10$ decreases compared to those two. This is caused by that $H_L = 10$ sounds less impulsive. When $H_P = 80$ the overall Annoyance starts at a higher level for all $H_{S/N}$-levels compared to $H_P = 60$. Again when $H_{S/N} = 100$ the level of Annoyance fall with increasing H_L. The highest drop is seen with $H_R = 70$. In most cases the ratings between the different H_R for each H_L, $H_{S/N}$-combination does not vary much.

In figure C.2 the ratings of Impulsiveness in terms of hanning length and repetition frequency is shown. In the $H_P = 60$ cases the level of impulsiveness generally increases with increasing $H_{S/N}$. When $H_P = 80$ and $H_{S/N} = 10$ the general impulsiveness increases with increasing H_L and that is mainly due to that the impulses are hardly audible when $H_L = 0.1$. When $H_P = 80$, $H_{S/N} = 20$ and $H_{S/N} = 100$ the ratings does not vary much with increasing H_L. In all $H_{S/N} = 20$ and $H_{S/N} = 100$, $H_R = 70$ decreases when $H_L = 1$ and $H_L = 10$, compared to the others. In general the variations within the same H_P, $H_{S/N}$ and H_L does not vary much.

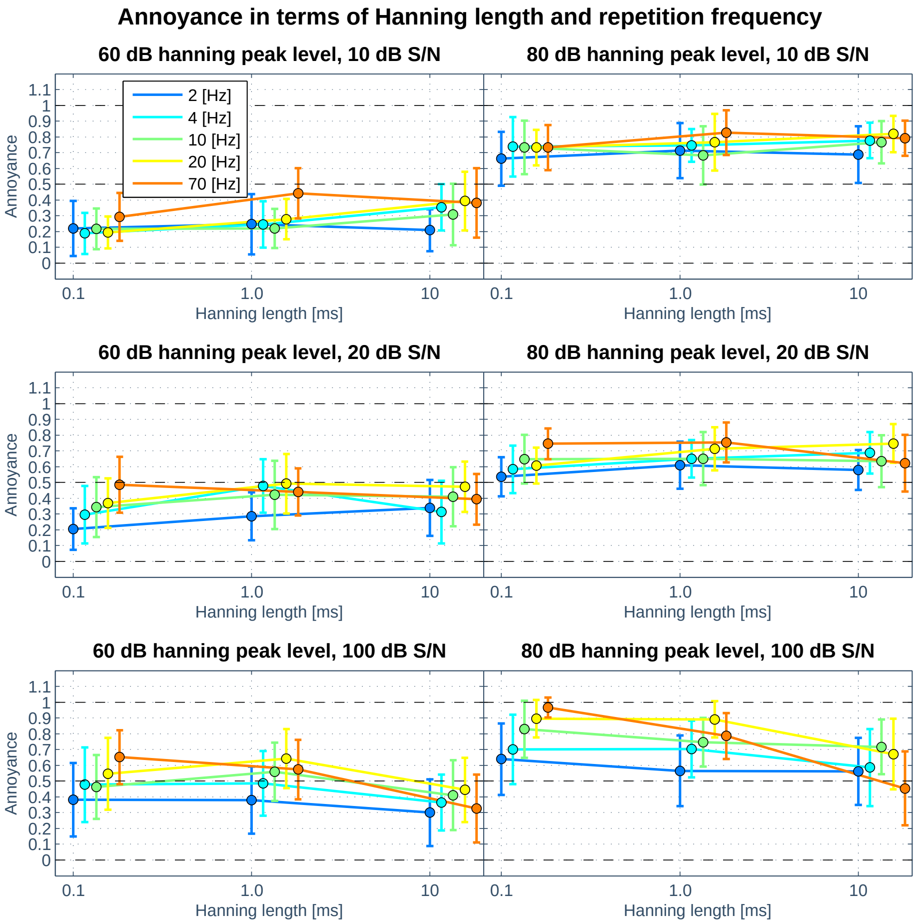


Figure C.1: *Annoyance in terms of hanning length and repetition frequency. Each point are the mean values through all subjects and the bars is the standard variation. For each hanning length the points has been shifted to better distinguished between them.*

C.2 Ratings in terms of repetition frequency and hanning length

In figure C.3 the ratings of Annoyance in terms of repetition frequency and hanning length is shown. The general trend is that the ratings increases with increasing H_R. The $H_P = 80$ are rated more annoying than $H_P = 60$. When $H_{S/N} = 100$, $H_L = 10$ is rated lower than the other H_L. In general the ratings within the same H_P, $H_{S/N}$ and H_R, does not vary much

In figure C.4 the ratings of Impulsiveness in terms of repetition frequency and hanning length is shown. In general the ratings of increasing H_R are almost constant except when $H_{S/N} = 10$, where $H_R = 70$ exhibits the overall highest ratings and when $H_{S/N} = 20$ and $H_{S/N} = 100$ where $H_R = 70$ falls in rating compared to the others.

