AALBORG UNIVERSITY

Master Thesis

Optimal Music Reproduction in Noisy Environments

Authors:
Johnni Thomsen Pedersen
Louise Baldus Vestergaard

Supervisor:
Flemming Christensen

Aalborg University
Department of Electronic Systems
Frederik Bajers Vej 7
9220 Aalborg Ø
Telephone 96 35 86 00
http://es.aau.dk/

Title: Optimal music reproduction in noisy environments

Project term:
ACO10, Spring 2010

Project group:
Group 1060

Members of the group:
Johnni Thomsen Pedersen
Louise Baldus Vestergaard

Supervisor:
Flemming Christensen

Number of copies: 5

Number of pages: 99

Attachments:
1 CD-ROM

Completed 3/6 2010

Synopsis:

This master thesis examined the effect noise can impose on the perception of music. Focus was put on achieving naturalness in the reproduction and minimizing the annoyance caused by noise.
Focusing on transit noise in trains in particular, two methods for optimizing music reproduction in noisy environments was suggested.
One model primarily focused on making the music audible at all frequencies.
The other method contained the functionality of the first one, as well as a subtle equalization to avoid the perceived change in the tonal balance.
The two methods were tested using human subjects in a paired comparison test. The test was carried out with music processed using the two methods.
Results were rather inconclusive, but a straight ranking was achieved. Overall, method two were preferred for optimizing the naturalness and decreasing the annoyance.

Authors

This thesis is written by:

Johnni Thomsen Pedersen

Louise Baldus Vestergaard

Preface

This report documents the work done by group 1060 at the Section of Acoustics, Department of Electronic Systems at Aalborg University during 4th semester of the acoustics master program of Aalborg University in the spring semester.

The documentation consists of the following parts:

Part I contains some of the basic theory on related to the project.

Part II describes the design of the methods proposed.

Part III describes the method and results of a listening experiment, and discusses the results.

Part IV concludes on the report, and outlines which future studies in this field might be interesting.

Part V includes the appendices for the report.

CD contains all MatLab - and sound files mentioned in the report, as well as a copy of the report and the collected data

The group would like to thank all the subjects that voluntarily have participated in the experiment.

Contents

From the concert hall to the truly portable musical experience, the possibilities for listening to music when and where the listener prefers have never been more unrestricted. But the respectful silence of the audience in a concert hall is often not part of the public reality.

1 Introduction

Methods for music reproduction has evolved tremendously from the late 19th century's phonographs to todays variety of analog and digital playback equipment. Even in the early days of reproduced music, developers were designing portable equipment, such as travel gramophones. Even though it was portable, it was still not suitable for everyday use while traveling back and forth to work, or while exercising.

In 1979, Sony Corporation launched the "Walkman", a portable platform to replay cassette tapes. It was delivered with headphones and enabled the user to listen to music basically anywhere. People now had the opportunity to pass time to and from work, listening to music. The emergence of the "discman", solid state memory, the MPEG2 - layer 3 technology and popular products like the Ipod have moved the target market of transportable music from a select few individuals to almost anyone.

The first transportable platforms used supra-aural headphones and these are still quite common. They do, however, still allow a significant amount of external sound to pass. Later, the circumaural type mitigated this problem to some extent and the newer in-ear types even more so. A total isolation from outside noise is in most cases not possible, and as a consequence the listener will either have to accept some amount of annoyance or, alternatively, turn up the volume. In the first case the quality of the listening experience is degraded. In the second case hearing damage can be a possible consequence, as newer MP3 players and in-ear headphones are capable of outputting extensive acoustic power.

Various possible approaches for optimizing the listening experience could be chosen. In the following, the approach of this research will be outlined.

2 Approach

This thesis investigates how noisy environments influences the listening experience, and aims for an optimization of the music perception in a noisy environment. In order to meet these goals, different steps must be taken as to examine the issues related to this problem. Perception, noise characteristics, playback equipment and effects of either one of these on the other needs to be examined before a method for possible improvement can be determined.

In Chapters 3, 4 and 5 the physiology of the ear, as well as the perceptual effects of masking and loudness will be examined. The key point in relation to optimization is to avoid that the noise masks the signal to a crucial degree. The loudness perception plays an important role, since uncorrelated signal and noise increases the sound pressure level, and might constitute to a listening level that could cause hearing damage for the listener. Therefore, both the physiology and the perceptual effects play an important role in determining a method for improving the listening experience.

In Chapter 6 some characteristics of importance regarding the noisy environment will be outlined. These theoretical considerations will form the basis for a problem statement - and description.

As stated, also the playback equipment plays a significant role in determining the sound experience for the listener. In Chapter 7 the headphone types available on the marked will be outlined, since the variation here is of significance to a possible solution.

All these theoretical considerations will form the basis for a problem statement, which will be given in Chapter 8.

Part I

Theory

A quick tour of the physiology of the ear, namely the concept of "critical bands".

3 Physiology of the Ear

To be able to asses the importance of different parameters, it is necessary to understand what mechanisms that govern these parameters. For the discussion of both masking and loudness models, a basic understanding of the physiology of the ear is needed. The topic has been covered extensively in many books, e.g. [Moo03, Chapter 1], [Moo95, Chapter 2] and [Mø06].

3.1 The Outer Ear

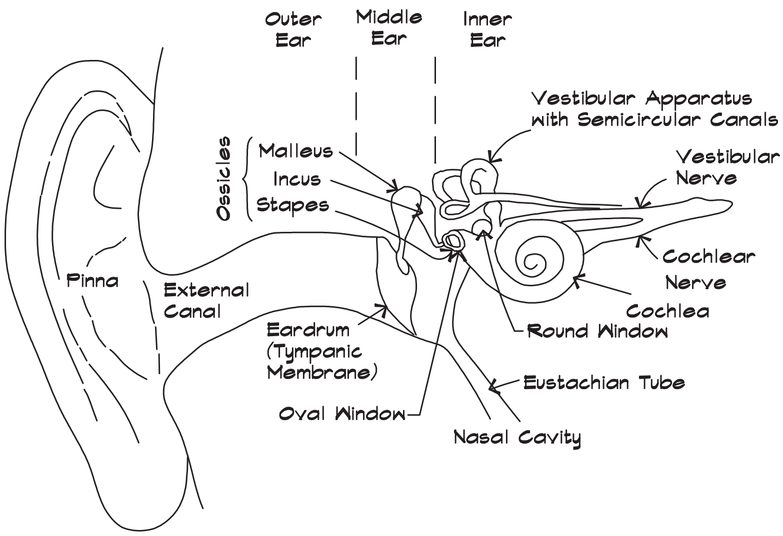


Figure 3.1: The auditory system (not necessarily to scale), auditory cortex excluded [Lon05, p.74].

The Pinna (see Figure 3.1) offers both amplification of the incoming sound and plays an important role in sound localization. Especially the high frequency content of an incoming signal will be altered. The Pinna, together with the Auditory Canal, make up the outer ear.

3.2 The Middle Ear

The middle ear consists of the Tympanic Membrane and the Ossicles, Malleus, Incus and Stapes. The Malleus is connected to the Tympanic Membrane and the Stapes is connected to the inner ear via the Oval Window. The main function of the middle ear is a mechanical impedance transformation, matching the area of the Tympanic Membrane to that of the Oval Window. The Ossicles conductance of sound can be (unconsciously) controlled, decreasing the sensitivity to ones own voice. This Auditory Reflex is also activated, when exposed to extreme sound levels.

3.3 The Inner Ear

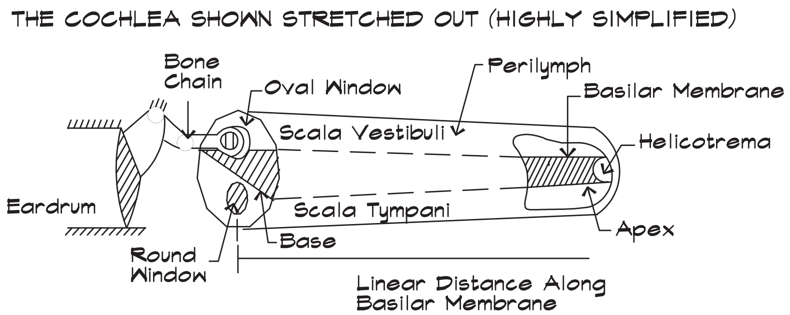


Figure 3.2: The unfurled cochlea [Lon05, p.74].

The inner ear is is made up from the Cochlea. The Cochlea contains Scala Vestibuli and Scala Tympani, that are cavities filled with almost incompressible fluid. The cavities are separated by the Basilar Membrane which has a small hole, the Helicotrema, that allows passage of the fluid. If the Cochlea is "unfurled", it is basically a tube, as seen in Figure 3.2. The Helicotrema is placed at the very end of this tube, called the Apex. Vibration from the Stapes is transmitted to the Cochlea through the Oval Window. Since the fluid in the ear is not easily compressible, the Round Window serves as pressure relief. A given sound will result in a given excitation pattern on the Basilar Membrane which can, somewhat simplified, be viewed as a frequency analyzer as in Figure 3.3. A high frequency will excite the Basilar Membrane close to the Stapes, while lower frequencies will excite the BM closer to the Apex.

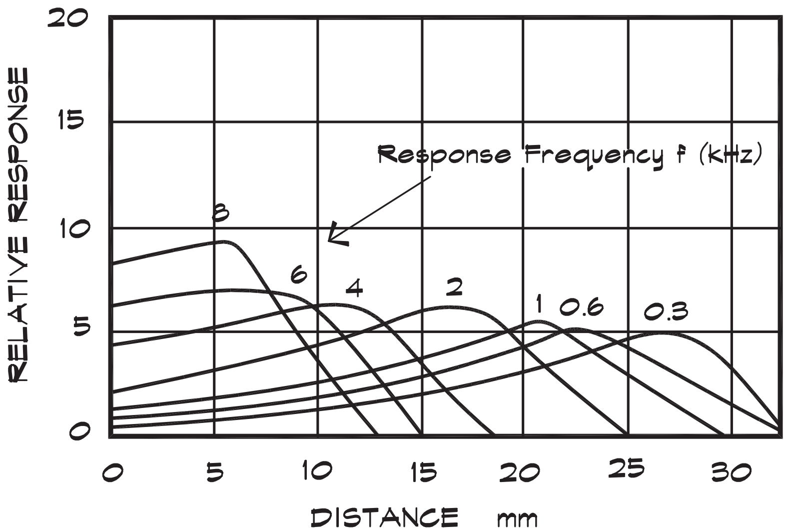


Figure 3.3: The displacement of maximum excitation on the Basilar Membrane [Lon05, p.77]

3.4 Critical Bands

While the former description of the ear have primarily been a consequence of the mechanical construction and measurements, the exact neurological workings are still being discussed. There does however seem to be a general agreement on the concept of "critical bands". The idea of frequency selective bands have been mentioned by Fletcher and Munson [FM37] and in 1940 they *"measured the threshold of a sinusoidal signal as a function of the bandwidth of a bandpass noise masker"* [Fle40],[Moo03, p. 66]. Fletcher found that, as the bandpass limited noise increased in bandwidth, the sinusoid became harder to detect. But after a certain increase in noise bandwidth, the detection of the sinusoid *"did not become harder"*. He concluded that when the bandwidth of the noise exceeded that of the auditory filter used to detect the sinusoid, it was not (or only to a small degree) masking the tone. For sake of ease, these filters are often recalculated/estimated as their "equivalent rectangular bandwidth", ERB. This is practical, as no matter what shape these filters have, they can be put into an comparable form. Later, these filters and their shape will make a base for calculations of perceived loudness from complex waveforms as described by Zwicker, Flottorp and Stevens in [EZS57].

The presence of a loud sound makes it harder to hear others. This is why we only realize the phone is ringing the second we turn off the vacuum cleaner.

4 Masking

This chapter includes the basic theory on masking and masking models, relevant for this thesis. Masking can make music passages completely inaudible. According to Moore [Moo03], masking is defined as: *"The amount (or the process) by which the threshold of audibility for one sound is raised by the presence of another (masking) sound."*

Depending on the environment in which sound is perceived, masking could be both positive and negative. When a disturbing or annoying sound is reduced due to masking, the masking is considered a positive effect. In an environment where it e.g. inhibits speech perception, the noise is an undesired masker.

Masking induces a shift of the hearing threshold. The hearing threshold can be determined experimentally in a noise-free environment. When adding a masking sound source, the hearing threshold can be increased. When masking sounds are present, the perceived loudness of the target sound as well as the hearing threshold are affected. When the loudness of a target sound is reduced but not entirely masked in the presence of another sound, the sound is said to be partially masked [ZS65].

For the signal as well as for the masker two variables are of relevance: the level and the frequencies, at which it occurs. The temporal relationship between the the two constitute the masking pattern.

4.1 Masking Methods

There are two main types of masking; spectral masking and temporal masking. Spectral masking is used for simultaneous masking where the masker is present simultaneously with the sound. Temporal masking is when the masker and the sound are consecutive to each other instead of being present simultaneously. The different masking methods are suitable for different causes as will be described in the following.

4.1.1 Simultaneous Masking

Simultaneous masking is a frequency domain phenomenon [HC02]. Here the masker and the sound are present at the same time.

The amount of simultaneous masking is dependant on how spectrally far from each other the signal and masker are. The closer, the more profound the masking effect. Due to the excitation patterns in the cochlea, a masker is more efficient at masking frequencies higher than the masker itself. This means that low frequency noise has a high potential of masking a given signal.

4.1.2 Temporal Masking

Temporal masking is, as stated, non-synchronously played masker and sound. If the sound is played before the masker the masking effect is called backwards masking or post-masking. If the masker on the other hand is played before the sound it is called forward masking or pre-masking. Zwicker illustrated this as in Figure 4.1. It is a phenomenon first and exclusively described by Zwicker [FZ90, p. 82-93].

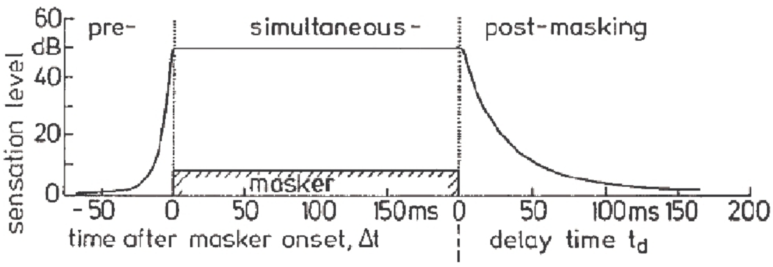


Figure 4.1: Frequency and temporal masking effects.[FZ07]

4.1.3 Masking Sounds

Sounds used for masking can span from the simple pure tone, which has the narrowest spectra to narrowband, broadband and white noise, the latter having a flat power spectra. The masking sound can also be either stationary or exhibit spectrally dynamic properties.

The perceived tonal balance is dependent on the sound level. This is part of the explanation of why "louder" often sounds better.

5 Loudness

Loudness is defined as that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud [Moo03, p.127]. In other words it is the perceived intensity of a given sound.

5.1 Equal Loudness Level Contours

Mapping of loudness versus frequency was pioneered by Fletcher, Munson and, to some degree, Steinberg [FM33, FM37, FS24, Ste25]. Fletcher and Munson investigated the perceived loudness of both pure and complex sinusoids. A result of their work is the famous Fletcher-Munson curves, as seen in Figure 5.1. These types of curves is called *equal loudness level contours*, ELLC's.

They found that the sensitivity versus frequency is far from linear. At all recorded sound pressures, a pronounced rise in sensitivity at the middle frequencies is seen. At low sound pressures, frequencies below 300-400 Hz and above 6-7 kHz are perceived much less intense, compared to the 1 kHz reference. As the sound pressure increases, the sensitivity becomes somewhat more linear in a process that can best be described as compression.

Another pioneer in this field was Zwicker, who in the late 1950's examined the correlation between loudness, masking and intensity differences of a signal, and how to evaluate these subjective characters [Zwi58].
His comparison of the subjective experience of loudness led to some of the first loudness curves.

Later these curves have been further refined and updated, most recently in the ISO 226:2003 standard. These revised equal loudness contours can be seen in Figure 5.2 and differs quite a lot from the earlier contours.