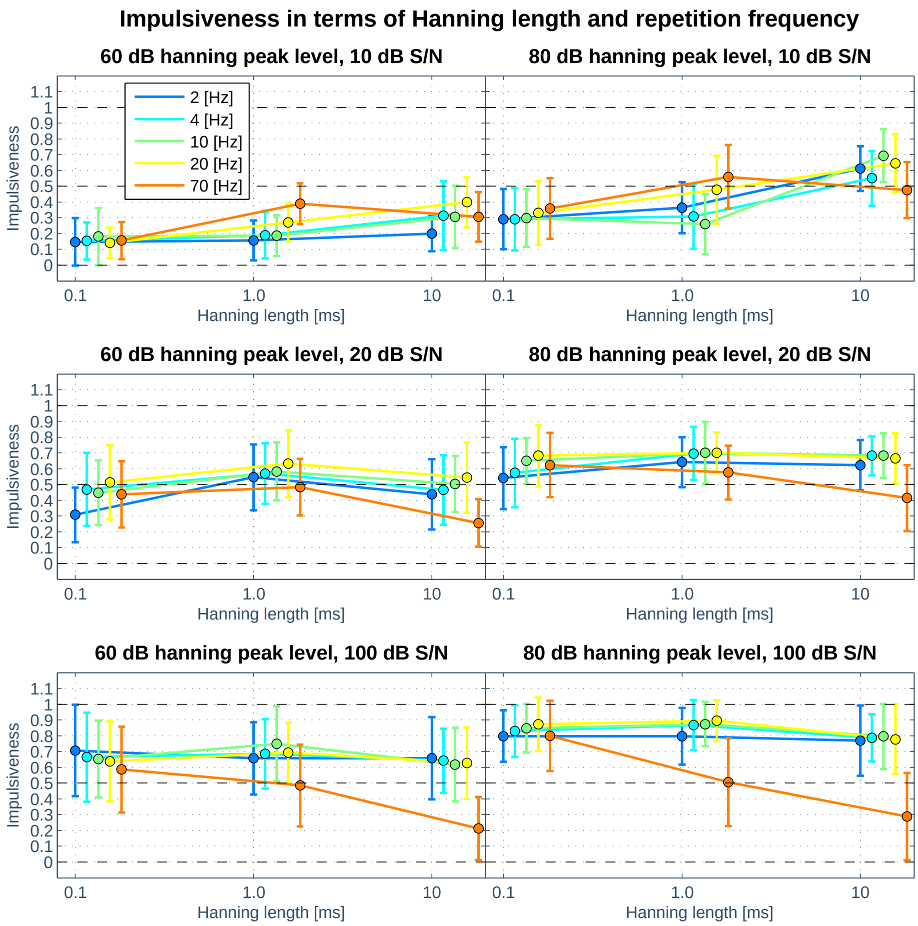


Figure C.2: *Impulsiveness in terms of hanning length and repetition frequency. Each point are the mean values through all subjects and the bars is the standard variation. For each hanning length the points has been shifted to better distinguished between them.*

C.3 Ratings of Annoyance and impulsiveness in terms of S/N-ratio and Hanning peak level

In figure C.5 the ratings of Annoyance and impulsiveness in terms of S/N-ratio and Hanning peak level is shown. In terms of Annoyance when $H_P = 60$ the ratings is increasing when $H_{S/N}$ is increasing. When $H_P = 80$ the highest rating is seen when $H_{S/N} = 10$, then it falls at $H_{S/N} = 20$ due to the lowered noise level and then increases a little when $H_{S/N} = 100$ because the impulses are clearly audible. In terms of Impulsiveness the trend are that the ratings increases with increasing $H_{S/N}$.

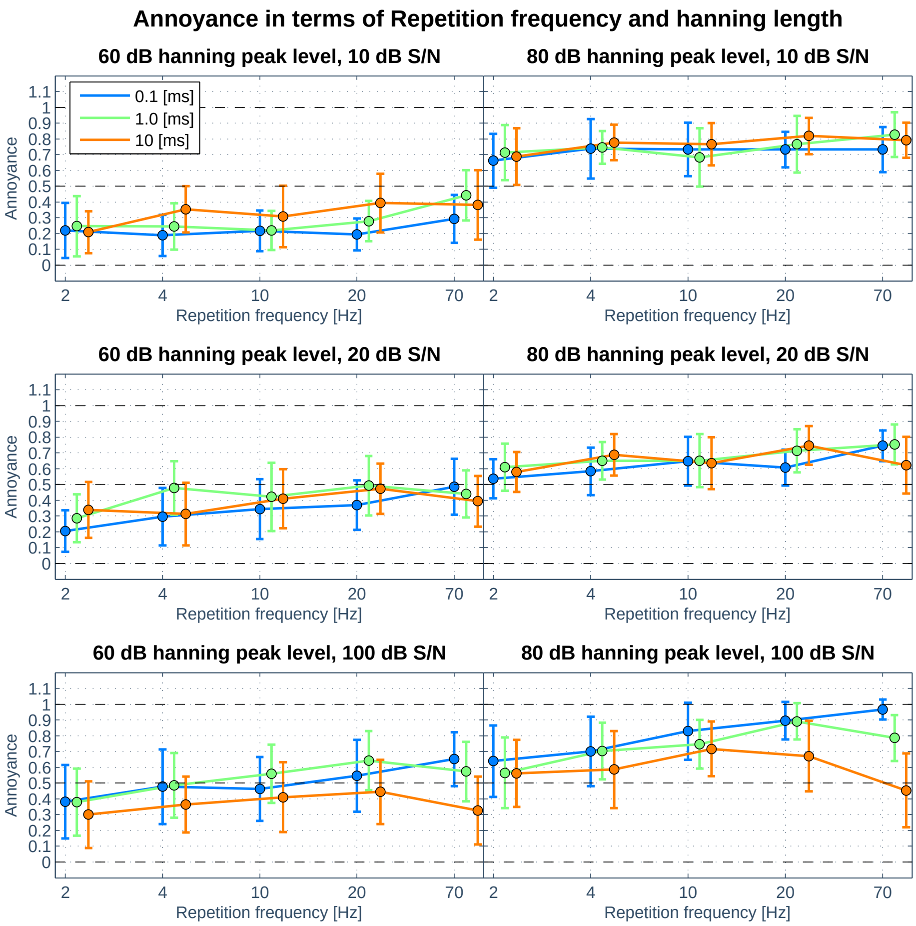


Figure C.3: *Annoyance in terms of repetition frequency and hanning length. Each point are the mean values through all subjects and the bars is the standard variation. For each repetition frequency the points has been shifted to better distinguished between them.*

Conclusion

In terms of Annoyance

Noise is a little bit more annoying than impulses when $H_P = 80$. When $H_P = 60$ the Annoyance increases with increasing $H_{S/N}$. Samples are generally perceived more annoying when increasing H_R except when $H_{S/N} = 100$ for $H_L = 10$.

In terms of Impulsiveness

The perception of Impulsiveness increases with increasing H_P and $H_{S/N}$. The variation in H_R does not vary much within the same H_P, $H_{S/N}$ and H_L except when $H_R = 70$, $H_{S/N} > 10$ and $H_L > 0.1$ where the perceived impulsiveness decreases.

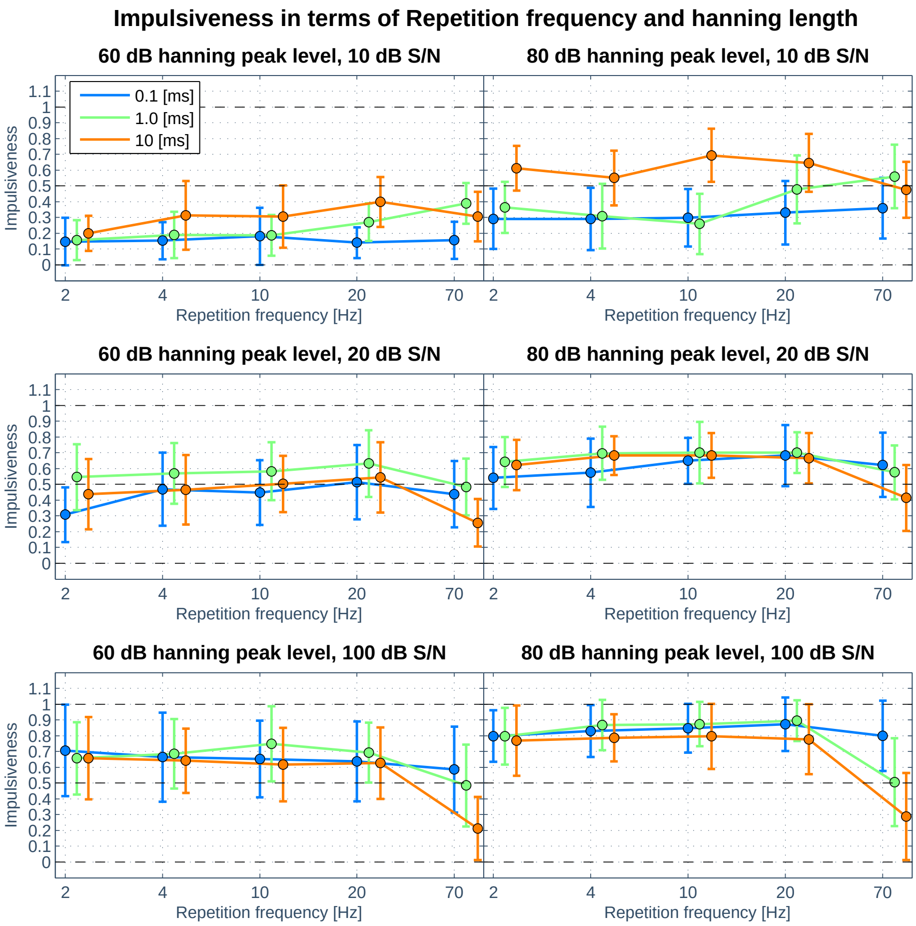


Figure C.4: *Impulsiveness in terms of repetition frequency and hanning length. Each point are the mean values through all subjects and the bars is the standard variation. For each repetition frequency the points has been shifted to better distinguished between them.*

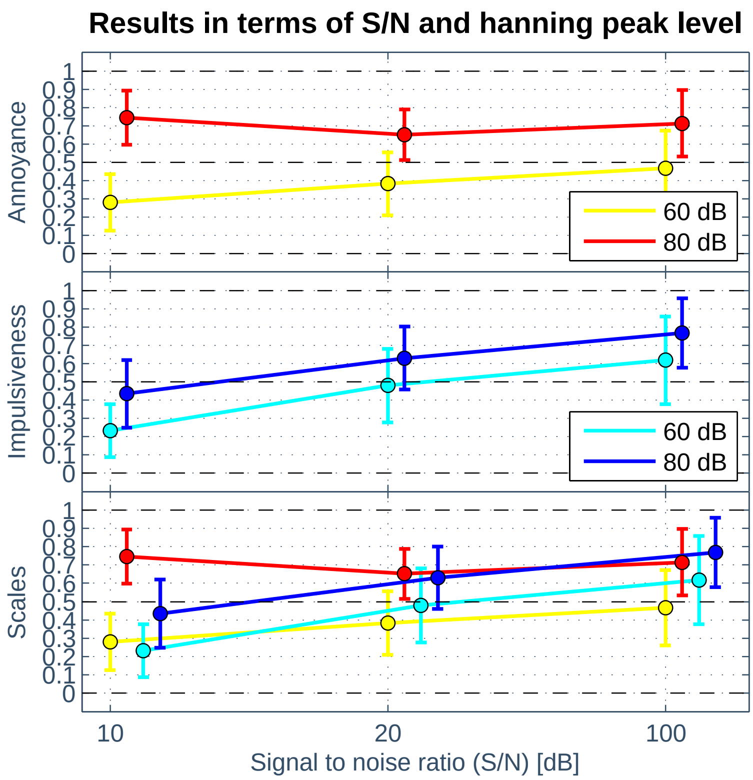


Figure C.5: *Annoyance and impulsiveness in terms of S/N-ratio and Hanning peak level. The last subplot is a combined plot. Each point are the mean values across subjects, hanning impulse length and repetition frequency. The bars shows the standard variations. Each point has been shifted to better distinguished between them. The legend shows the hanning peak level in dBSPL*

APPENDIX D

Measurement of BeyerDynamic DT990 Pro headphones + Ear transfer functions

The objective of this measurement journal was to measure the transfer function from BeyerDynamic DT990 Pro Headphones to the Ears of Valdemar Sejr (AAU custom made mannequin with artificial ears, with build in microphones) (Christensen et al., 2000), which was done in the Audiometry room at Aalborg University. The results obtained are used in chapter 3 were equalisation to the measured transfer function is applied.