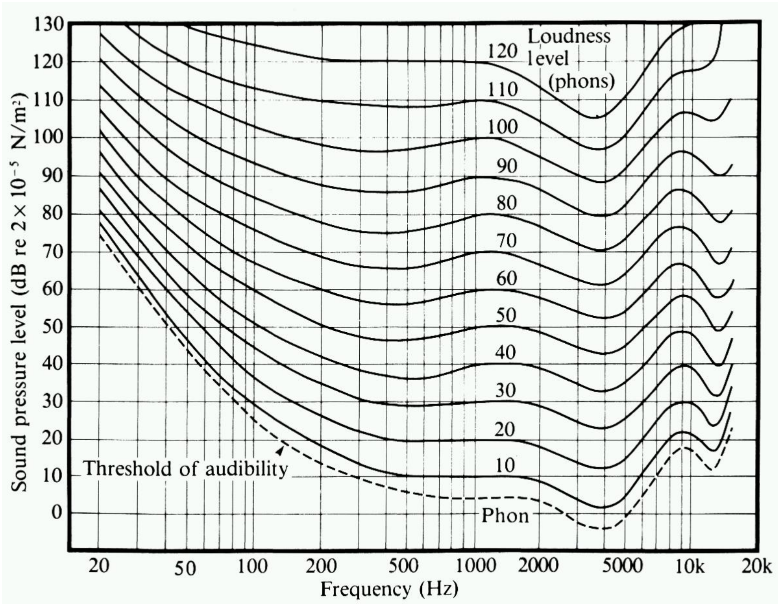


Figure 5.1: The Fletcher-Munson equal loudness contour curves [FM33]

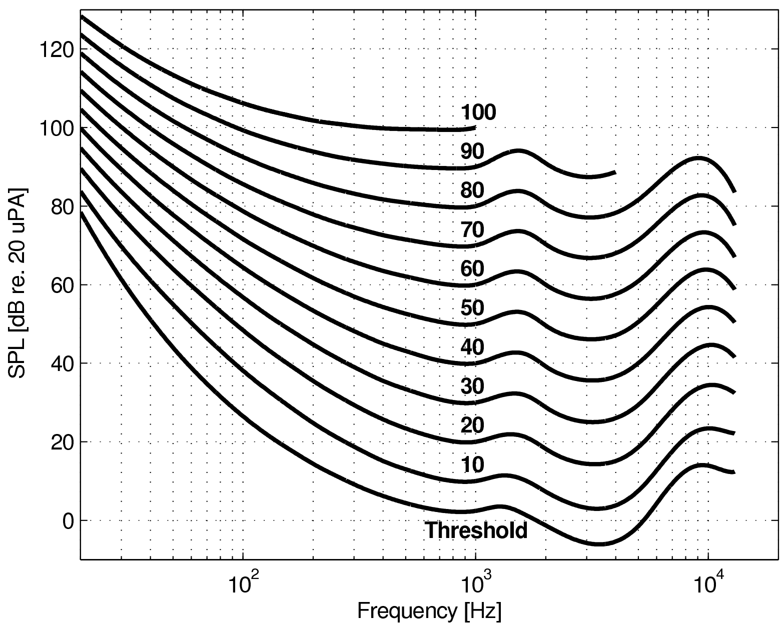


Figure 5.2: Revised equal loudness level contours from ISO 226:2003. Bold values are in Phons.

It is important to notice that these curves does not necessarily give the whole picture, as every possible acoustical stimuli have not been used. Nevertheless, they have been the base for later weighting curves.

5.2 Loudness Assessment

The original weighting curves have been used to approximate the perception of sound pressure vs. frequency and for estimating possible long term hearing damage. The most used curves are the A and C weighting curves seen on Figure 5.3 and to some extent the D curve (not shown) for measuring airplane noise.

5.2.1 Weighting Curves

The A curve is a 4th order approximation to the equal loudness level contours at relatively low sound pressure levels. It is normally used up to around 80 dB(A). As the sound pressure rises, the C curve is a closer approximation to perceived loudness. The C curve is often used when assessing industrial noise.

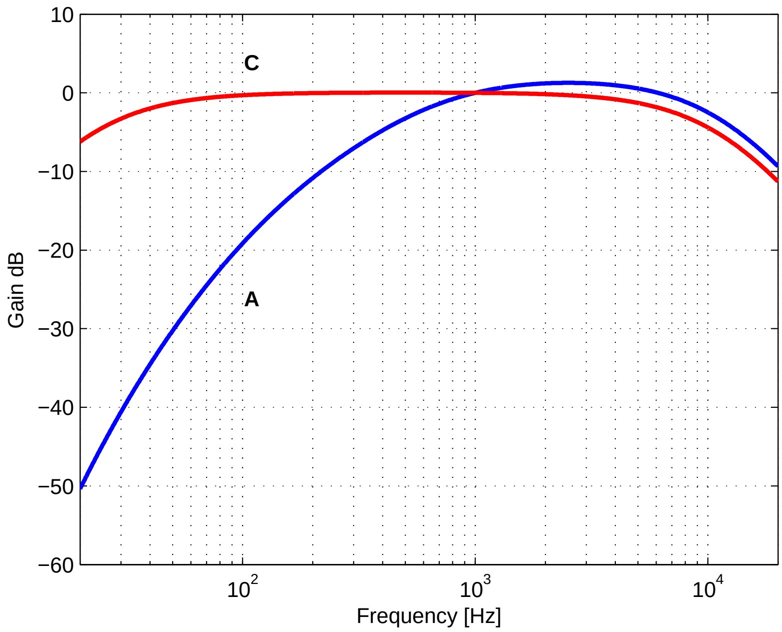


Figure 5.3: Transfer function of the A and C weighting.

Both A and C weighting have a great advantage as it is rather simple to do both analog and digital implementations of these. They have also been used for a long time, making it quite easy to compare older and newer data. But if a better loudness approximation is needed, the simple weighting functions are no longer adequate.

5.2.2 Zwicker, Moore and Glasberg's Models

Zwicker suggested an alternate method for loudness assessment in [Zwi60], based on the shape of the auditory filters and the effect of masking. Recommendations on using the method are given in [ISO75]. The SPL in each of the frequency bands are recorded and weighed by an ELLC. If the SPL rises during the transition from a lower frequency to a higher frequency, the rise is instantaneous. If the SPL drops in this transition, a smoothing of the transition is made. This corresponds to the masking effect of tones of slightly higher pitch than the masking tone. An example of this can be seen on Figure 5.4. In other words, the SPL of the neighbour higher frequency is increased, if it is lower than the neighbour.

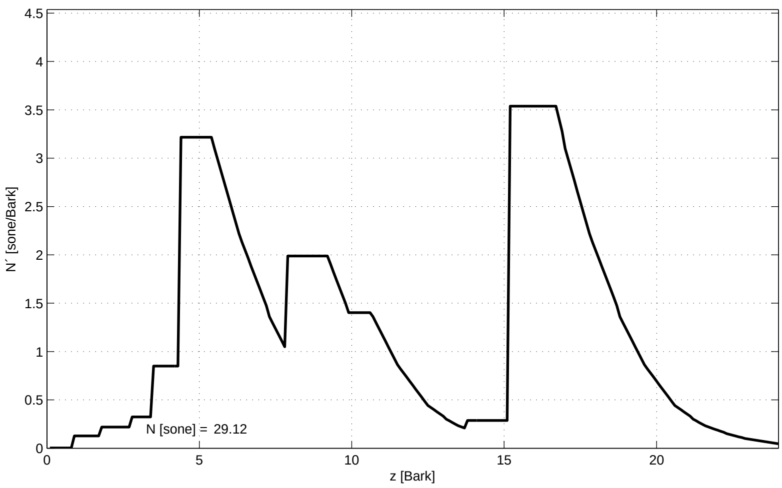


Figure 5.4: The excitation pattern for a signal composed of 500, 1100 and 3400 Hz.

When this has been done for all frequency intervals, the curve is integrated and a corresponding loudness value in Sones or Phons can be manually read. Basically the method is a weighted energy summation across frequency bands, taking the both masking phenomena and the loudness contour into account. In 1996, Glasberg and Moore suggested a revised method in [MG96] by, most importantly, using the much newer loudness curves found in ISO 226:2003.

5.2.3 Stationary vs. Non-Stationary Sound

For all of the previous models, the sound used for loudness assessment have been stationary. While this is adequate for sound e.g. from ventilation or other sources with a relatively constant frequency spectrum, loudness of music is a very different matter. As music is often very rhythmic, the "instant loudness", whatever that is, is wildly fluctuating. As the ELLC's are based on steady state sounds, it make little sense to assess the music as a stationary source. In 2002, Moore and Glasberg suggested another model that was better able to predict the "longterm loudness" of time-varying sounds [MG02]. "Longterm" can in this case be more than a second.

What types of noise is present when listening to music and what can be done to reduce the annoyance of said noise?

6 Mitigating Noise

In general noise is considered anything *but* the part of signal that is of interest. In our case everything other than the music reproduced is noise. Noise will do more than just cause annoyance. As the music and noise is uncorrelated, the net SPL will rise. As seen on the ELLC's on Figure 5.2, this will result in a slightly more linear contour, meaning that the perception of music will be tonally altered, possibly in frequency areas not in the vicinity of the noise that altered the contour in the first place.

6.1 Noise Spectrum

The term "relative bandwidth" is defined as $\frac{\Delta f}{f_c}$ and is strongly connected to the logarithmic character of hearing different frequencies. It is one of the reasons that octave intervals are often more used than decades in the audio frequency range. There are no hard distinction between narrow- and broad-band audio signals, so this must be estimated taking circumstances into account. In the case of noise super positioned on a musical signal, the term "narrow banded" will be used if more than half of the noise energy is contained in one 1/3 octave band. The reason this distinction is made is that 1/3 octave filter banks have been used to approximate the critical bands in other work and standards, notably [ISO75]. Although Zwicker suggested the Bark scale, for assessing loudness [Zwi61], the lower 1/3 octave bands can be added together, for a rather close estimate of said scale, as in [ISO75] that gives advice on using Zwicker's loudness summation.

6.1.1 Narrow Band Noise

In the case of narrow band noise, there are several possible music/noise scenarios:

1. Audible noise in a frequency range where very little or nothing is spectrally present in the music signal.
2. Audible noise in a frequency range where the music signal is spectrally present and audible.
3. Audible noise in a frequency range where the music signal is spectrally present but no longer audible, due to masking.

In case 1, very little can be done to hide the noise, besides raising the overall sound level. In worst case, this will not even be able to mask the noise, if the music is spectrally far from the noise, as the effect of masking is more profound, the closer the two notes are [Moo03, p. 87].

In case 2, the altered equal loudness contour could be taken into account, and a more tonally "correct" music presentation is possible using either equalization or compression and gain.

In case 3 it is not feasible to focus on loudness contours, instead the main focus should be to make the inaudible frequency areas at the very least audible.

6.1.2 Broad Band Noise

Roughly speaking, broadband noise is anything that isn't narrow band i.e. with less than half the noise power in any 1/3 octave band. In this case, the altered equal loudness contour curve could perhaps be reconstructed, using subtle equalization.

6.2 Processing Possibilities

As one of the goals is to achieve a better fidelity while maintaining a relatively low increase in SPL, the processing of the music can possibly be done in only rather small spectral intervals. Two of the methods commonly used is equalization and narrow band compression and subsequent gaining.

6.2.1 Equalization

Equalization is amplification or attenuation of a given frequency interval, with respect to another. The common consumer audio amplifier will have a "bass" and "treble" knob, to emphasize the low or high frequencies. Professional sound technicians will probably have experience with the parametric and/or graphic equalizer. A parametric equalizer will typically have 3 to 5 filter sections, where it is possible to adjust the specific parameters of each section, hence the name. These parameters are center frequency, Q/bandwidth and gain [Z̈02, p. 50-55]. A graphical equalizer is split into specific fixed center frequencies and bandwidth. It consist of many filter sections, typically one for each 1/3 octave band, where the gain is the only controllable parameter. The gain is often controlled by a vertical "slider" potentiometer, giving a rough graphical presentation of the transfer function.

For both types, the process is linear, among other things implying that the output is proportional to the input at a given frequency.

6.2.2 Narrow Band Compression

Compression is a non-linear process that attenuates the signal based on the amplitude of the waveform. The two main controllable parameters are threshold and compression ratio. An example is shown in Figure 6.1, with a threshold of -30 dB and a compression ratio of 3:1. As long as the signal does not exceed -30 dB of full scale, the gain is 0 dB. When the threshold is exceeded, the gain is reduced. If f.ex. the input is -21 dB's, 9 dB's over the threshold, the output will be -30 + 9:3 = -27 dB's. The dynamic gain is controlled by the input level, using either peak detection or, more commonly, short-time RMS values [Z̈02, p. 95-102]. It should be noted that while the dynamic range of the signal is decreased, the overall sound level can be, and very often is, increased by amplification without risk of clipping. This results in an overall loudness increase.

A narrow band compression is only performed in a specific frequency range. Other important parameters are "attack" and "release" time. Attack time allows a slight delay of the compression, so some dynamics is preserved and "release" is the delay or recovery time from a given compression. Sometimes it's also possible to choose between "soft" and "hard" knee, referring to the transition between the gain before and after the threshold value. In Figure 6.1 a hard knee type is depicted. A soft knee type would have had a round transition instead of the very direct change in gain.

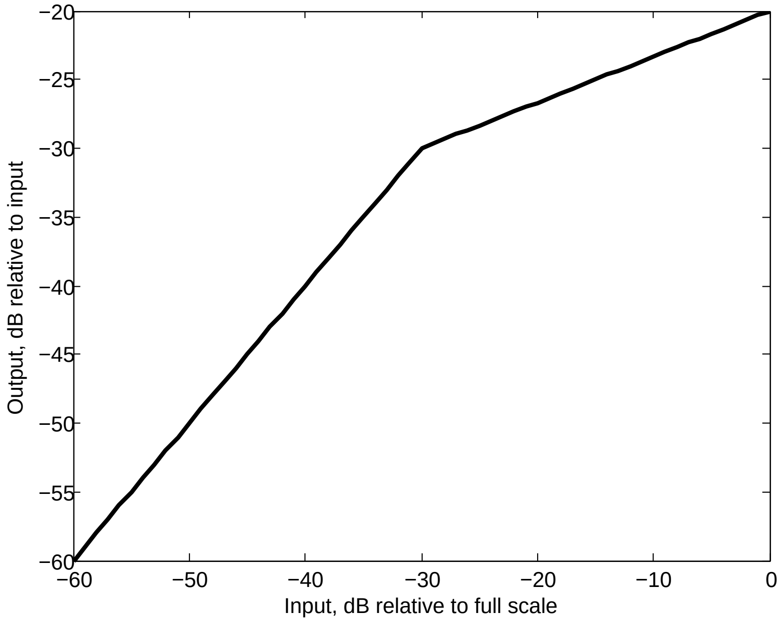


Figure 6.1: Output as a function of input.

6.3 Passive Mitigation

No matter how intricate digital or analogue signal processing solutions are available, the easiest and best solution is often to eliminate the root of the problem. Headphones comes in many different types and shapes. The insertion loss that the headphone causes is in fact a free noise mitigation. To be able to truly attenuate the lower frequency, the coupling from headphone to the tympanic membrane must be airtight. In most cases, this is not possible, but some newer headphone types, namely the "in-ear" type, are getting close. Some have a soft plastic tube to insure a very direct coupling. Besides attenuating external low frequencies, it also enhances the specific in-ear capability of reproducing low frequency audio.

As is the case for stationary consumer music reproduction systems, it is most often the transducers that have the largest impact on the reproduction quality. For portable personal music reproduction systems, this is almost exclusively in some form of headphones.

7 Headphones

Limitations have been done in Chapter 1 to focus only on music playback via portable devices and headphones. This induces a choice of headphone types, and possibly further limitations regarding playback environment.

Headphones provides a different playback environment opposed to loudspeakers, since the sound field is confined to a very small and limited volume.

A brief introduction to the different types of headphones will be outlined.

7.1 Closed Headphones

Closed headphones strives to eliminate the interference of the surrounding noisy environment thereby improving the listening environment. Both the circumaural headphones, that covers the entire outer ear and the canal phones, which are close fitting in-ear phones, often made with a rubber or plastic shield to be as close fitting in the ear canal as possible, are in this category.

7.1.1 Noise Cancelling Headphones

A somewhat new group of products on the market is the Noise Cancelling Headphones(NCH). Various companies[1] have developed both circumaural and canal headphones with an active noise cancellation(ANC). ANC is obtained by recordings from a microphone placed close to the ear and from the recorded signal "anti-noise" is generated, thereby reducing the noise floor and enabling lower sound power levels for a good music reproduction. The NCHs requires an independent power supply, and are often constructed so that the ANC can be switched on and off by the listener.

[1]Eg. Sony, JVC, Bose and Sennheiser.