Equipment

Control room A

- Computer with Matlab R2010a, PortAudio and Playrec preinstalled.
- Rotel Six channel power amplifier RB-976MKII, with fixed gain-setting of 0dB (AAU nr.: 33978)
- Edirol USB Audio Capture UA-25EX (Soundcard) (AAU nr.: 78377)
- 20 dB Attenuator box (AAU custom made)
- LD Systems LDI Active 02, Low noise active linedriver (DI). Converts a unbalanced to a balanced signal (External device, not from Aalborg University).
- Fluke 37 Multimeter (AAU nr.: 08285)
- PM 5193 programmable synthesizer/function generator (AAU nr.: 08125)
- Brüel & Kjær sound calibrator type 4231. 94 & 114 db SPL@1000Hz (AAU nr.: 78301).

Audiometry

- BeyerDynamic DT990 Pro[1], 2x250Ω (AAU nr.: 02036-77)
- Valdemar Sejr. AAU custom made mannequin with artificial ears, with build in microphones (AAU nr.: 2150-00) (Christensen et al., 2000).
  - Left microphone: G.R.A.S, type 40AD (AAU nr.: 33958). Mic.sens@1kHz/94dB = 35,4mV (-29,0dBV)

[1] BeyerDynamic DT990 Pro has PDR close to unityBlauert et al. (2005, p. 232)

– Right microphone: G.R.A.S, type 40AD (AAU nr.: 33959). Mic.sens@1kHz/94dB = 34,6mV (-29.2dBV)

Theory

Valdemar has been used instead of human subjects to measure the transfer function of DT990 + Ear because the need to gather subjects was avoided. Various experiments gave results of good validity, and that artificial head provide the better performances compared to other artificial heads (Blauert et al., 2005, p. 239).

A MLS-sequence (Rife and Vanderkooy, 1989) has been used to reveal the transfer function of DT990 + Ear

Main procedure

In figure D.1 the measurement setup is shown.

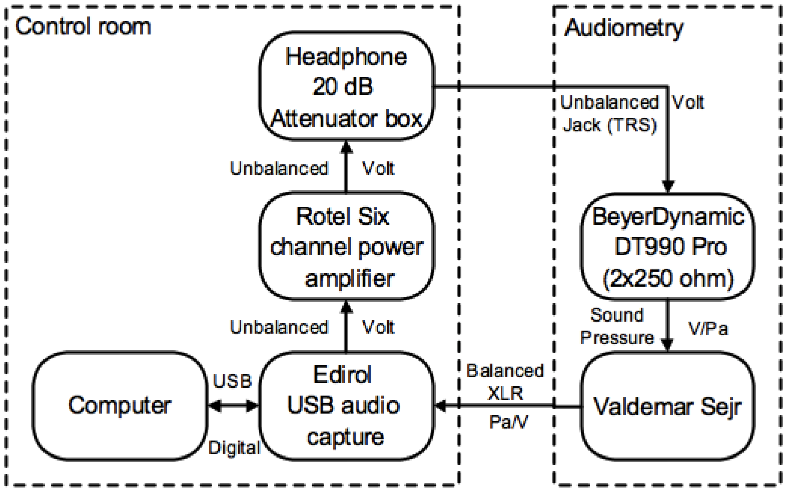


Figure D.1: *Setup for measuring the transfer function of BeyerDynamic DT990 Pro Headphones + Ears of Valdemar Sejr + Loop-back gain.*

To archive the transfer function of BeyerDynamic DT990 Pro Headphones + Ears of Valdemar Sejr, 5 steps is needed.

1. Measure transfer function of complete system[2].
2. Measure transfer function of loop-back-gain[3].
3. Measure microphone sensitivities[4].
4. Subtract full loop-back-gain from complete system.
5. Subtract microphone sensitivities from complete system.

It can be described in equations in the following way

$$H_C = H_{SC} + H_{Amp} + H_{Box} + H_{DT990} + H_{Ear} + H_{Mic}$$

$\Downarrow$

$$H_{DT990} + H_{Ear} = H_C - (H_{SC} + H_{Amp} + H_{box} + H_{Mic})$$

$\Downarrow$

$$H_{DT990} + H_{Ear} = H_C - (H_{Full} + H_{Mic}) \tag{D.1}$$

Where (see table D.1)

Variable	**Description**
H_C	TF of complete system
H_{SC}	TF of Edirol sound card
H_{Amp}	TF of amplifier
H_{Box}	TF of 20db attenuator box
H_{DT990}	TF of DT990 Headphones
H_{Ear}	TF of Valdemars´ Ears
H_{Mic}	TF of microphones
H_{Full}	$H_{SC} + H_{Amp} + H_{box}$

Table D.1: *Description of transfer function variables.*

[2]BeyerDynamic DT990 Pro Headphones + Ears of Valdemar Sejr + Loop-back gain
[3]Soundcard + Amplifier + Attenuator box
[4]Build-in in Valdemar Sejr

Procedure

The procedure for conducting the 5 steps is listed as follows.

Step 1: TF of complete system - Procedure

1. Connect all equipment as illustrated in figure D.1.
2. Turn Edirol Output knob to maximum and both microphone input sensitivities around half up.
3. Send out a MLS-sequence with the parameters listed in table D.2 to each headphone speakers (one at the time) and record the sound simultaneously.

MLS parameter	Number
Order	16
Repetitions of measurements	1
Repetitions of each measurements	4
Digital amplitude	0.25

Table D.2: *MLS settings*

4. If Soundcard peaks at the input or the recorded signal is distorted, turn the sensitivity down or lower the digital amplitude of the MLS signal.
5. Repeat the measurement 4 times by remounting the headphone each time.

Step 1: TF of complete system - Results

In figure D.2 5 repetitions and the averaged transfer function of left and right channel of the complete system is shown. As seen each repetitions are almost equal.

Step 2: TF of Loop-back-gain - Procedure

Step 2 requires more steps and can not be directly measured. The problem is that the signal from the headphone attenuator box (see figure D.1) is unbalanced. If the unbalanced signal is fed directly into the Edirol sound card the internal gain is different than feeding it with a balanced signal which is the case when it receives the signal from the microphones in Valdemar. The solution is to convert the unbalanced to a balanced signal. This is done by using a linedriver. The problem by introducing a linedriver to the feedback-loop is that it introduces a new unknown gain-factor to the system. In figure D.3 the procedure for measuring the gain at 1kHz is shown. The gain of the linedriver at 1kHz was measured to be -12dB.

The next stage is to know the gain introduced by the Edirol sound card. In figure D.4 the experimental setup for measuring the transfer function of the

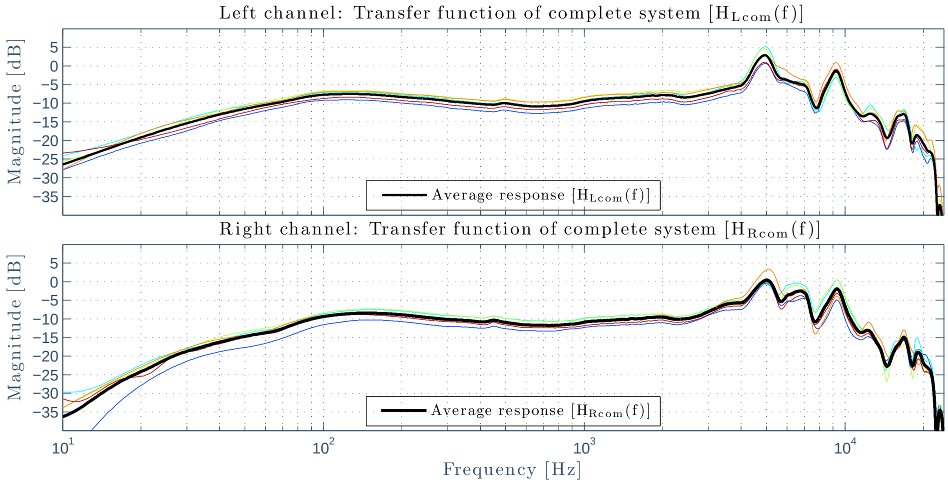


Figure D.2: *5 repetitions and the averaged transfer function of the complete system of left and right channel*

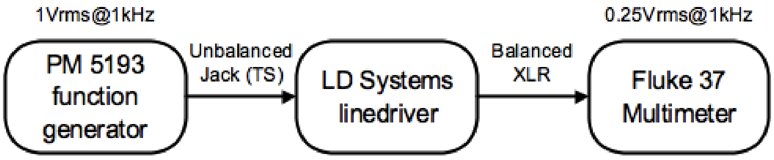


Figure D.3: *Procedure for measuring the Linedriver gain at 1kHz*

Edirol sound card is shown. The output knob and sensitivity knobs should

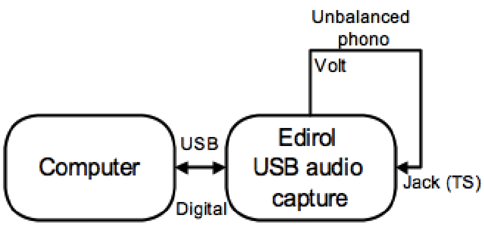


Figure D.4: *Experimental setup for measuring the TF of Edirol sound card.*

be the same as used in step 1. The shape of the transfer function is correct, but the gain is wrong since the input signal into Edirol is unbalanced. In figure D.7 on page 81 the measured transfer function is shown (labeled as "Soundcard (Wrong gain)").

In figure D.5 the experimental setup for measuring the TF of Edirol sound card + Linedriver is shown. Now the input signal into Edirol is balanced, but the gain at 1kHz is attenuated by 12dB caused by the Linedriver. In figure D.7 on page 81 the measured transfer function is shown (labeled as

"Soundcard + DI").

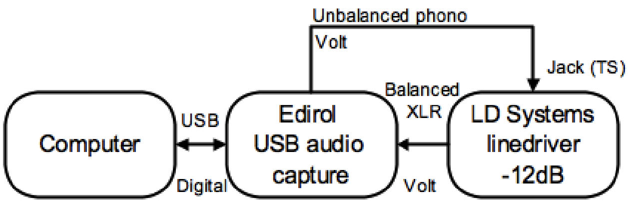


Figure D.5: *Experimental setup for measuring the TF of Edirol sound card + Linedriver.*

In figure D.6 the experimental setup for measuring the Full loop-back gain + Linedriver is shown. The load caused by BeyerDynamic DT990 Pro Headphones + Ears of Valdemar is included to have nearly the same burden to the system. In figure D.7 on the facing page the measured transfer function is shown (labeled as "Full + DI").