7.2 Open Headphones

Most portable music players come with an enclosed set of open headphones, such as earphones or supra-aural headphones. Open headphones, opposed to closed headphones, does not exclude the sounds from the surrounding environment since they are not close-fitting. For open headphones it is often harder to create a controlled and undisturbed listening environment, however they are more suitable if you need to keep an awareness of the surrounding sounds while listening to music.

In this chapter the problem statement, specifications and limitations, which forms the underlying basis of the research, will be stated.

8 Problem Description

The use of portable music players have increased and are used in various environments with different noisy backgrounds. Though a proper choice of headphones might increase the listening experience in a given listening environment, it would be interesting to examine whether a signal processing approach to the problem could increase the quality of the listening experience for the listener.

There are two main parameters; the music signal and the noise. Multiple headphone companies have already produced active noise cancelling (ANC) headphones to counteract the effects of the noise, thereby improving the listening experience.

As listening to music is common during transit, it is of interest to optimize the sound quality in this kind of noisy environment. Reproduction of music in public transportation is relevant, as this mode of travel is particularly noisy. One could argue that it is the background noise itself that makes listening to music so appealing. Whatever the case, a reproduction method for optimization of the listening experience in noisy environments is the issue that this thesis will address.

8.1 Problem Statement

> **Problem statement**
> *The overall goal is to make it possible to reproduce music in a noisy transit environment with the same perceived tonal balance as without any noise.*

This goal ideally means to obtain a situation where the noise has no influence on the listening experience, but as a minimum it is to make the music noticeable in all the spectra, despite the noise. Between these "best" - and "worst" - cases, there is a reverse proportional scale, where the annoyance of the noise for the listener presumably decreases, as the clarity of the desired signal increases. Where the ANC headphones strive to block out all the noise, the approach in this project will be to compensate for the noise, regardless of headphone type. A sketch of the proposed method is displayed in Figure 8.1.

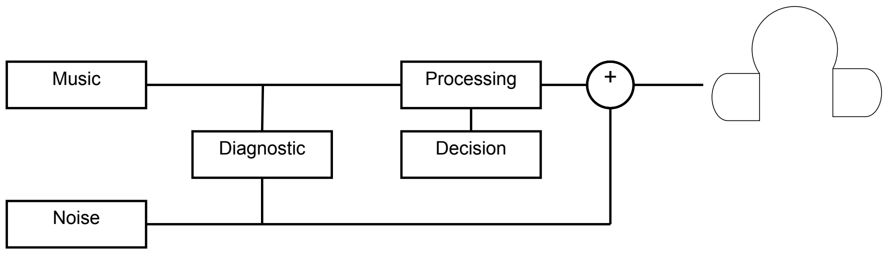


Figure 8.1: Block diagram of the system

Based on an evaluation of the music signal and the noise, a diagnosis can be made on how to counteract the effects of the masking, induced by the noise, without raising the overall loudness level. The processed signal, which takes the influence of the noise into account, is then fed to the headphones, enabling a "natural" listening experience.

8.2 Specifications

There are several cases to consider when dealing with this problem:

- The composition of sound and noise might constitute to a sound level above 85 dB. If so, the system should not try to compensate for the noise.
- There might be huge spectral differences between the sound and the noise. If this is the case, some sort of broadband approach must be taken, unless the spectral differences are so great, that the masking by the noise is irreversible.
- The noise could either be dynamic or stationary. Dynamic noise prompt an adaptive system, whereas stationary noise could be handled in various ways. In this thesis only stationary noise will be examined and

taken into account. If Stationary noise can effectively be mitigated, quasi-stationary noise mitigation can be considered.

- The frequency content of the noise is of great importance. In transit noise the low frequency noise is particularly crucial, hence this thesis will be limited to examining and handling only the low frequency noise, meaning noise below a few hundred hertz.

- In a real situation the noise must be known. It could be acquired using a number of methods, but throughout this thesis it will be assumed that it simply is ”known”.

8.3 Limitations

The primary travel modes that will be considered is by bus and over ground train. These are widely used (together with the subway- and shuttle trains) to transport the working force in many countries. The type of noise in the vehicles must be examined to asses the level of impact it has on music reproduction.

A reasonable estimation/measurement of the attenuation of noise that in-ear, on-ear and circumaural headphones causes, must also be taken into account.

Furthermore, it is imperative to propose a solution that does not expose the listener to harmful sound pressure levels.

Part II

Design and Implementation

How is the system supposed to work? Essential building blocks are defined.

9 System Identification

In this part, the needed elements in the system are defined and design demands identified. In each subsystem the chosen method will be argued and it's implementation explained. Furthermore, to give a better impression of the effect of each subsystem, an example waveform will be processed and the impact of a given system will be visualized. This waveform is more or less arbitrarily chosen and **not** identical to any of the chosen music samples.

A sequential breakdown of the system goes like this:

1. Register the "target" equal loudness level contour. This is very possibly an interpolation between two defined ELLC's.
2. Register the SNR between music and noise in 1/3 octave bands.
3. Gain the music in a given band to around the threshold where it is no longer masked by the noise.
4. If any band band is causing clipping, compression will be applied.
5. Register the actual ELLC, based on the processed signal and the noise.
6. Use equalization to achieve a final tonal balance closer to the "target" curve.

Some of these subsystems have experimental parameter values. Throughout the design it has been the aim to find "adequate" theoretical parameters, but in some cases, this has not been possible. Initial values has been chosen and later, before and perhaps during the pilot test, these values has been fixed for optimum performance.

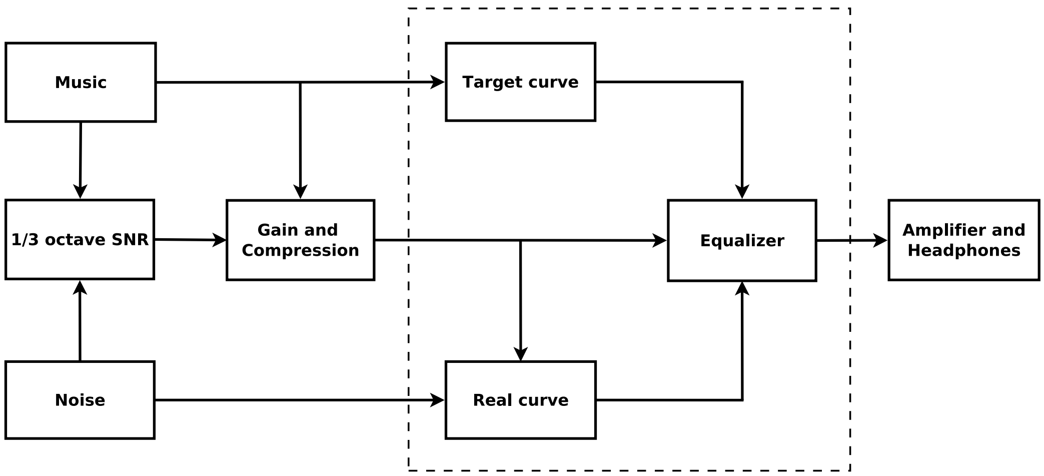


Figure 9.1: A sketch of the system. The minimum system can be implemented without the stippled box.

What does the different parts consist of? Demands are defined and the solution is argued.

10 System Design

A description of the signals used, the subsystems designed and the choices made.

10.1 Signals Used

Some kind of noise needs to be chosen and a range of adequately different pieces of music found.

Music

It is the aim to be able to test the implemented system on different types of music with distinct differences in the frequency spectrum. Based on that, it was decided to find 4 pieces of music with a close resemblance of contemporary pop, classic, pop/rock and Jazz. Based on this, the following sound samples were chosen:

- Madonna - GHV2, track 9, "Frozen" from 1:11 to 1:21. Warner Bros. records 2001.
- Händels Messias, track 24, "Surely, He hath borne our griefs" from 1:16 to 1:26. Danica records 1998.
- Sting - Ten Summoner's Tales, track 8, "It's probably me" from 2:33 to 2:43. A&M records 1995.
- Miles Davis - Some kind of blue, track 1, "So What?" from 1:34 to 1:44. Columbia 1997.

Each track was further cut down to 6 seconds length. It was deemed a reasonable compromise between getting a clear impression of the sound, while not stressing the "acoustical memory" unnecessarily. Next objective was some kind of normalization of the sound level. A level of 70 dBA is loud enough to sound "clear", while still well within safe limits, even under long-term exposure. This value was chosen and the tracks were normalized by first A-weighting the given sample, calculate the RMS sound level and correcting it to 70 dBA.

Noise

The noise signal could be viewed from a purely statistical standpoint. It would be possible to noise shape white noise to a given set of characteristics. Instead, it was decided to use actual noise from a relevant environment, so the later testing could be done using a realistic sound. A 7 minute sample of cabin noise from the passengers' compartment was recorded as in Appendix A.

A 6 second sample was arbitrarily chosen in an interval that sounded homogeneous. By doing this, the noise can be considered quasi-stationary.

To get a SNR that would surely require correction, the noise was gain corrected to 70 dBA using the same method as for the music samples. This value also corresponds to the maximum recommended value for in-vehicle transportation noise of 72 dBA [Mil].

10.2 S/N Measurement

The S/N measurement and the later small band gaining is basically a "brute force" approach when a signal needs to be amplified to a level, where detection is possible.

Some demands must be met:

1. The bandwidth of the filters used for S/N measurement should be practically implementable.

2. The filters must have a reasonable resemblance to the critical bands, to emulate the leakage in the auditory system.

3. The S/N must be calculated in suitable timeframes.

Bandwidth

Several possibilities are present. Of the perceptual scales and consecutive bandwidths, two seems to come to mind. The `Mel` scale, suggested by Stevens, Volkman and Newman in 1937 [SSSN37], is a scale representing the perceptual "distance" between two pitches. Even though this can perhaps be used to asses tonal regions or ranges of interest, no specific bandwidths or center frequencies are suggested.

Another candidate would be the Bark scale suggested by Zwicker in 1961 [Zwi61]. It consist of 24 frequency bands, corresponding to the critical bands of hearing from 20 Hz up to around 15 kHz. The cut off frequencies are 20, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000 and 15500 Hz. As is evident, the relative bandwidth is inconsistent.

Finally, a relatively simple 1/3 octave band solution could be implemented. This solution is used in ISO 532, where some of the lower bands are added together, effectively mimicking the Bark scale. It is this solution that have been chosen, mainly because it is "close enough" to the bark scale and because the filters are easy to generalize and implement.

Leakage

The purpose of the SNR registration is to assess if a music signal is audible in the chosen frequency range. That means that the frequency components in nearby bands can also mask the music, in addition to noise in the same band as the music. According to Zwicker and Fastl: *"Generally used third-octave band filters show a leakage towards neighboring filters of about -20dB. This means that a 70dB, 1-kHz tone produces the following levels at different center frequencies: 10dB at 500 Hz, 30dB at 630Hz, 50dB at 800Hz and 70dB at 1kHz"*, [FZ90, p. 211]. To conform to, what in this context is referred to as "generally used third-octave band filters", a roll off corresponding to a 4th order filter is adequate. As these filters are of the bandpass type, effectively 8 poles are needed. A filter order of 8 has therefore been chosen.

Filter Type

For the filter banks, Butterworth filters are used. They have a harder roll-off than Bessel types, and the constant group delay of the Bessel type is without relevance in this application. Chebychev and elliptic filters features a more aggressive roll-of at the expense of linear amplitude and phase characteristics, so this option has also been omitted.

Time Resolution

The SNR will be updated once a second. According to Moore and Glasberg: *"...it is generally agreed that, for a fixed intensity, loudness increases with increasing duration up to 100-200 ms, and then remains roughly constant"*, [MG02, p.333]. But music is most often dynamic by nature, so such a simple time division is not plausible. It is also not desirable to process the gain settings too fast, as this can result in audible "pumping" or "breathing" artifacts. For this reason 1 second has been the initial value chosen. During the pilot test, this value has performed reasonable.

Processing

The signal to noise ratio measurement can be broken into two 1/3 octave filter banks, followed by rather simple algebra. signal to noise ratio is defined as

$$\mathrm{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \left(\frac{A_{\text{signal}}}{A_{\text{noise}}}\right)^2$$

where

SNR is the signal to noise ratio
P_{signal} is the signal power in RMS
P_{noise} is the noise power in RMS
A_{signal} is the signal amplitude in RMS
A_{noise} is the noise amplitude in RMS

To get a measure of power distribution within each time frame, the crest factor, $\frac{A_{\text{peak}}}{A_{\text{RMS}}}$, for each band will also be detected.

The function is called "SNR" and has been implemented in MatLab. The signal flow can be seen on Figure 10.1.

In this example, two input signals, "music" and "noise" are used. Both are 3 seconds long samples at 48 kHz sample rate. In Table 10.1 the results are presented for each of the bands. It should be noted that these values are representing 3 seconds of sound instead of 1, as used in the implementation.

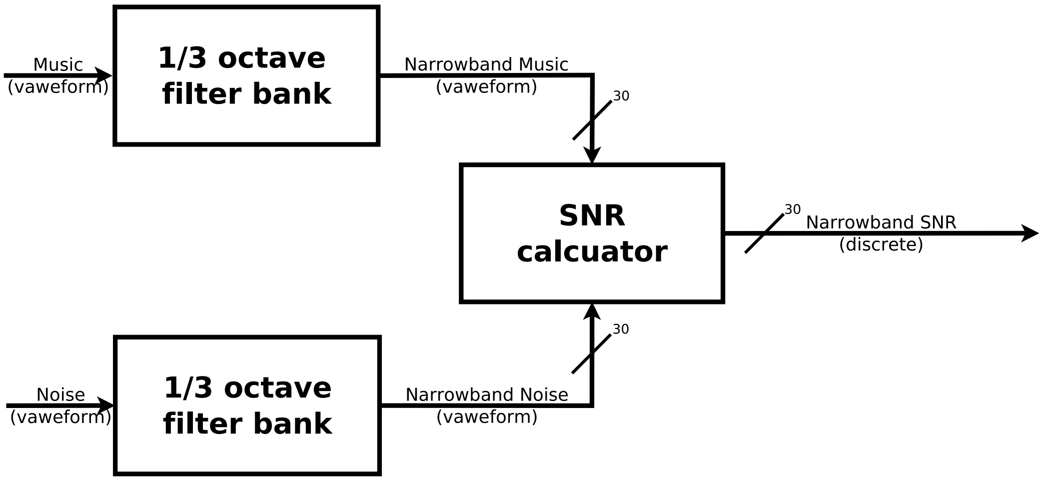


Figure 10.1: Signal flow for the SNR detection

This is to better represent the data. The filter numbering used in Table 10.1 will be used throughout the design documentation, as this is easier to present graphically, than the actual frequencies.

Filter	f_c [Hz]	Music [SPL]	Noise [SPL]	SNR [dB]	Crest [dB]
1	25	36	82	-46	13
2	32	43	81	-38	12
3	40	57	75	-19	13
4	50	63	74	-12	14
5	63	63	78	-15	15
6	79	67	78	-12	15
7	100	64	78	-13	17
8	125	64	83	-19	15
9	158	71	72	-1	11
10	200	66	70	-4	14
11	251	63	67	15	-3
12	316	62	63	12	-1
13	398	55	60	15	-5
14	501	54	58	14	-4
15	631	53	54	19	-1
16	794	54	53	21	1
17	1000	52	51	18	1
18	1585	60	49	20	7
19	1259	57	49	17	11
20	1995	63	50	20	13
21	2512	61	46	15	21
22	3162	57	44	13	25
23	3981	55	42	13	24
24	5012	54	41	13	31
25	6310	51	39	11	34
26	7943	46	38	8	31
27	10k	45	37	7	29
28	12.6k	47	37	11	22
29	15.9k	46	38	8	23
30	20k	37	36	1	27

Table 10.1: Output of the SNR routine.

10.3 Gain and Compression

This stage in the signal processing will apply gain to selected frequency bands. As the noise in question is of low frequency, the gaining will most probably only be needed from a couple of hundred hertz and down. The main goal of this processing block is to bring a masked frequency band from inaudible to audible. There is, however, a couple of important issues to consider.

Crest Factor and Threshold

If the frequency band under examination exhibits a very high crest factor, it is possible to detect a rather low amount of signal power while the important part of the waveform is easily audible. This could be the case when percussive instruments are dominating, e.g. a snare drum. If the SNR is then viewed purely from an RMS perspective, the peak power of the waveform will dominate that particular time frame and actually achieve the opposite of a "natural" reproduction. for this reason, the gain for a given band will be calculated as:

$$\text{gain}_{\text{frequencyband}} = -(CF_{\text{frequencyband}} + \text{SNR}_{\text{frequencyband}}) + \text{Threshold}$$

where:

$\text{gain}_{\text{frequencyband}}$ is the gain in the specific frequency band		[dB]
$CF_{\text{frequencyband}}$ is the crest factor in the specific frequency band		[dB]
$\text{SNR}_{\text{frequencyband}}$ is the noise SNR in the specific frequency band		[dB]
Threshold is the minimum SNR where a signal is audible		[dB]

For the pilot test music different from the experiment samples was used, and this gain scheme has performed reasonably.