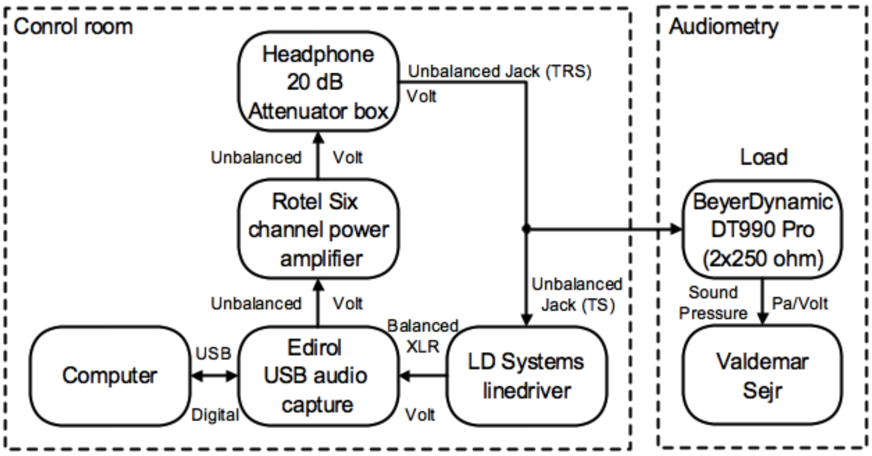


Figure D.6: *Experimental setup for measuring the Full loop-back gain + Linedriver.*

Step 2: TF of Loop-back-gain - Results

In figure the measured left and right channel transfer functions of "Soundcard (Wrong gain)", "Soundcard + DI" and "Full + DI" is shown. It can be seen that all transfer functions is reasonable flat in the frequency area from about 30 Hz→20kHz. The influence of the DI is that it attenuates the signal by 12dB and it adds a steeper high-pass-filtering below 30 Hz.

By adding 12dB to the measured transfer function of "Soundcard + DI" the correct gain at 1kHz of Soundcard can be found. By re-scaling "Soundcard (Wrong gain)" with the gain difference at 1kHz from itself and "Soundcard + DI" + 12dB, the correct transfer function of Soundcard with the correct gain can be achieved. This corrected response is shown in figure D.8.

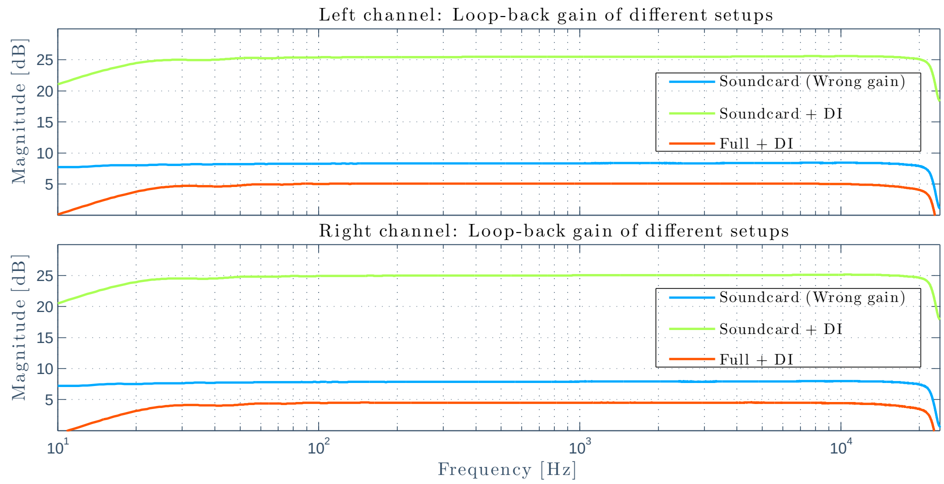


Figure D.7: *Left and right channel transfer functions of "Soundcard (Wrong gain)", "Soundcard + DI" and "Full + DI"*

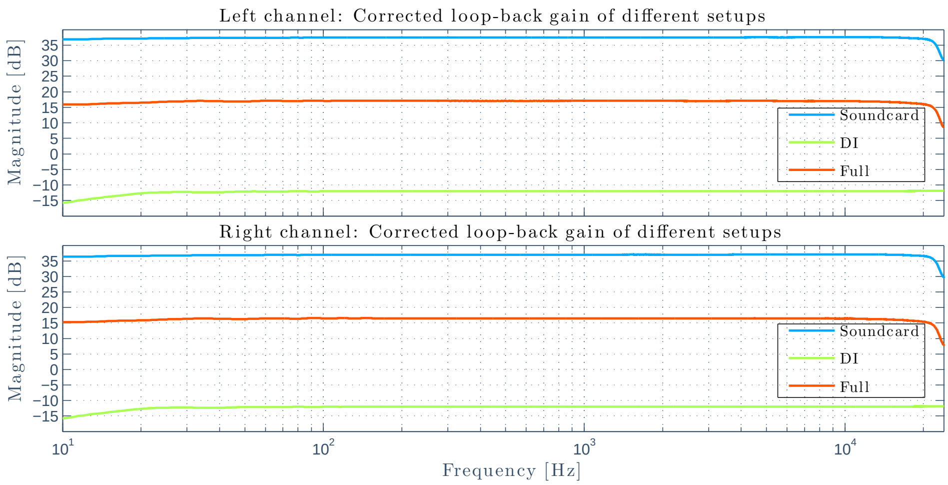


Figure D.8: *Left and right channel transfer functions of "Soundcard (Wrong gain)", "Soundcard + DI" and "Full + DI"*

By subtracting the correct gain of Soundcard from "Soundcard + DI" the correct transfer function of the DI is achieved. By subtracting the correct DI gain from "Full + DI" (Seen in figure D.7) the correct transfer function of the full loop-back-system can be found. Both responses is shown in figure D.8.

Step 3: Measurement of microphone sensitivities - Procedure

1. Connect left channel output from Valdemar to mulitmeter
2. Screw both ears of Valdemar off
3. Mount sound calibrator to left channel microphone

4. Read measured Vrms

5. Redo for right channel

Step 3: Measurement of microphone sensitivities - Results

The sensitivities of Valdemar build-in microphones at 1kHz/94dB were measured to:

Left channel: 35,4mV (-29,0dBV)
Right channel: 34,6mV (-29.2dBV)

From the Calibration Chart (G.R.A.S) the frequency response of G.R.A.S, type 40AD is approximately flat up to about 10 kHz as seen in figure D.9.

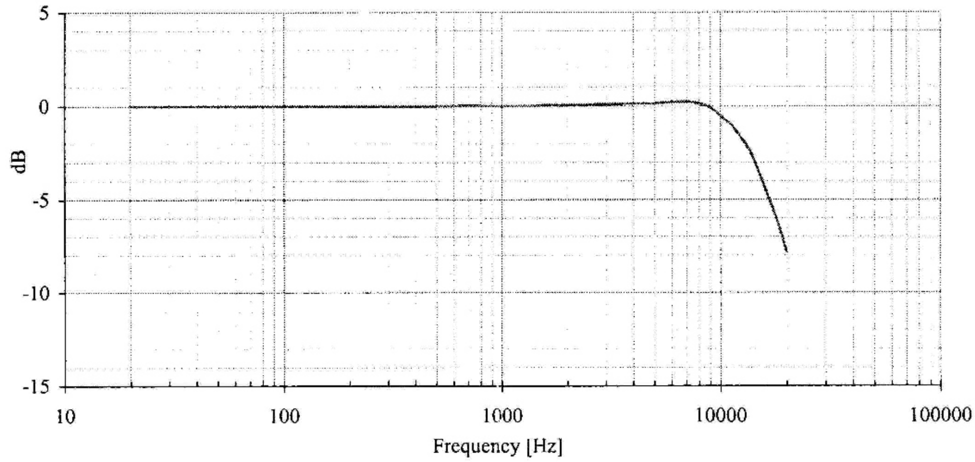


Figure D.9: *Frequency response of G.R.A.S, type 40AD. Build-in microphones in Valdemar*

Results

The TF measurement of the complete system H_C has been done in step 1. The TF measurement of the full loop-back-gain H_{Full} has been done in step 2. The measurement of microphone sensitivities has been done in step 3. Left is step 4 and 5, to fulfill equation D.1 on page 77. In figure D.10 the transfer function of $H_{DT990} + H_{Ear}$ is shown. Above 10kHz the transfer function is not only affected by the $H_{DT990} + H_{Ear}$ because of the step roll-off caused by the microphone transfer function shown in figure D.9.

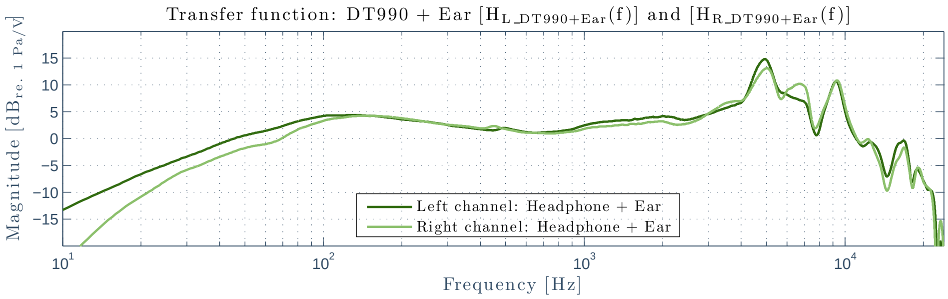


Figure D.10: *Left and right channel transfer functions of $H_{DT990} + H_{Ear}$*

Discussion

In (Blauert et al., 2005, p. 231) human variations above 7 kHz across subjects are highly individual. This suggest only to apply headphone equalisation up to around this frequency, since further compensation will not be true for the general case.

Conclusion

The transfer function of DT990+Ear has been measured using Valdemar, with 5 repetitions on each ear. The average of these has been scaled according to the full loop-back-gain and by the sensitivity of G.R.A.S, type 40AD microphones. This yields the true acoustic/electrical relationship of DT990+Ear in the units of $[\text{dB}_{\text{re. 1 Pa/V}}]$, from 20Hz→≈10kHz where the frequency response of G.R.A.S, type 40AD starts to roll-off.

APPENDIX E

Enclosed DVD contents

Aco10main.pdf: Digital version of the report.

Code/

- **GUI/**
  - gui_impulsiveness.m:
    Program used for the psychoacoustic experiment (requires playrec).
  - gui_annoyance.m:
    Program used for the psychoacoustic experiment (requires playrec).
  - gui_example.m:
    Example given to the participants before the actual session (requires playrec).
- **Minimum_Phase_Filter/:**
  - mpf2go.m:
    Function for correcting the transfer function of an input and creating minimum-phase filter impulse response from it.

Plots/

- **Misc/:**
  A variety of plots used in the report
- **Sample_Ratings/:**
  All subjects' ratings for each sample.

References/
Papers and manuals used in the report

Samples/
Samples used for the subjective and objective evaluation

Bibliography

Music Acoustics, The University New South Wales, Sidney Australia, Maj 2012. URL http://www.phys.unsw.edu.au/jw/hearing.html.

Residual Analysis, Maj 2012. URL http://www.weibull.com/DOEWeb/residual_analysis.htm.

ANSI. ANSI S1.11-2004: Specification for octave-band and fractional-octave-band analog and digital filters, 2004.

S. Bech and N. Zacharov. *Perceptual Audio Evaluation - Theory, Method and application.* Number ISBN: 0-470-86923-2. John Wiley & Sons, Ltd, 2006.

J Blauert, J Braasch, H Fastl, V Hamacher, D Hammershoi, U Heute, I Holube, U Jekosch, G Klump, A Kohlrausch, A Lacroix, H Moller, S Möller, J N Mourjopoulos, P Novo, and S van der Par. *Communication Acoustics.* Springer, 2005.

M. Blommer, A. Eden, and S. Amman. Sound Quality Metric Development and Application for Impulsive Engine Noise. *Processing*, 2005.

W. Bray. Comments about loudness measurement, and filtering. SAE SQ Workshop, 2007.

G. Cerrato. Automotive Sound Quality - Accessories , BSR , and Brakes. *Sound & Vibration*, 2009. URL http://www.sandv.com/downloads/0909cerr.pdf.