Moore states that: *"for medium frequencies, the criterion amount [for avoiding masking effects] corresponds to a signal-to-noise ratio of about ... -4 dB's"*, [Moo03, p. 86]. This SNR Moore is referring to, is within a critical band and not a 1/3 octave band. Furthermore, the frequencies in question is lower than "medium". Nonetheless, this value was chosen as an experimental threshold value, since it makes a reasonable starting point for later adjustments.

Example:

In the 100 Hz band, SNR is -20 dB and the crest factor is 12 dB. The gain for that region will then be:

$$\text{Gain} = -(12 - 20) - 4 = 4[\text{dB}]$$

Obviously, if the SNR is better than the Threshold value, no gain is applied.

Maximum Gain

A problem may occur if no music signal is present in a given frequency band. The result would then be an aggressive amplification of the ”empty” band, resulting in added noise. For this reason, a given band will never be amplified by more than 12 dB’s. This have not posed a problem in the initial testing phase, but music can have a huge diversity, so it is possible that a scheme must be made, to further mitigate the problem.

A reasonable approach would be to make a decision based on the ratio of the music energy in the band in question vs. the total music energy. Another approach could be perceptual coding, thereby taking computational advantage of the human perception. If the system where to run on a platform using the MPEG2 layer 3 lossy compression, there would obviously be no need for processing the omitted information.

Windowing

Another issue is the temporal transition from one filter to the next. If a given filterbank is gained or attenuated aggressively from one timeframe to the next, discontinuities on the waveform can be generated. To avoid this, a rather simple fade in/out of the filters is performed. This gain function can be seen on Figure 10.2.

In its essence, this is a windowing function. Hann, Hamming, Blackman, Tukey and others could all be used, but the triangular, or Bartlett, is the one with least computational complexity, bar the rectangular, and so is chosen.

This block will output the altered waveform.

Figure 10.3 shows the 1/3 octave band energy in the original and gained signal. It should be noted that the small gain increase in the middle and high frequencies is due to leak between filter banks.

Compression and Limiting

Since the signal is possibly gained, there is a risk of clipping. This can be avoided by applying compression or limiting. Limiting is a special case of compression, where the signals envelope is forced within a given dynamic range. This is done using a ”soft knee”. If the ”hard knee” limiting is used, the artifacts would closely resemble those from (hard or digital) clipping. If compression is to be used, peak detection should be used instead of RMS

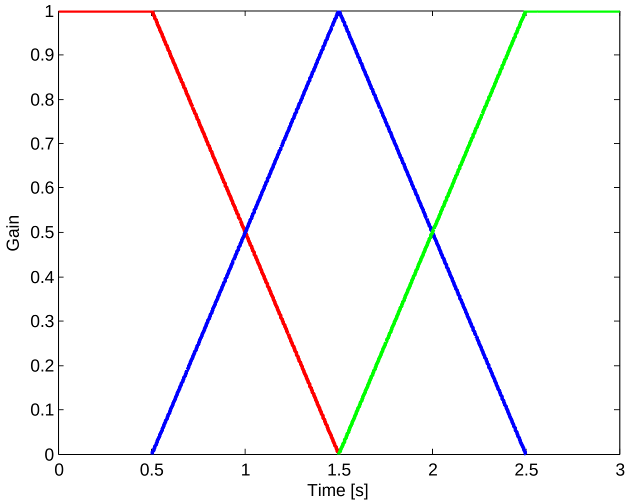


Figure 10.2: The first time frame will use the red gain function. Consecutive time frames will use blue gain function and the last time frame will use the green gain function.

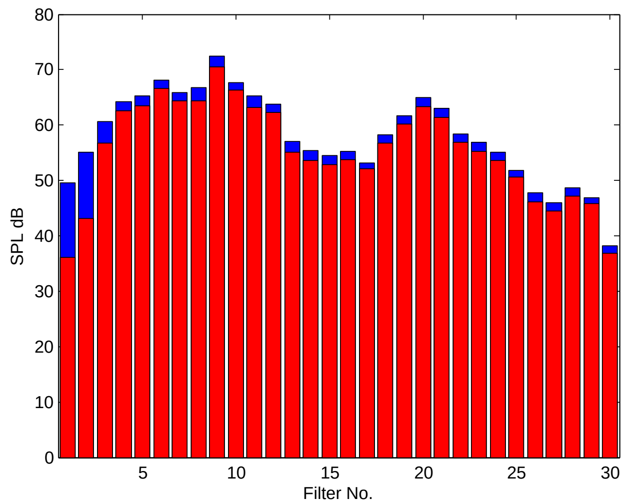


Figure 10.3: The red bars represents the SPL before gaining the blue shows the actual gain.

detection. In the music industry, compression is a tool for shaping and possibly "coloring" the sound. In this case, a transparent sound is the target. Small band compression is recommended. It gives superior control over the dynamics of each specific band, and if broadband compression where to be used, energy levels from other bands can very easily have an unpredictable effect on the target area. On the other hand, it is impossible to predict the total envelope size, when simultaneously compressing multiple frequency bands, so a soft limiting of the whole waveform is possibly needed.

Since the DT990 Pro headphones used can output tremendous sound levels [Bey], there is more than ample headroom for gaining without risk of clipping. The sensitivity is 96 dB SPL @ 1 mW input, and with a power handling capability of 100 mW, a theoretical (and dangerous) sound level of 136 dB is achievable. Therefore compression and limiting will be omitted in this implementation.

10.4 Target Function

The overall function of "target" and "real" curve, is to asses what hearing curve is most applicable. "Target" predicts the tonal balance of the music, based on the a loudness model and interpolation of the equal loudness level contours in ISO 226:2003. It is this target curve that the overall system tries to achieve. "Real" predicts what the actual tonal balance is, including the noise.

Loudness Assessment

There are a lot of options available. By far, the bulk of these methods are developed using single or complex tone stationary signals. One big flaw, however, is that very few are actually tested on dynamic signals, making it hard to choose a suitable method. The method describes in [MG02] have been designed and used, with reasonable results, to predict loudness of time-varying waveforms. This model takes both the temporal and amplitude characteristics into account and as such is considered an open-and-shut candidate for loudness assessment. Figure 10.4 depicts a breakdown of the model.

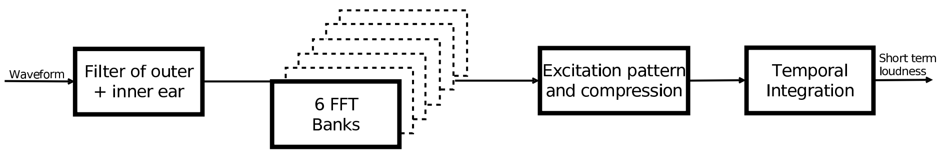


Figure 10.4: Moore and Glasbergs model for assessing short-time loudness.

The first filter is a FIR representation of the transfer function of the outer and middle ear. It is quite similar to the loudness contours in shape, emphasizing the frequencies between 1 and 3 kHz. The signal is then processed using 6 different FFT's. These run at different time resolutions, to get a quick response at high frequency, while maintaining a reasonable frequency resolution at the lower frequencies. Table 10.2 shows the different time and frequency subdivision.

FFT No.	Time Resolution	Frequency Area
1	2 ms	4050 - 15000 Hz
2	4 ms	2540 - 4050 Hz
3	8 ms	1250 - 2540 Hz
4	16 ms	500 - 1250 Hz
5	32 ms	80 - 500 Hz
6	64 ms	20 - 80 Hz

Table 10.2: Data for the FFT's

The next step is an approximation of the excitation pattern on the Basilar Membrane, including a compression similar to that of the auditory system. Lastly, the "instantaneous" loudness is calculated by temporal integration. This instantaneous loudness, S_n is more or less an empty expression, as Moore and Glassberg refers that it: *"is an intervening variable which is not available for conscious perception"*[MG02]. It is, however, the base for calculation the "Short-Term Loudness", $S_n^{'}$. The time frame is denoted n, meaning that "n - 1" is the earlier time frame. $S_n^{'}$ is calculated based on $S_{n-1}^{'}$ and S_n as follows:

$$S_n > S_{n-1}^{'} :$$

$$S_n^{'} = \alpha_a S_n + (1 - \alpha_a) S_{n-1}^{'}, \qquad \alpha_a = 1 - e^{-T_i/T_a}$$

$$S_n \leq S_{n-1}^{'} :$$

$$S_n^{'} = \alpha_r S_n + (1 - \alpha_r) S_{n-1}^{'}, \qquad \alpha_r = 1 - e^{-T_i/T_r}$$

T_a and T_r represents an attack time and release time with values 0.045 and 0.02, respectively. According to [MG02], these values have been found by experiment. This is indeed short term loudness assessment, as T_i is 1 ms.

The longterm loudness assessment, $S_n^{''}$ is calculated somewhat identical as:

$$S_n^{'} > S_{n-1}^{''}:$$
$$S_n^{''} = \alpha_{al} S_n^{'} + (1 - \alpha_{al}) S_{n-1}^{''}, \qquad \alpha_{al} = 1 - e^{-T_i/T_a}$$

$$S_n^{'} \leq S_{n-1}^{''}:$$
$$S_n^{''} = \alpha_{rl} S_n^{'} + (1 - \alpha_{rl}) S_{n-1}^{''}, \qquad \alpha_{rl} = 1 - e^{-T_i/T_r}$$

The values of α_{al} and α_{rl} is set to 0.01 and 0.0005, respectively.

As seen on the α values, the perceived loudness rises very quickly but recedes a lot slower. This, in effect, is somewhat identical to the "temporal masking" phenomenon.

Execution

The implemented code is helpfully available from Glasberg's homepage on Cambridge university's web page[1] and have been executed from MatLab. The function is executed by means of the *dos prompt emulator* available in MatLab. The input arguments is:

- A waveform sampled at 32 kHz. A resampling program is available from the web page.
- A calibration value, stating the SPL represented by full swing. Since the waveform format only allows a signal of ± 1, the signal is attenuated 20 dB and the calibration set to 111 dB's, the original 91 dB added to the 20 dB's attenuation.
- A filter to take into account headphone representation. There are filter available for both diffuse field and free field equalized headphones. DT990 Pro is of the diffuse field type [Bey] and so, the corresponding filter is used.

The program outputs a text file containing both the $S_n^{'}$ and $S_n^{''}$ values. A small script is used to extract the longterm loudness $S_n^{''}$. It is then evaluated as the time average for each second of music presented. In the case of the test music, results is shown in Table 10.3.

[1] `http://hearing.psychol.cam.ac.uk/Demos/demos.html`

Time [s]	1	2	3
Phons	85.6	89.4	89.5

Table 10.3: Output from the loudness program

For all the music used in the listening test, this model consequently gives higher results than simple A weighting. However, if a pure 1 kHz sine is used as input, The results are identical with almost no variation. As music is often estimated by pink noise, having equal power in all octave bands, it is clear that this model is more sensitive to the lower frequencies, than A weighting.

The Equivalent Loudness Level Contour Used

As the data for the ELLC's are readily available in ISO226:2003 and easily interpolated using spline interpolation, this function is in reality just a linear interpolation between two defined curves seen of Figure 10.5.

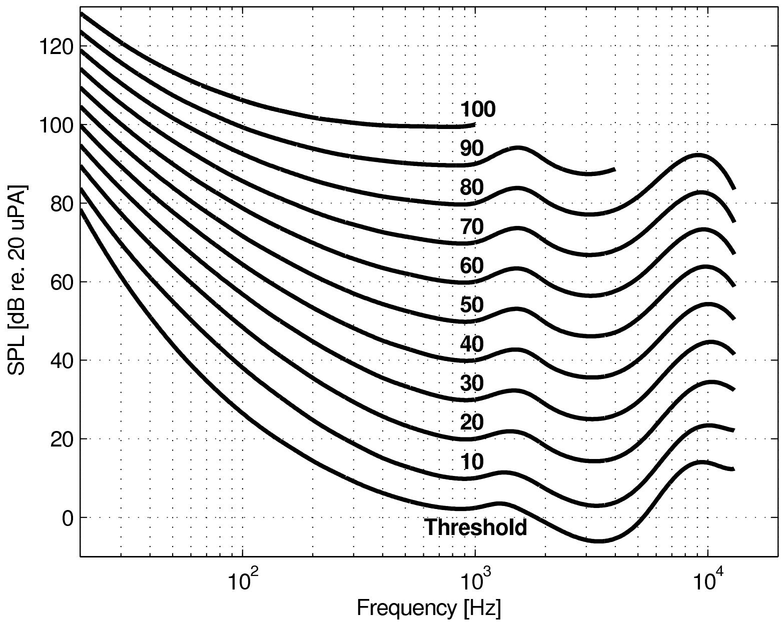


Figure 10.5: Revised equal loudness level contours from ISO 226:2003. Bold values are in Phons.

If, eg. the 63 Phon curve is used, it is calculated as:

$$C_{63}(f) = C_{60}(f) + (C_{70}(f) - C_{60}(f)) \cdot \left(\frac{63 - 60}{10}\right)$$

where

$C_x(f)$ is the ELLC for a reference tone of x Phon [dB]

The new curve depicted in Figure 10.6. It should be noted that the 90 and 100 Phon curves are incomplete. In these areas the earlier "usable" graph will be used, meaning that the 80 Phon curve will be interpolated to give the contour or the 90 Phon curve at frequencies 4 kHz and up.

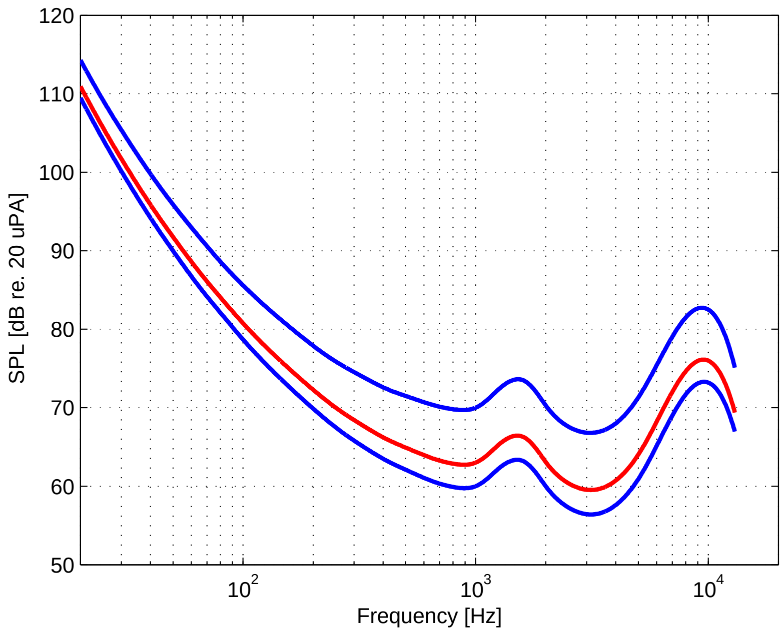


Figure 10.6: The "new" 63 Phon contour is red, 60 and 70 Phon contours are blue.

The output will be a 1000 point ECCL logarithmically spaced.

10.5 Equalization

As in the SNR subsystem, a 1/3 octave filter bank will be used. The filter order for each flank will be 4, for a total of 8 poles in each passband filter. The function takes in the waveform from the gain/compression stage.

The target tonal balance will be applied to frequencies between 400 Hz and 13 kHz. This tonal balance aimed for is the target ELLC subtracted from the real ELLC. Figure 10.7 illustrates this. In this case, the ELLC without noise corresponds to the 60 Phon curve. When the listener is affected by the noise, the ELLC changes to the 70 Phon curve. To negate this tonal shift, the black line can be used to extract the values to equalize with.