U. K. Chandrika and J. H. Kim. Development of an algorithm for automatic detection and rating of squeak and rattle events. *Journal of Sound and Vibration*, 329(21):4567–4577, October 2010. ISSN 0022460X. doi: 10.1016/j.jsv.2010.05.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0022460X10003081.

F. Christensen, C. B Jensen, and H. Møller. The design of valdemar - an artificial head for binaural recording purposes. In *Audio Engineering Society Convention 109*, 9 2000. URL http://www.aes.org/e-lib/browse.cfm?elib=9085.

W. Davies, M. Adams, N. Bruce, R. Cain, P. Jennings, A. Carlyle, P. Cusack, K. Hume, and C. Plack. A positive soundscape evaluation tool. In *Proceedings of Euronoise*, October 2009.

DIN. Draft DIN 45631/A1: Calculation of loudness level and loudness from the sound spectrum — Zwicker method — Amendment 1 : Calculation of the loudness of time-variant sound, 2007.

ETSI. *ETSI EG 202 396-3 Speech Processing Transmission and quality aspects Part3 background noise transmission objective test methods*, volume 1. 2011. ISBN 2005012513.

H Fastl and E Zwicker. *Psycho Acoustics - Facts and models*. Number 978-3-540-23159-2. Springer, 3rd edition, 2006.

J. Feng and J. Hobelsberger. Detection and Scaling of Squeak & Rattle Sounds. In *Society of Automotive Engineers, Inc.*, number 724, page 12, Traverse City, Michigan, 1999.

Genesis. History and description of loudness models, 2009. URL `www.genesis.fr`.

Genesis. Loudness toolbox, Maj 2012. URL `http://www.genesis-acoustics.com/en/index.php?page=32`. http://www.genesis-acoustics.com/en/index.php?page=32.

K Genuit. Objective evaluation of acoustic quality based on a relative approach. *Internoise 1996, Liverpool*, pages 1–6, 1996. URL `http://www.head-acoustics.de/downloads/publications/hearing_related/Internoise96_Objective_Evaluation_of_AQ.pdf`.

K. Genuit and W. Bray. Dynamic Acoustic Measurement Techniques Considering Human Perception. *Journal of Vibration and Acoustics*, 130(3): 031005, 2008. ISSN 07393717. doi: 10.1115/1.2827453. URL `http://link.aip.org/link/JVACEK/v130/i3/p031005/s1&Agg=doi`.

B. R. Glasberg and B. C. J. Moore. A Model of Loudness Applicable to Time-Varying Sounds. *Audio Eng. Soc.*, 50(5):331–342, 2002.

G.R.A.S. *Calibration Chart - G.R.A.S, type 40AD 1/2 inch Prepolarized Microphone*.

S. W. Hainsworth and M. D. Macleod. Analysis Of Reassigned Spectrograms For Musical Transcription. *Applications of Signal*, pages 1–4, 2001. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=969533`.

M. O. Hawksford. Digital signal processing tools for loudspeaker evaluation and discrete-time crossover design. *Journal of the Audio Engineering Society*, 45:37–53, 1997.

S. I. Hayek. Nearfield Acoustical Holography. *Handbook of Signal Processing in Acoustics*, 2009. URL `http://www.springerlink.com/index/nk062m46477362n1.pdf`.

LMS International. Sound quality and psycho-acoustic metrics. Technical report, LMS International, 2012.

F. Kavarana and B. Rediers. Squeak and Rattle-State of the Art and Beyond. *Sound and Vibration*, 35(4):56–65, 2001. URL `http://www.sandv.com/downloads/0104kava.pdf`.

S. G. Mallat. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 1993. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=258082`.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-time Signal Processing*. Number ISBN: 0-13-083443-2. Prentice Hall International, Inc., 3rd. edition, 2008.

J. Plunt. Impact sound quality of consumer products evaluation by sound quality-metrics and wavelet time-frequency analysis. *Proceedings of Euronoise*, pages 1–6, 2006. URL `http://www.afconsult.com/upload/konsulttjanster/LjudoVibrationer/pdf/euronoise2006_Paper417.pdf`.

D. D. Rife and J. Vanderkooy. Transfer-function measurement with maximum-length sequences. *J. Audio Eng. Soc.*, 37(6), June 1989.

S. H. Shin and C. Cheong. Experimental characterization of instrument panel buzz, squeak, and rattle (BSR) in a vehicle. *Applied Acoustics*, 71(12):1162–1168, December 2010. ISSN 0003682X. URL `http://linkinghub.elsevier.com/retrieve/pii/S0003682X10001702`.

W. Song and H. Saito. Improved Noise Source Identification Using Sound Quality Metrics Mapping in Vehicle Noise Measurements. *SAE Technical Paper*, 2011. doi: 10.4271/2011-01-1671. URL `http://papers.sae.org/2011-01-1671`.

R. Sottek. *Modelle zur Signalverarbeitung im menschlichen Gehör*. PhD thesis, RWTH Aachen, Germany, 1993.

R. Sottek and K. Genuit. Models of signal processing in human hearing. *AEU - International Journal of Electronics and Communications*, 59(3):157–165, June 2005. ISSN 14348411. doi: 10.1016/j.aeue.2005.03.016. URL `http://linkinghub.elsevier.com/retrieve/pii/S1434841105000701`.

R. Sottek, P. Vranken, and G. Busch. Ein Modell zur Berechnung der Impulshaltigkeit.

R. Sottek, W. Krebber, and G. R. Stanley. Tools and Methods for Product Sound Design of Vehicles. *Technology*, 2005.

M. Trapp and F. Chen. *Automotive Buzz, Squeak and Rattle: Mechanisms, Analysis, Evaluation.* Oxford : Butterworth-Heinemann, 2008.

B. D. V. Veen. Beamforming: A versatile approach to spatial filtering. *ASSP Magazine, IEEE,* (April), 1988. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=665`.