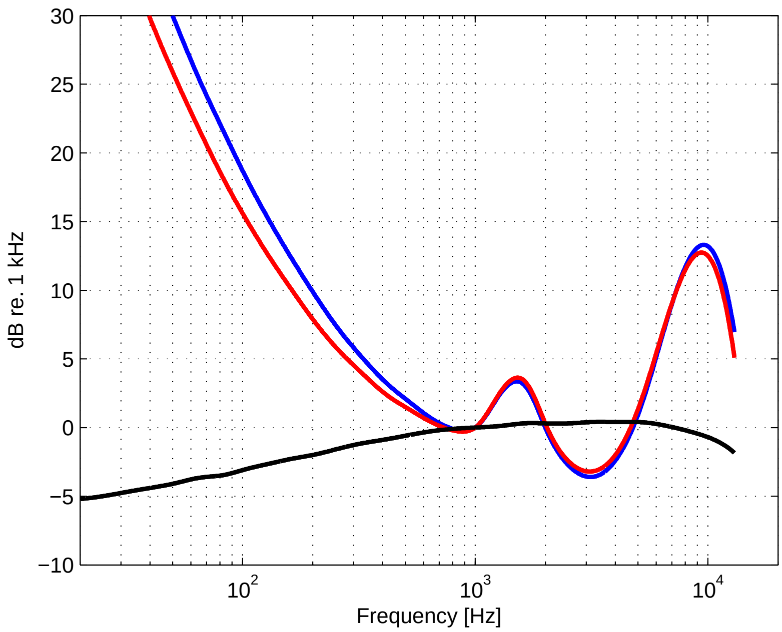


Figure 10.7: The tonal balance aimed for in black. The blue curve is the 60 Phon curve, the red curve is the 70 Phon curve.

As evident from Figure 10.7, an increased total loudness will actually result in a target equalization curve that attenuates the lower frequencies. These are the exact frequencies that are possibly masked by noise. For this reason, the equalization will only be carried out on the middle and high frequencies, namely 400 Hz and up. It is at this frequency, in the 13th filterbank, that the level will be referenced to 0 dB. Figure 10.8 shows the effect of the equalization on the test signal. This equalization is extremely subtle for the test signal used, as there is only a very small change in the ELLC curve, when noise is added. Also, the change will be more profound at low sound levels, as the compression at higher levels will result in a more linear curve, leading to a more flat equalization curve.

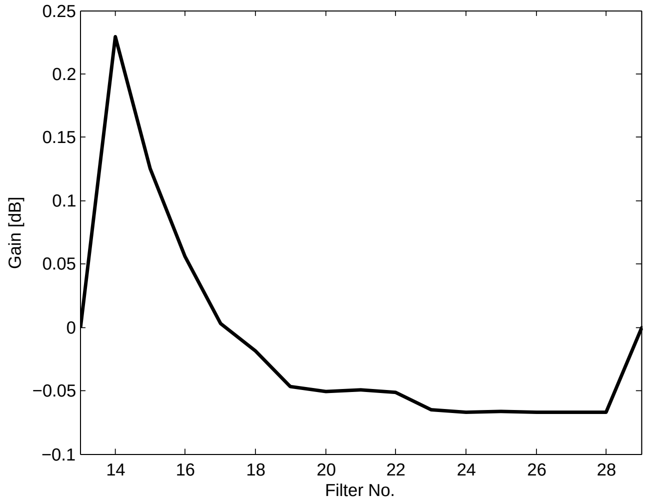


Figure 10.8: The gain in filterbanks 13 to 28.

10.6 Amplifier and Headphones

To be able to present the processed signal to the user, means of amplification for a suitable set of transducers are needed.

Amplifier

Headphone impedance can vary a lot from very low to 600 ohm's or more. It was decided to compare a relatively cheep in ear type and a relatively expensive circum-aural type. The two headphones had very different impedance. The in ear type as low as 15 Ω at DC, while the circum-aural are more than 500 Ω at DC. This impedance can, and probably will, change near resonance. It requires an amplifier capable of driving low impedances. First, a Pioneer power amplifier with fixed gain were used to supply power. To achieve this fixed gain, the speaker outputs had to be used. This was not a usable solution. The noise on the output terminals combined with the very high sensitivity of both types of headphones resulted in excessive noise. Instead an cheap external soundcard were used. The soundcard have a variable headphone output that was modified to full gain setting. This soundcard is used as the amplifier. Testing of the soundcard is found in Appendix B.1.

Headphones

It was decided to match up a couple of headphones of the in ear type with a pair of high quality circum-aural headphones. Test of linearity was made for both and the in ear headphones were further tested for insertion loss. The test was carried our in Appendix B.2 and B.3. Ultimately, the Beyerdynamic DT990 PRO was chosen, since it was deemed that these would fit consistently to the test subject. A low frequency correction was also carried out to achieve a higher degree of linearity, see Appendix B.3.

10.7 Results

The four pieces of music were processed using the system. From the 4 uncorrected samples, 4 samples with the SNR correction and 4 samples with full compensation were produced. To check that the loudness did not increase aggressively an A-wighting was conducted on all tracks. The RMS value for the full 10 second long unprocessed and processed sample were then calculated. The results is shown in Table 10.4.

Style	**No correction**	**SNR correction**	**Full correction**
Pop	70.0 dBA	71.7 dBA	73.6 dBA
Classic	70.0 dBA	71.8 dBA	73.8 dBA
Pop/Rock	70.0 dBA	71.4 dBA	73.1 dBA
Jazz	70.0 dBA	71.6 dBA	73.5 dBA

Table 10.4: The A-weighted results of music SPL from the processing.

As can be seen, the SNR correction roughly raises the SPL 2 dB's, while the full correction raises the SPL somewhere around 3 dB's. This is off course nor telling anything about the perceived loudness, as the increase in SPL primarily is in more silent passages and not overall.

To get an idea of the increase in total loudness, Table 10.5 shows the A-weighted results of the noise super positioned on the unprocessed and processed samples. Again, these values are just for reference, they do not necessarily represent the perceived loudness.

Style	No correction	SNR correction	Full correction
Pop	73.0 dBA	74.0 dBA	75.2 dBA
Classic	73.0 dBA	74.0 dBA	75.3 dBA
Pop/Rock	73.0 dBA	73.8 dBA	74.8 dBA
Jazz	73.0 dBA	73.9 dBA	75.1 dBA

Table 10.5: The A-weighted results of music and noise SPL from the processing.

Part III

Listening Experiment

To examine whether the proposed solution provide improvements to the sound experience for the listener, a listening experiment should be conducted. In the following, the motivation for the listening test will be described.

11 Motivation

As stated in the problem description, Chapter 8, the overall goal is to make it possible to reproduce music in a noisy transit environment with the same perceived tonal balance as without any noise. In order to examine more thorough the effects of the proposed method, as described in Part II, a listening experiment is conducted.

The shifting playback environment for portable music devices introduce various possible degradations of the listening experience. By applying the noise compensating method, described in the previous chapters, it is of interest to see whether applying the method improves the listening experience when the playback environment contains transit noise. In this thesis, the focus will be on two characteristics that seems of particular relevance to playback in a noisy environment; the naturalness of the sound and the annoyance of the noise on the listening experience. If the perceived annoyance of the noise is not decreased by applying the proposed methods, the methods are not useable in the proposed form and setup. On the other hand, if the processing the sound signal introduces artificial attributes this might degrade the listening experience unacceptably. Therefore, when examining the effects of the method through a listening experience, following questions is raised:

Naturalness: To which extend is the naturalness of the sound maintained, despite the background noise and the alterations made?

Annoyance: To which extend is the annoyance of the background noise decreased for the proposed solution?

These issues motivates a two-part listening experiment. In the following chapter the procedure, method and setup will be described.

The subjects task will be evaluation of both naturalness and annoyance. A clarification of the different tasks is of importance. *Naturalness* is defined as the one of the two sounds most like the reference, a noise free sample. *Annoyance* is defined as where the music experience is degraded the most by the noise.

What choices were made? What are the potential sources of error and how can they be minimized? The perceived naturalness and annouyance were measured.

12 Experiment Procedure

The listening experiment took place at Aalborg University, listening cabinet B, Frederik Bajers vej 7, B5, 9220 Aalborg Ø from May 14th to May 17th 2010. The experiment setup is described in Appendix E. The subjects were asked to fill out a questionnaire, as found in Appendix D[1]. It was chosen to compare the unprocessed signal and noise to a bass compensated and a full compensated method. There are four cases of sound stimuli:

- *Original signal.* This is used as reference for evaluation of naturalness.
- *Signal with noise.* The unprocessed signal and noise.
- *Signal with noise, bass compensated model.* This simplified model consists of 1/3 octave band gaining to compensate for the masking, but does not take the loudness model into account
- *Signal with noise, fully compensated model.* This model is the full scale processing method, which compensates for the masking, without raising the overall loudness level considerably[2].

In order to eliminate the possible influence of personal preferences on music genres in the responses from the subjects, it was decided to examine four different music samples of different genres.

12.1 Choice of Experiment Method

The naturalness and the annoyance will be examined individually in two consecutive parts.

[1]Results from the questionnaire can be found on the attached CD.

[2]Figure 9.1, page 27 shows the devision of the proposed method into the bass compensated (without the stippled box), and the full system (with the stippled box).

12.1.1 Naturalness Experiment

For the naturalness experiment, the subject should examine how the specific characteristics, that may be altered by influence of the noise, and possibly by the processing, influence the listening experience.

Since it is of interest to examine the naturalness in the unprocessed and processed noisy scenarios, the original and noise free sample should be presented as a reference.

This evaluation could be done either by asking the subject to rate signals related to each other on a scale from e.g. best to worst. In order to do so, the subjects would need to listen to the samples more than once, and even with a thorough subdivision of the rating scale, the task might be difficult for the subjects. A simpler task for the subjects would be to compare just two samples to each other and pick one of them; the paired comparison method. A drawback of this method is the greater number of subjects needed; at least six times the number of stimuli compared. It was chosen to use the paired comparison method due to its simplicity in task for the subjects. It was also chosen to allow the subjects to listen to the samples and the reference as many times as they wanted before stating their choice.

Hypothesis

The hypothesis examined for the naturalness experiment is:

There are no change in the naturalness of the sound relative to the noise free sound sample.

If this hypothesis can be rejected, further investigation can be done as to whether the simple bass compensation or the full compensation method provides an improvement of the impression of the sound in noisy environments.

The graphical user interface displayed to the subjects can be viewed in Figure 12.2.

The subjects will be presented to one single judge[3], i.e. three pairs, for each kind of music.

[3]The concept *single judge* describes a complete judgment of $\binom{t}{2}$ pairs of t stimuli, according to the notation used by *H. A. David* [Dav88].

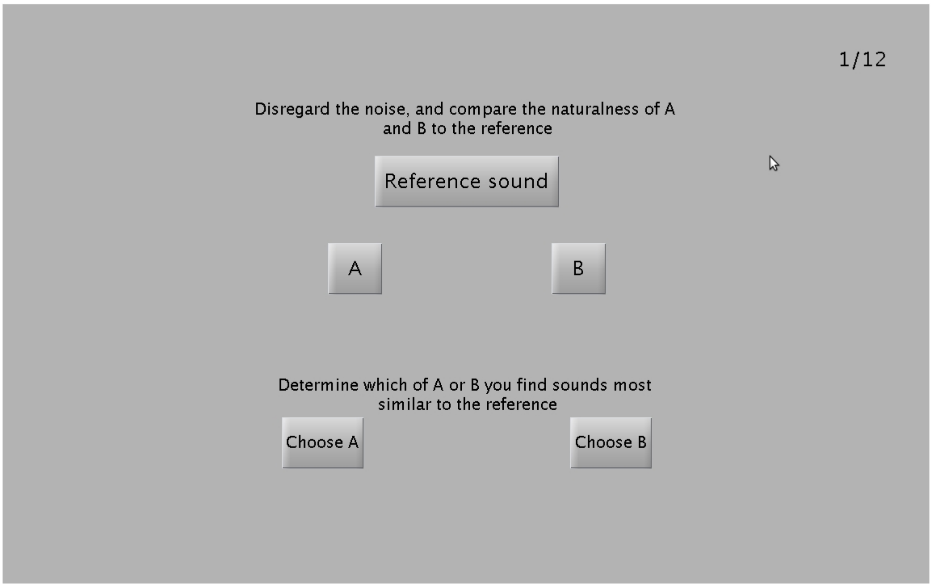


Figure 12.1: Graphical user interface for part 1 of the experiment.

12.1.2 Annoyance Experiment

The annoyance part of the experiment examines the degradation of sound quality due to the simultaneous present noise. It is of interest to examine whether application of the proposed method will decrease the annoyance of the noise on the listening experience.

Again, the unprocessed and processed noisy samples should be compared. By examining the annoyance of noise, it is not relevant to compare the samples to a reference. Instead each sample could be rated on e.g. a 5 or 7 scale from, ranking the noise from "not annoying at all" to "very annoying". A drawback of the rating scale methods are that people tend to avoid the endpoints, which leaves only limited room for different perceptual scale values. The paired comparison method also seems yet again suitable, due to its simplicity for the subjects, and the possibility of obtaining at least a straight ranking from the results obtained from the subjects. It was chosen also to use the paired comparison method for the annoyance experiment.

Hypothesis

The hypothesis examined for the annoyance experiment is:

No method will decrease the annoyance of the noisy background better than *others.*

If this hypothesis can be rejected, further investigation can be done to examine if the bass compensation or the full compensation method provides an improvement, i.e. a more optimal listening experience.

Though the statistical method is the same for both parts of the listening experiment, the task differs slightly. In this experiment the graphical user interface changes to the one displayed in Figure 12.2.

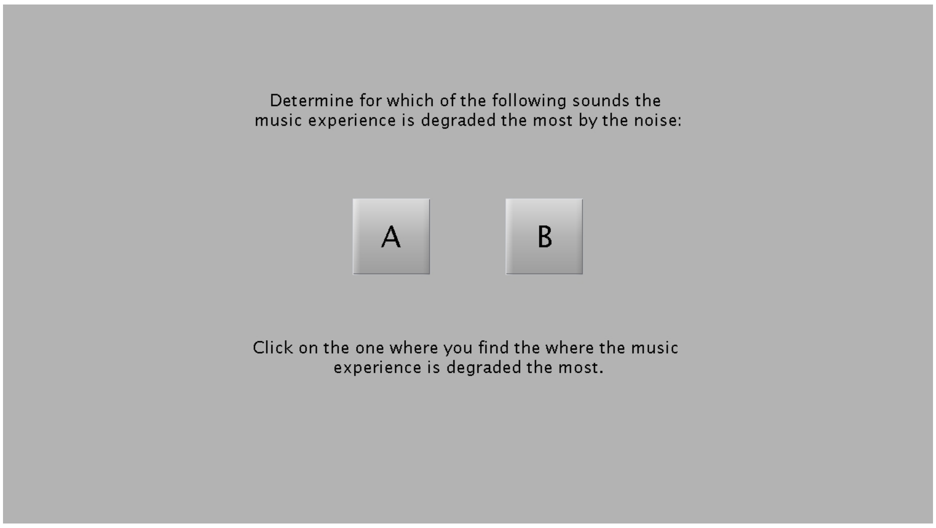


Figure 12.2: Graphical user interface for part 2 of the experiment.

In this part, the subject will be presented to two single judges, i.e. 6 pairs for each kind of music. It was chosen only to present the subjects to the stimuli once for each pair of sounds, and then ask them to make a choice based on the presented stimuli.

12.1.3 Population of Subjects

For the paired comparison experiments it is recommended to use a number of subjects at least six times greater than the number of stimuli [ZE03]. Since this method is applied for both experiment parts, and three stimuli are compared, a minimum of 18 subjects is required.

It was chosen to use untrained subjects, since it is of significance that the possible users should be able to detect the difference. Another issue is time constrains, as training subjects requires more time.

12.2 Statistics on Paired Comparison

It was chosen to do a paired comparison test, comparing the unaltered signal and noise to a bass compensated and fully compensated version, as described in the proposed method. The paired comparison method seemed suitable for evaluating whether the proposed methods had a significant effect both on the naturalness and on the perceived annoyance by the subjects.

By paired comparison the subjects are asked to choose the one of two. It is a forced choice, and in order to determine whether the subjects are actually able to establish a logic preference order or not, both the order of the sounds presented and repetitions must be considered. By paired comparison, one obtains indirectly scaled responses, which can be used to establish a ranking of the stimuli, which can not be established by objective measurements.

In order to examine the data, following steps must be taken [ZE03]:

- Individual consistency check, to examine whether or not the individual subject is consistent in his or her responses.
- Pooled Data consistency check, to examine whether or not the subjects as a group are consistent.

And if the data are consistent:

- Estimation of parameters and model fit, to determine the model most suitable for description of behavior.
- Assignment of ratio scale values.

12.2.1 Individual Consistency Check

The subject was asked to compare $\binom{t}{2}$ pairs to complete a single judge. For the naturalness experiment one single judge is made, and for the annoyance experiment two single judges. Since the subjects are able to hear the stimuli as many times, as needed in the naturalness experiment, it was chosen not to repeat the experiment. In the annoyance experiment the subjects were presented to the stimuli only once before they were asked to make a choice. Since this task was estimated harder for the subjects, it was chosen to do two presentations of the same pair of stimuli. The number t is the total number of stimuli, which is chosen to 3. A typical way of representing the data from a single judge is shown in Table 12.1. A "1" in the matrix represents that the stimulus in the row is preferred over the stimulus in the column. In this example stimulus 2 is preferred over 1 and 3, stimulus 1 is preferred over 3 and stimulus 3 is not preferred over any of the two others.

Stimulus	1	2	3	scores α_i
1	0	0	1	1
2	1	0	1	2
3	0	0	0	0

Table 12.1: Example of a single judge with 3 stimuli.

The null hypothesis is evaluated by examining whether the single judge(s) are consistent or not. For the subject to be able to answer consistently, it is necessary that enough differences exist between the stimuli. The results from the individual subject should be examined for inconsistencies, namely circular triads. A circular triad is when stimuli A is preferred over B, B is preferred over C but C is preferred over A.

For Individual consistency check Zimmer and Ellermeier [ZE03] refers to a method proposed by Kendall, where at least 8 stimuli for comparison is needed in order for the method to be valid. Since only three stimuli are examined in this case, another form of individual consistency check must be made.

The individual test subject can have different preferences or it is possible that the subjects are not able to hear any difference between the samples. To determine if the latter is the case, a working hypothesis is needed.

H_0*: The test subjects can not perceive any difference between the samples.*

This implies that they give random answer, meaning that they will prefer method X over Y with $P = 0.5$. Since the method preferred for processing a given signal doesn't necessarily apply to another type of music, it is of interest to see how many subject results of this type are consistent with themselves.

For a subject to be consistent, it is demanded that $\alpha_1 \neq \alpha_2 \neq \alpha_3$ for the naturalness experiment. The only situation where this is not the case is when all α_i are equal to 1. For the annoyance experiment, an arbitrary α must assume the value 4, another 2 and the last α must equal 0.

If consistency for a given subject and musical style is established as a "success" and no consistency a "failure", the consistency can be viewed as a Bernoulli random variable with a chance of success equal to p, as calculated in Appendix C. To estimate the probability of achieving at least the number of consistencies, the number of trials, n, must be found.

The probability of i successes out of n trials is:

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \tag{12.1}$$

12.2.2 Pooled Data Consistency Check

If the individual data sets are sufficiently consistent, the relationship between the different subjects responses should be examined by pooling them to form a cumulative preference matrix. The consistency of this matrix can be examined by evaluating the weak stochastic transitivity (WST). If the WST holds, it is possible to establish an ordering of the stimuli with respect to the perceived naturalness or annoyance, according to the respective experiment..

If for three stimuli: A, B and C, both the probability of preferring A over B is $P_{AB} \geq 0.5$ and the probability of preferring B over C is $P_{BC} \geq 0.5$ then the WST is violated if A is **not** preferred over C, $P_{AC} \geq 0.5$. There are no statistical test available to examine if the WST condition is violated. Instead, the number of times the WST is violated should be compared to the total number of times, where a pair of $P_{AB} \geq 0.5$ and $P_{BC} \geq 0.5$ occurs in the cumulative matrix [ZE03].

If the WST condition is fulfilled, it is possible to establish a preference order, and apply a model. In the case where the WST condition is not fulfilled, the ratio scale values u are derived from the cumulative matrix by straight ranking, ie. counting the number of times one stimulus is preferred over the others.

The cumulative matrix can be represented as:

$$M = \begin{bmatrix} \mathrm{N}_{11} & \cdots & \mathrm{N}_{1t} \\ \vdots & \ddots & \vdots \\ \mathrm{N}_{t1} & \cdots & \mathrm{N}_{tt} \end{bmatrix} \tag{12.2}$$

where N_{ij} is the number of time that the subjects prefer the stimulus in row i over the stimulus in the column j, and t is the number of stimuli ($t = 3$). The total number of scores is:

$$\alpha_i = \sum_{j=1}^{t} N_{ij} \tag{12.3}$$

Since the preferences are given in a ratio scale, the signal and noise, unprocessed, is chosen as a reference value. Thus, the ratio scale values u will be given by:

$$u_i = \frac{\alpha_i}{\alpha_1} \tag{12.4}$$

The standard error in this case is:

$$se_i = u_i \pm 1.96 \frac{\delta(M)}{\sqrt{t}} \tag{12.5}$$

where $\delta(M)$ is the variance of the cumulative matrix.

12.2.3 Estimation of Parameters

If the WST is not violated, a model can be applied to obtain a ratio scale for the compared stimuli. Both the Bradley-Terry-Luce (BTL) model, and one of the generalizations of this, the Pretree model, can be applied [ZE03]. An example of the model structures are shown in Figure 12.3.

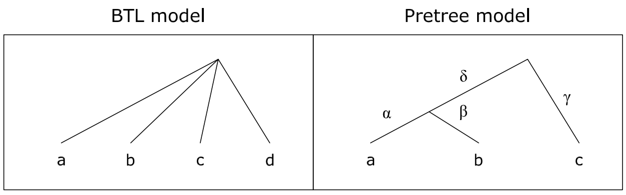


Figure 12.3: Schematic structure of the BTL and Pretree models [ZE03].

For the BTL-model, the probability that stimulus a is preferred over stimulus b relates to the ratio scale values as:

$$P_{ab} = \frac{u(a)}{u(a) + u(b)} \tag{12.6}$$

where $u(a)$ and $u(b)$ are the ratio scale values for stimulus a and b respectively [ZE03].

For the Pretree model, the probability that a is considered more pleasant than b, P_{ab}, relates to the characteristics of the stimuli that influence the subject's decision. Thus, the probability that stimulus a is considered more pleasant than stimulus b is:

$$P_{ab} = \frac{u(a' - b')}{u(a' - b') + u(b' - a')} \tag{12.7}$$

where $u(a' - b')$ and $u(b' - a')$ are the ratio scale values of preferred pleasantness for the stimuli a and b respectively [ZE03]. a' denotes the attributes that the stimulus a does not share with b. In the example shown in Figure 12.3, the relevant attributes are $a' = \{\alpha, \delta\}$, $b' = \{\beta, \delta\}$ and $c' = \{\gamma\}$, therefore P_{ab} becomes:

$$P_{ab} = \frac{u(\{\alpha, \delta\} \,/\, \{\beta, \delta\})}{u(\{\alpha, \delta\} \,/\, \{\beta, \delta\}) + u(\{\beta, \delta\} \,/\, \{\alpha, \delta\})} \tag{12.8}$$

In both cases, the ratio scale values $u(a)$, $u(b)$, $u(a' - b')$ and $u(b' - a')$, are estimated by maximizing the likelihood data, given by:

$$L = \prod_{i<j} P_{ij}^{\alpha_{ij}} (1 - P_{ij})^{N - \alpha_{ij}} \tag{12.9}$$

where α_{ij} is the number of times that the stimulus in the row i is considered more pleasant than the stimulus in the column j and N is the total number of comparison of a pair of stimuli.

To calculate the goodness of fit, the likelihood of the model chosen L_{model} is compared to the likelihood of the unrestricted model L_{sat} which is assumed to have independent binomial distribution [ZE03].

The χ^2 distribution:

$$\chi^2 = -2 \cdot \ln \frac{L_{model}}{L_{sat}} \tag{12.10}$$

with $\frac{t(t-1)}{2} - (t - 1 + c)$ degrees of freedom, is used to test the goodness of model fit. t is the number of stimuli and c is the number of branches of the model. The significance level typically used for the test statistics is 10% [ZE03].

The choice of the model follows the Akaike's Information Criteria (AIC) [WS04]. The AIC is defined as:

$$AIC = -2 \log L_{model} + 2(t - 1 + c) \tag{12.11}$$

The AIC takes into account the likelihood data of the model tested L_{model} and the number of free parameters.The model chosen should be the one that gives the lower AIC.

12.2.4 Assignment of Ratio Scale Values

If the BTL model is applied, the ratio scale values u of the stimuli are equal to the parameter estimates $\hat{u}$ $(u(a), u(b), ...)$ [ZE03].

On the other hand, in the case of the Pretree model, the ratio scale values are given by the sum of the length of the branches of the tree. Thus, in the Pretree structure presented in Figure 12.3, the ratio scale value of the stimulus a is equal to the sum of the length of the branches α and δ.

As stated, the signal and noise, unprocessed, is the reference value. The ratio scale values u for the i stimuli $(i = 1 \ldots t)$ will be given by:

$$u_i = \frac{\hat{u}_i}{\hat{u}_1} \tag{12.12}$$

The standard error of the parameter estimates $\hat{u}$ is obtained as the square root of the diagonal of the covariance matrix [WS04]:

$$se_i = u_i \pm 1.96\sqrt{\mathrm{Diag}\left[\widehat{\mathrm{cov}}(\hat{u}_i)\right]} \tag{12.13}$$

A listening experiment has been conducted, in order to examine whether the proposed solution decreases the level of annoyance related to the background noise, and maintains and impression of naturalness. The data from the listening experiment are examined and discussed in this chapter.

13 Results and Discussion

As stated in the Chapter 11, examination of paired comparison experiment data includes an initial consistency check to determine, whether model specifications and model fit can be calculated. If the responses cannot be considered consistent, only a straight ranking can be determined.

To eliminate the personal preference on types of music, the subjects are presented to four different kinds, representing a broad spectrum of music[1].

It is of interest both to see if the method has influence on the preferences of the subjects for specific music types, or if one method is preferred significantly for all kinds of music.

13.1 Examination of Naturalness

The first part of the listening experiment was designed to examine whether the noise or the application of the proposed method for optimizing the listening experience in noisy environments introduced some crucial attributes to the original signal. The subjects were asked to compare the original noise free signal, denoted as the reference, to two noisy samples, and determine whether the one or the other sounded most like the original[2].

The subjects are asked to compare three pairs, and are able to hear the samples as many times as they like.

[1]The sound samples are of a length of 6 seconds, and can be found on the attached CD.

[2]The MatLab script used for this part of the listening experiment can be found on the attached CD: *test_Gui_part1.m*

13.1.1 Consistency Check

An evaluation of the consistency of the data obtained from the subjects must be made in order to determine whether a ranking model can be applied. Even though this is not the case, one of the advantages of the paired comparison method is, as stated previously, that a straight ranking can always be obtained from the results. In order to evaluate the consistency, both the individual data and the pooled data is examined.

Individual Check

The individual check was done by examining if the responses for a single judge constituted a circular triad, meaning that A was preferred over B, B over C and C over A, as such as the response shown in Table 13.1. Alternatively, the responses are consistent as the response shown in Table 13.2. Since the single judge consists of only 3 samples, only one circular triad is possible pr. subject, meaning that either the subject is fully consistent or fully inconsistent.

Stimulus	1	2	3	α_i
1	0	1	0	1
2	0	0	1	1
3	1	0	0	1

Table 13.1: Example of an inconsistent single judge with 3 stimuli. (subject 2, sound sample 1)

Stimulus	1	2	3	α_i
1	0	0	1	1
2	1	0	1	2
3	0	0	0	0

Table 13.2: Example of a consistent single judge with 3 stimuli. (subject 2, sound sample 4)

The responses for the naturalness part of the experiment is shown in Table 13.3.

Style	Consistent single judges	Inconsistent single judges
Pop music	14	4
Classic	13	5
Pop/Rock	13	5
Jazz	14	4

Table 13.3: Responses from the naturalness experiment.

By applying the method described in Section 12.2.1 and Appendix C, the probability that a subject is consistent for a single music style is:

$$P(\text{consistent}) = 0.875 \tag{13.1}$$

If consistency for a given subject and musical style is established as a "success" and no consistency a "failure", the consistency can be viewed as a Bernoulli random variable with a chance of success equal to p = 0.875. To estimate the probability of the number of successes or more, the number of trials, n, must be found.

$$\begin{aligned} n =& \text{No. of subjects} \cdot \text{No. of music styles} \\ n =& 18 \cdot 4 \\ n =& 72 \end{aligned} \tag{13.2}$$

In our case, we have 54 successes out of 72 possible. The chance that we should achieve this result, or better is calculated from Equation 12.1, page 56:

$$\begin{aligned} P(X \geq 54) =& P(X = 54) + P(X = 55) + ... + P(X = 72) \\ P(X \geq 54) =& 0.9989 \end{aligned} \tag{13.3}$$

Based on this, we cannot reject the working hypothesis. A rejection of the hypothesis is necessary to conduct further analysis. A number of successes equal to 68 is needed to be comparable to a confidence interval of 95%. 68 successes only occurs with probability 0.0445.

Pooled Check

Not only the individual consistency, but also the consistency between the subjects, ie. the consistency of the cumulative matrices, are relevant in order to determine whether the hypothesis can rejected or not.

To evaluate the pooled data, a Weak Stochastic Transitivity(WST) check, as described in Section 12.2.2, has been applied. If the WST is violated, the data should be viewed as inconsistent. There are no specific measure as to when the WST is violated crucially, but a count of the violations can be used to determine whether the responses can be considered consistent or not [ZE03]. The results from the WST check can be viewed in Table 13.4.

Style	Violations
Pop music	2 of 6 possible combinations
Classic	4 of 9 possible combinations
Pop/Rock	4 of 9 possible combinations
Jazz	4 of 9 possible combinations

Table 13.4: Weak Stochastic Transitivity check for the cumulative matrix from naturalness experiment.

Since there for all four sounds are violations, the responses must be considered inconsistent, hence it is not possible to extract model specifications or fit the data into a model.

13.1.2 Discussion

The consistency check revealed that the responses were inconsistent, and the hypothesis therefore cannot be rejected. It is not possible to rank the stimuli in any other way than a straight ranking. There might be many reasons as to why consistency is not present in the data. The stimuli might not differ significantly enough for the subjects to either hear the difference, or, if the difference is noticeable, not determine a logic preference order among the three stimuli. Another reason could be that the subjects are changing their minds throughout the experiment due to some sort of learning effect. The learning effect has been minimized by randomizing the order of the stimuli.

Even though 7 subjects[3] are consistent for all four kinds of music presented, they do not have the same preference order for the different kinds of music.

[3]The 7 fully consistent subjects are subject 3, 4, 6, 7, 8, 12 and 18.

This could be because of their normal preference on music types, but it could also be because of the characteristics of the specific type of music.

Another point is that though they are able to perceive differences between the samples, they are not able to establish a logic preference order.

13.1.3 Straight Ranking Results

Since the hypothesis of randomness cannot be rejected, and the subjects cannot be considered consistent, only a straight ranking of the methods can be obtained. Table 13.5 contains the straight ranking results. It appears that the full compensation method are chosen as the one giving the most natural replica of the sound signal.

Style	Most natural		Least natural
Pop music	Full comp.	Bass comp.	No comp.
Classic	Full comp.	No comp.	Bass comp.
Pop/Rock	Full comp.	Bass comp.	No comp.
Jazz	No comp.	Full comp.	Bass comp.

Table 13.5: Straight ranking results for the naturalness experiment. Best means the one closest to the original, and worst the one furthest away from the original.

13.2 Examination of Annoyance

In the second part of the listening experiment the noise annoyance was examined for the unprocessed, bass - and fully compensated cases for four different kinds of music. As the naturalness experiment, the annoyance experiment also was conducted as a paired comparison experiment, however this time no reference sound was presented, and the sound pairs were only presented one time pr. choice. The experiment consisted of two single judges, randomized to eliminate any learning effect. The subjects were asked to pick the one of the two presented sounds, where the noise disturbed the music experience the most.

13.2.1 Consistency Check

Since the method for the naturalness experiment is the same as for the annoyance experiment, the same steps must be taken in order to evaluate the consistency of the data; both the individual data and the pooled data are examined, to examine whether the results are consistent enough for a model fit to be applied.

Individual Check

The individual check was done as for the naturalness experiment. The data this time consists of a three times three matrix with results from 2 single judges. As described in Appendix C, the probability that a given subject will make consistent choices for a given music style, P(c), is approx. 0.1. Since there are 13 consistent cases of 72 possible, it follows that the chances of this happening by pure randomness is:

$$\begin{aligned} P(X \geq 13) =& P(X = 13) + P(X = 14) + ... + P(X = 72) \\ P(X \geq 54) =& 0.0157 \end{aligned} \tag{13.4}$$

Based on this number, it is assumed that the subjects are capable of distinguishing the different processing methods in this part of the experiment. It is, however, in no way implied that they agree from subject to subject or style to style.

Style	Consistency	Inconsistency in 1 pair	Inconsistency in 2 pairs	Inconsistency in 3 pairs
Pop music	5	4	9	0
Classic	2	5	6	5
Pop/Rock	2	8	7	1
Jazz	4	8	4	2

Table 13.6: Responses for the annoyance part of the experiment. With 2 single judges pr. subject, it is possible to examine the agreement.

Pooled Check

To evaluate the pooled data, a Weak Stochastic Transitivity(WST) check, as described in Section 12.2.2, has been applied. If the WST is violated, the data should be viewed as inconsistent. There are no specific measure as to when the WST is violated crucially, but a count of the violations can be used to determine whether the responses can be considered consistent or not. The results from the WST check can be viewed in Table 13.7.

Style	Violations
Pop music	4 of 9 possible combinations
Classic	4 of 9 possible combinations
Pop/Rock	6 of 9 possible combinations
Jazz	4 of 9 possible combinations

Table 13.7: Weak Stochastic Transitivity check for the cumulative matrix from annoyance experiment.

Since there for all four sounds are violations, the responses must be considered inconsistent, hence it is not possible to extract model specifications or fit the data into a model.

13.2.2 Discussion

The consistency check revealed that the responses were not random, but anyway inconsistent, and the hypothesis therefore cannot be rejected. It is not possible to rank the stimuli in any other way than a straight ranking.

13.2.3 Straight Ranking Results

The straight ranking results are displayed in Table 13.8.

Style	**Least degraded**		**Most degraded**
Pop	Full comp.	No comp.	Bass comp.
Classic	Full comp.	Bass comp.	No comp.
Pop/Rock	Full + Bass comp.		No comp.
Jazz	Full comp.	Bass comp.	No comp.

Table 13.8: Straight ranking results for the annoyance experiment. Since the subjects are asked to choose the sample, where the noise is the most annoying, the data are presented inverted, so the one that is rated as the "least annoying" is denoted the best and visa versa. For sound 3 the bass compensation and the full compensation obtained the same rating value.

13.3 Concerted Discussion

There might be many reasons as to why consistency is not present in the data. The stimuli might not differ significantly enough for the subjects to either hear the difference, or, if the difference is noticeable, not determine a logic preference order among the three stimuli. Another reason could be that the subjects are changing their minds throughout the experiment due to some sort of learning effect. The learning effect has been minimized by randomizing the order of the stimuli.

Even though 7 subjects[4] in the naturalness experiment are consistent for all four kinds of music presented, they do not have the same preference order for the different kinds of music. This could be because of their normal preference on music types, but it could also be because of the characteristics of the specific type of music.

[4]The 7 fully consistent subjects are subject 3, 4, 6, 7, 8, 12 and 18.

Obviously, inconsistency is unwanted for a number of reasons. To minimize inconsistency, it make sense to examine possible factors. One reason could be that the subjects simply cannot perceive any difference between the processing methods. In the first part it has not been possible to reject this possibility but in the second part it seems reasonable to assume that there is indeed detectable differences. Since the pieces of music and the processing used is absolutely identical between part 1 and 2, it could be argued that it probably is not the inability to single out the processing method that causes inconsistency.

Other scenarios then comes to mind: does the listener change preferences throughout the first part of the listening experiment? Is the possible interpretations of "naturalness" to wide, making it too hard for the test subject to make consistent choices. More consistent data is perhaps achievable using a professional panel used to single out the different instruments.

E.g. for the jazz music, the unprocessed signal is highest ranked. The instrumentation consists of trumpet, double bass, piano and drums with no apparent bass drum playing. This sound sample is very natural in the sense that no synthesized sounds are included and it is also only processed to a very small degree by todays standards. Overall, the sound sample contains a small amount of energy in the bass region, compared to the pop and pop/rock styles. A trained listener would perhaps be better to single out the different instruments, thereby having an easier time to detect the bass compensation provided by the processing methods. An untrained listener might primarily focus on e.g. the lead instrument, the trumpet, thus reducing the attention to the double bass.

For further testing, it would make sense to acquire test subjects that normally pays attention to music, it could be musicians or HiFi enthusiasts. Training of the subjects is not recommended, as there are a risk of teaching the subjects "what the optimal method is".

Part IV

Conclusion and Future Studies

14 Conclusions

The goal of this thesis was to suggest one or more methods for optimizing the music reproduction in a noisy environment. It was decided to focus primarily on portable reproduction systems influenced by transportation noise. This type of noise is primarily found at the low end of the frequency spectrum. Two methods were suggested. The first focuses on amplifying the frequency contents of the music to a level were it could at the very least be heard. The second method also takes into account the possibly altered perceived tonal balance of the music and is considered a full compensation.

For the first method, a simple SNR approach was chosen. Any 1/3 octave band with music information was amplified to a level were it was deemed detectable in the noisy environment. This was performed on time frames of 1 second, mixing the changing filter gains using Bartlett windowing. With the noise used, this resulted in a significant boost of the lowest frequencies.

The second method also included a loudness model. The model was fit for assessing the loudness of time varying waveforms. This assessment were used to negate the altered perception of the music due to the noisy environment. The result was, in addition to the bass boost, a very subtle rise in the middle frequencies.

Four short pieces of music was processed using the methods. To test whether the methods were more or less suitable for different types of music, the music styles used consists of contemporary pop, pop/rock, classic and Jazz. It was argued that two important factors in optimal reproduction was the "naturalness" of the music and the degree of annoyance caused by noise. A listening experiment examining these two factors was conducted with 18 subjects, to establish if either method would increase the perceived naturalness and/or decrease the annoyance of the noise. The two methods were compared to each other and to the case of no alteration of the music. Comparison was made using the "paired comparison" method.

The two methods were compared to the case, where no method of compensation was applied, using paired comparison.

The results are rather inconclusive, but a straight ranking was achievable. Since straight ranking offers no measure of the degree of preference, these results should be investigated further. For achieving the highest degree of naturalness, the most preferred method was the full compensation for all pieces, except for the jazz music. To minimize the annoyance the full compensation was preferred for all pieces of music.

15
Further Studies

It is always possible to improve. In the following some of the areas of this thesis, that could be interesting points for further studies is outlined.

Perceptual Coding

Perceptual coding can be an effective method of reducing computational complexity and/or optimizing the sound processing. A lot of mobile music reproduction systems already have the MPEG 2 layer 3 codec. Maybe it is possible to reduce the processing power needed by using the information in the format.

Determination of Threshold Values

In the tests the SNR aimed for is set to -4 dB's. Perhaps this value is suboptimal. Further Research and perhaps experiments could be carried out, to achieve a better result.

Determining Just Notable Difference of Processing Methods

A possible reason for the inconclusive results could be that the test subjects were unable to perceive any difference between the processing methods. It would make sense to examine how big a processing difference in noisy environments is needed for the subject to detect it. This could produce a baseline for further tests.

Processing Methods Versus Music Style

It goes without saying that music without any low frequency content can never achieve the target SNR. This could be the case in choral harmonies, only containing female voices. On the other hand, heavily synthesized music pieces can contain very powerfully bass. It would make sense to investigate, if certain processing methods are optimal for certain music pieces. Perhaps a style detection algorithm could be implemented. A great deal of music is already tagged with a lot of digital information, often stating the musical style. This could also be used for individual processing.

Training of the Test Subjects or Using a Professional Panel

Some test subjects experienced the task as being "very difficult". It is possible that a trained or even professional listener would provide a much more coherent set of answers. Although this at first seems very desirable, it should be assessed if the processing scheme has any relevance if only highly trained people notice the difference.

Clarification of Task Description

In the listening test, the subject was first asked to choose the sound that *most resembled the reference*. In the second part, they were asked to choose the sample that was *most degraded by the noise*. Some test persons felt that this difference was "confusing". Perhaps more consistent results is attainable, if questions were rephrased.

Noise Level

In the listening test, the noise were set to 70 dB SPL A-weighted. If lower noise- and listening levels were to be used, the relative change in the ELLC would be larger. It would be of interest to see if the performance of the methods are altered at other levels.

Noise Characteristic

The noise in the experiment is from a real life situation. Other types of noise, especially regarding spectral components, could be investigated. Also, the used noise is considered quasi-stationary or, at the very least, dynamic to a very small degree. Noise with dynamic characteristics, motor start/stop or similar, could be used to test the performance of different algorithms.

Preferred Listening Level

In noisy environments people tend to increase the sound level so much that it induces unhealthy environment. An interesting research could be done by examining whether the preferred listening level is decreased when the noise compensation is applied.

Part V

Appendices

A
Train Noise Measurements

As it has not been possible to get calibrated measurements from any known sources, calibrated data has been collected by the writers of this thesis. The collected data comes from an IC3 train on the Aalborg - Hobro route.

The demands for the equipment used has been:

- Portability. As a huge setup is not feasible, a goal will be to minimize it.
- Known characteristics. Linearity, or lack thereof, isn't a priority, if correction is possible.
- Preferably a pressure field type, as the sound is from random incidence.
- CD quality, meaning a bit depth of 16 and a minimum sample rate of 44.1 kHz.

To satisfy the first demand, a prepolarized type will be used. For this purpose, the G.R.A.S. type 40AZ is chosen. Unfortunately, it is a freefield type. The only prepolarized pressure field type available in-house, has a even harder roll-off from around 9 kHz. Using the type 40AZ, a set of measurements will be done. If, after appropriate correction, the higher frequencies (above 10 kHz) is less than -20 dB with reference to the most dominant contribution, these measurements will be accepted.

The preamp used will be the G.R.A.S. 26CC and the recording device will be the Zoom H4 mobile hard disc recorder. It is capable of recording 16 bits quality at a sample rate of 48 kHz. It can be set to record at constant gain, thereby making it possible to record a reference tone, using the B&K 4230 calibrator.

Equipment	Name	AAU number
Microphone	G.R.A.S 40AZ	75545
Preamplifier	G.R.A.S. 26CC	75571
Recorder	Zoom Handy Recorder H4	64675
Computer	Lenovo 3000 N200	N/A
Calibrator	B&K 4230	08373

Table A.1: Measurement equipment used for recording train noise.

Around 7 minutes of cabin noise from the train were recorded and subsequently analyzed. To extract an absolute value (in Pascals), the calibrator was first used as reference. The recorded noise were filtered in 1/3 octave bands and the higher frequencies where amplified to take into account the worst case scenario with regards to sound incidence. The RMS value in each band were calculated in 1 second intervals. The [Dab file can be found on the CD.

A.1 Method

The "Worst case correction filter" was made by using the *firls* function in MatLab. It makes and FIR filter Least Squares approximation of the (inverted) 90 ° incidence angle, that can be read on the document accompanying the 40AZ. The FIR filter is set to 100 taps. The noise signal was then filtered, using the FIR coefficients, to achieve a high frequency amplitude lift.

A.2 Results

The results show remarkable little activity beyond 1 kHz, with the far dominant power in 32 Hz range. It should however be noted that the signal has *not* been A weighted. Since, for validation purposes, it's mostly interesting to facilitate that frequencies higher than 10 kHz is irrelevant, Figure A.1 shows the 32 Hz area, with most activity, while Figure A.2 shows the activity for frequencies 10 kHz and up.

The rather large peak around 240 seconds is due to a loudspeaker announcement.

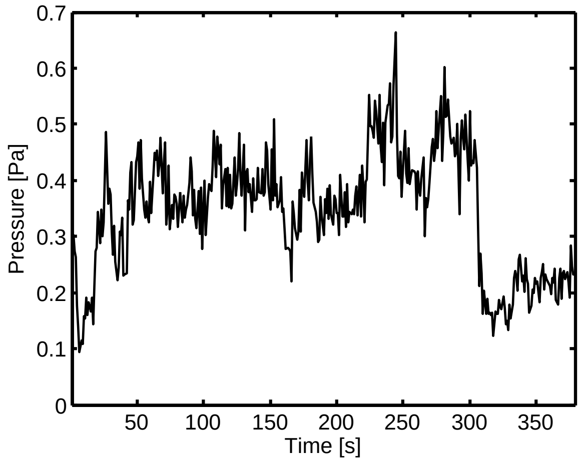
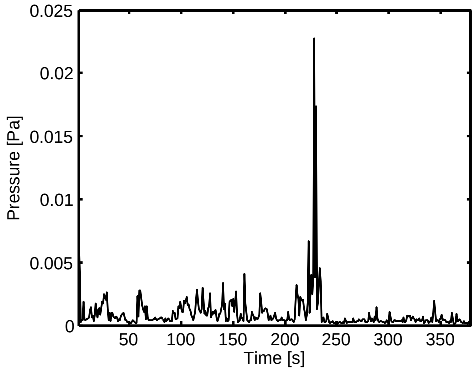


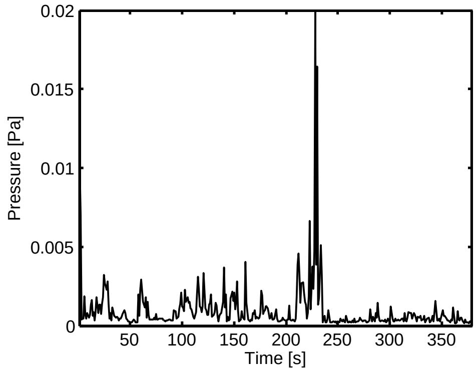
Figure A.1: The RMS pressure over time for 32 Hz center frequency.

A.3 Conclusion

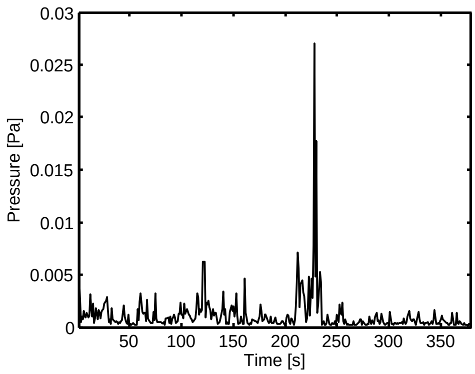
The used method of measurement is adequate.



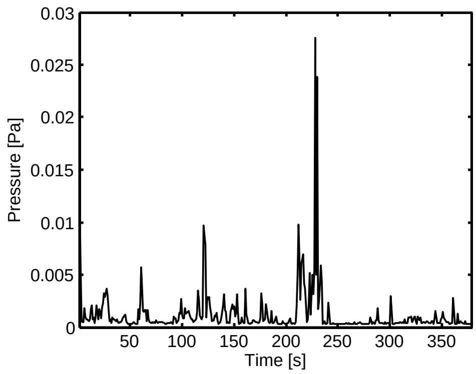
(a) 10 kHz center frequency



(b) 12.6 kHz center frequency



(c) 15.8 kHz center frequency



(d) 20 kHz center frequency

Figure A.2: The RMS pressure for frequencies from 10 kHz and up.

B

Headphone measurements

As the exact sound pressure must known to estimate perceived loudness and estimate ELLC's, a fully calibrated setup is needed. It was decided to investigate if in-ear types is more suitable than circumaural types to mitigate noise, primarily due to insertion loss. An rather inexpensive type, *Logitech PlayGear Stealth Headphones*, were tested, but this type of headphone is available in a huge price range and many quality grades, both sonically and with regard to user comfort.

B.1 The Soundcard

First, a validation of the external soundcard was made, using the setup in Figure B.1. The absolute levels have been ignored, as the parameter of interest is linearity.

Equipment	**Name**	**AAU number**
Soundcard	Creative soundblaster	2157-25
Computer	Lenovo 3000 N200	N/A
Multimeter	Fluke 37	08287

Table B.1: Measurement equipment used for testing and calibrating the sound card.

As the headphone output on the soundcard can be varied using a potentiometer, this was electrically shorted to provide full gain.

As seen in Figure B.2, the combined linearity of in- and output of the sound-

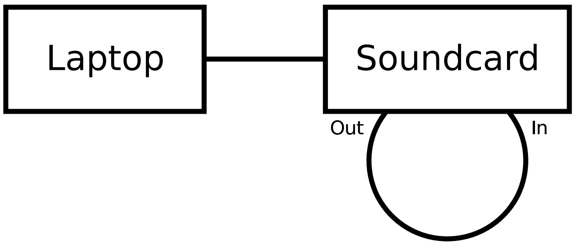


Figure B.1: Setup for test of the soundcard. The soundcard is connected to itself using a jack-jack cable.

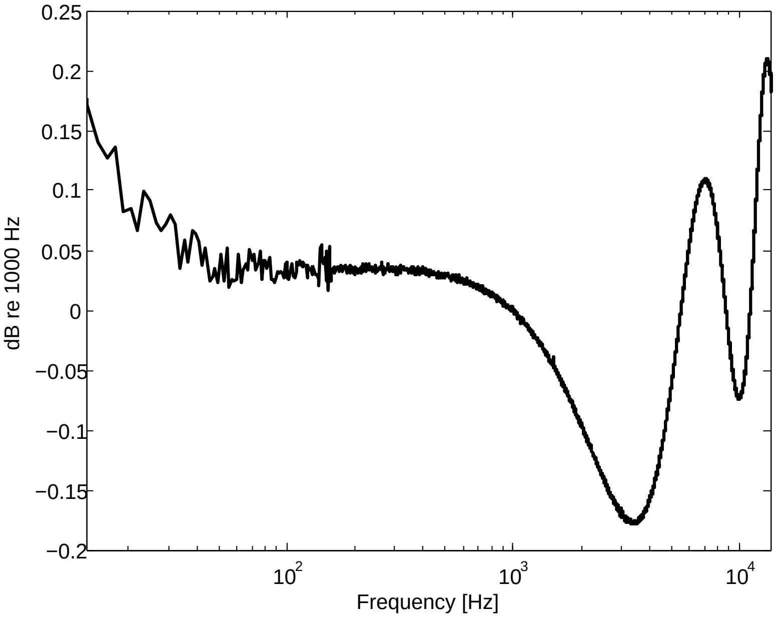


Figure B.2: Linearity vs. frequency from 20 to 20 kHz.

card is roughly ± 0.2 dB, relative to 1 kHz. This is considered well within reasonable limits.

Since both headphone transfer function and insertion loss is to be measured, it is necessary to measure or deduce, what the relationship between Pascal and MatLab units is. First, the output voltage as a function of MatLab units is measured, using the setup in Figure B.3.

A sine was generated in MatLab with peak values of 0.1 to 1 in 0.1 increments. The MM voltage was then recorded. results is seen in Table B.2.

As the actual voltage output is now known, the input and output on the soundcard is again shorted, as in Figure B.2, and the input units in MatLab

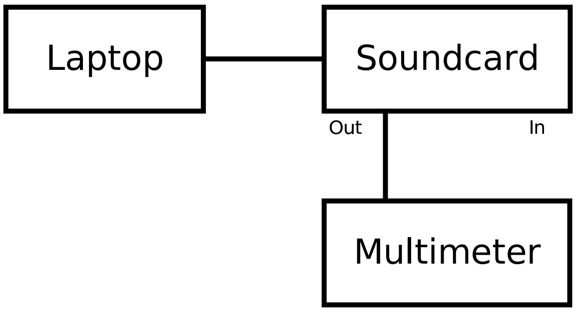


Figure B.3: Setup for calibration of the soundcard. The soundcard is connected to the multimeter.

Peak units []	**mV [rms]**	**mV [p]**	**mV/unit out**
0.1	61	86.3	862.7
0.2	122	172.5	862.7
0.3	183	258.8	862.7
0.4	243	343.7	859.1
0.5	304	429.9	859.8
0.6	365	516.2	860.3
0.7	427	603.9	862.7
0.8	487	688.7	860.9
0.9	548	775.0	861.1
1.0	609	861.3	861.3

Table B.2: Conversion from units in MatLab to voltage on the output of soundcard.

was recorded in Table B.3.

In both conversions, the gain results are all within 0.25 %, which is acceptable. It should also be noted that this is very close to the accuracy of the Fluke MM.

Results

A peak amplitude in MatLab of 1 will result in a peak voltage on the output of the soundcard of 0.86 volts. If 1 volt peak is presented to the input of the soundcard, a peak value of 0.95 units will be recorded in MatLab.

mV [p]	Units in [p]	Units in/V
86.3	0.082	0.952
172.5	0.164	0.951
258.8	0.246	0.951
343.7	0.328	0.955
429.9	0.410	0.954
516.2	0.492	0.953
603.9	0.574	0.950
688.7	0.656	0.952
775.0	0.738	0.952
861.3	0.820	0.952

Table B.3: Conversion from voltage on the input of soundcard to units in MatLab.

B.2 In Ear Headphones

The primary interest is linearity and insertion loss. For these tests, the equipment in Table B.4 were used.

Insertion Loss

As it is the low frequency region that is of interest, the insertion loss is simply measured for 30 Hz instead of the full frequency range. For a total insertion loss characteristic, a full range MLS measurement or similar is needed. This can be problematic to attain, as the sub woofer used has a variable low pass cut-off frequency of maximum 180 Hz and "normal" full range speakers often has very bad linearity around 40-50 Hz and lower. As the main insertion loss of interest is that around 30 Hz, this is the only frequency tested. The test setup is depicted on Figure B.4.

First, a test tone was played by the sub with the in ear headphone mounted in the earmold. It was checked, to the best of ability, that it was airtight sealed. The tone was recorded and the (unweighted) SPL were registered. This was repeated without the in ear headphone mounted. The results is shown on Table B.5. There is basically no insertion loss at very low frequencies, a little more that 1 dB.

Equipment	Name	AAU number
Soundcard	Creative soundblaster	2157-25
Computer	Lenovo 3000 N200	N/A
Multimeter	Fluke 37	08287
Microphone	B&K 4134	08129
Preamp	B&K 2669	56509
Conditioning amp.	B&K 2807	07305
Phantom power	ART phantom II	2157-55
Artificial ear	B&K 4153	07631
Head and torso simulator	B&K 4128	08453
Left and right ear simulator	B&K 4158/4159	08453-01/02
Power amp	B&K 2706	64654
In ear headphones	Logitech PlayGear Stealth	N/A
Headphones	Beyerdynamic DT990 PRO	2036-5
Earmold from art. head	N/A	N/A
Sub woofer	Dali SWA8	61413

Table B.4: Measurement equipment used.

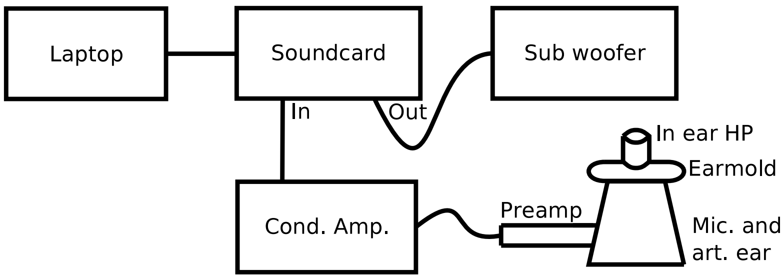


Figure B.4: Setup for testing insertion loss. The artificial ear is standing on a shock mount.

	Blocked	Unblocked
SPL	68.8 dB	70.2 dB

Table B.5: The insertion loss of the in ear headphones.

Linearity

The setup used for testing the linearity of the in ear headphones is seen on Figure B.5. For acquiring the impulse response a small MatLab MLS measuring system was made, based on [Van94] and [ea00].

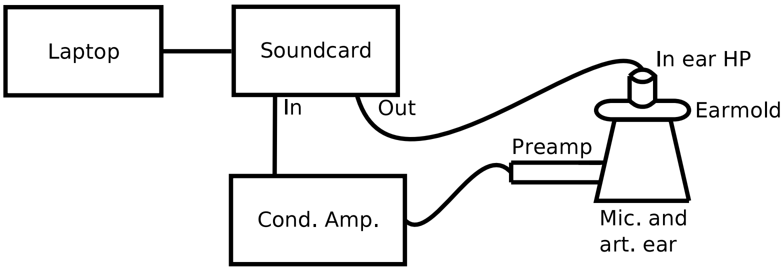


Figure B.5: Setup for measuring the linearity of the in ear headphones.

The earmold was used for two reasons. First, it is almost impossible to mount the in ear headphone on/in the artificial ear without it and secondly, if such a small transducer is required to reproduce very low frequencies, it must be in an airtight coupling condition. The result can be seen on Figure B.6.

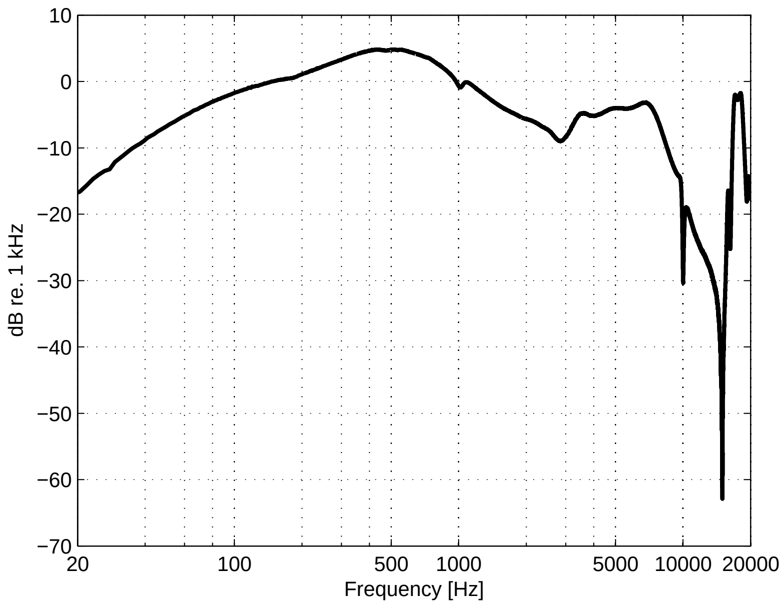


Figure B.6: The transfer function of the in ears.

In airtight coupling, the output drops around 12 dB between 20 and 40 Hz. Even a small leak will degrade bass reproduction further.

Wearing Comfort

The wearing comfort was tested by the authors of this thesis. One felt that the in ear headphones fitted reasonably, the other felt some discomfort. Further, and more importantly, one felt a distinct difference in the insertion loss to the surrounding environment, while the other did not experience much change. This could possibly be an issue of ear canal dimension.

Choice of Headphones Used

There is a number of disadvantages by using the in ear type. The most important is, that it can be hard to achieve the same degree of acoustic coupling between the headphone and the individual user. What is even worse, it is almost impossible to measure and thereby take into account this individual coupling. As it is paramount that the test subject are exposed to the same sound levels at a given frequency, it is decided to not use the in ear types for the listening experiment. However, if the noise is not primarily of low frequency nature, a user will probably still appreciate the possible insertion loss at mid and high frequencies.

B.3 Beyerdynamic DT990 PRO

The DT990 were chosen as the headphones to be used for listening tests. All tests were carried out, using the equipment in Table B.4. The insertion loss was not measured, due to expectation of an even smaller insertion loss, than that of the in ear type.

Linearity

As the headphones were rather difficult to mount on the artificial ear, a torso and head were used instead. The measurement setup is seen on figur B.7.

The ears were measured one at a time, using the MLS script used to test the in ear headphones. The results are seen on Figures B.8 and B.9.

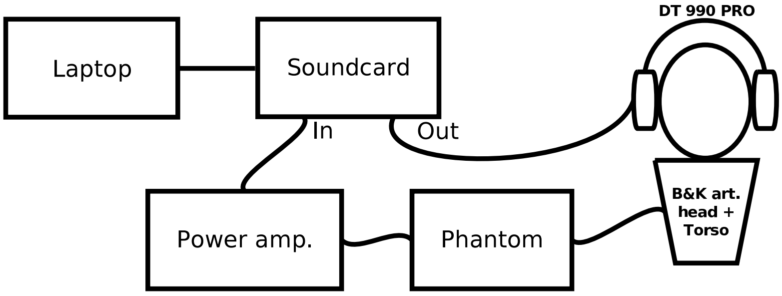


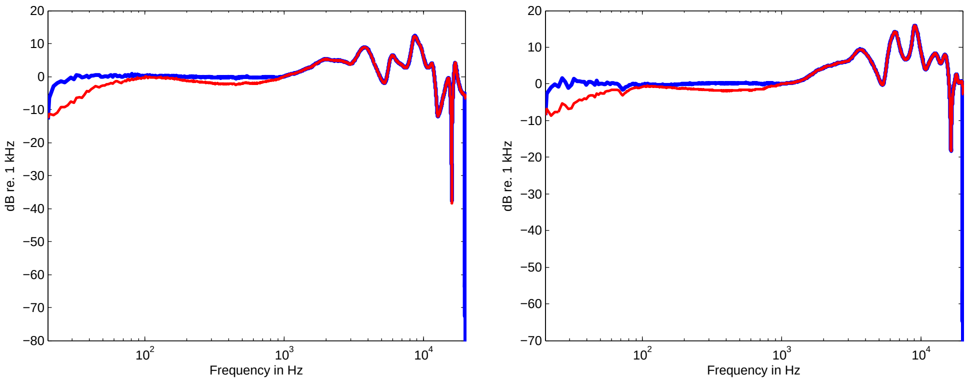
Figure B.7: Measuring the transfer function of the DT990 PRO.

Figure B.8: Transfer function for the left cup.

Figure B.9: Transfer function for the right cup.

Correction

To ensure correct bass reproduction,the transfer function of the DT990's were bass corrected. Since the headphones will be driven by one channel only, it is not possible to do a separate correction for left and right. Instead, the impulse responses were averaged and the FFT of this used as the base of correction. First, a FIR filtering correction were tried, using the "firls" function in MatLab. Since low frequency filtering is costly by filterorder, and even more so using no poles, a filterorder of around 10.000 were needed for reasonable results. As processing is done off line, this poses no problem in reality, so the filter order was set to 20000. The correction is only applied to frequencies lower than 1 kHz. Some of the variation seen in higher frequencies stems from the headphone being diffuse field equalized. Some originates in the mechanical shape of the ear, that acts as a resonator at certain frequencies. In this case, the most aggressive cancellation occurs around 15 kHz. Results of the correction are shown in Figure B.10.



(a) Left (b) Right

Figure B.10: The red line is uncorrected, the blue line is corrected.

The result is reasonable and this filtering is implemented in the "dt990_correction" script.

C

Calculations for the Individual Consistency Check

This is the full calculation for the probability to get the number of consistent answers in both part 1 and 2 of the listening test.

Part 1: Naturalness experiment

In this part of the listening test a matrix like Matrix C.1 will be generated by the subject's answers. This answer is inconsistent if and only if all row sums are equal to 1. The row sum α is defined in Equation C.2.

$$\begin{bmatrix} 0 & \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_3 & 0 & \mathrm{x}_4 \\ \mathrm{x}_5 & \mathrm{x}_6 & 0 \end{bmatrix} \tag{C.1}$$

$$\begin{bmatrix} \mathrm{x}_1 + \mathrm{x}_2 \\ \mathrm{x}_3 + \mathrm{x}_4 \\ \mathrm{x}_5 + \mathrm{x}_6 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{C.2}$$

If P(c) is the probability of a consistent answer and P(i) is the probability of an inconsistent answer, it follows that

$$P(c) = 1 - P(i)$$

Inconsistency is present if and only if all α are 1.

$$
\begin{aligned}
P(i) &= P(\alpha_1 = 1 \wedge \alpha_2 = 1 \wedge \alpha_3 = 1) \\
&= P(\alpha_1 = 1) \cdot P(\alpha_2 = 1 \mid \alpha_1 = 1) \cdot P(\alpha_3 = 1 \mid \alpha_1 = 1 \wedge \alpha_2 = 1)
\end{aligned}
$$

Where:

$$
\begin{aligned}
P(\alpha_1 = 1) &= P(x_1 = 1 \wedge x_2 = 0) + P(x_1 = 0 \wedge x_2 = 1) \\
&= 0.5
\end{aligned}
$$

and

$$
\begin{aligned}
&P(\alpha_2 = 1 \mid \alpha_1 = 1) \\
&= P(x_3 = 1 \wedge x_4 = 0) + P(x_3 = 0 \wedge x_4 = 1) \\
&= P(x_1 = 0 \wedge x_2 = 1) \cdot P(x_4 = 0) + P(x_1 = 1 \wedge x_2 = 0) \cdot P(x_4 = 1) \\
&= 0.25 \cdot 0.5 + 0.25 \cdot 0.5 \\
&= 0.25
\end{aligned}
$$

and

$$
P(\alpha_3 = 1 \mid \alpha_1 = 1 \wedge \alpha_2 = 1) = 1
$$

This implies that:

$$
P(c) = 1 - P(i) \qquad = 1 - 0.5^3 = 0.875 \tag{C.3}
$$

Part 2: Annoyance Experiment

In this part of the listening test a matrix like Matrix C.4 will be generated by the subject's answers. This answer is consistent if and only if the row sums are a permutation of Vector C.5.

$$
\begin{bmatrix} 0 & \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_3 & 0 & \mathrm{x}_4 \\ \mathrm{x}_5 & \mathrm{x}_6 & 0 \end{bmatrix} \tag{C.4}
$$

$$
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} \tag{C.5}
$$

This implies that

$$P(c) = 3! \cdot P(\alpha_1 = 0 \wedge \alpha_2 = 2 \wedge \alpha_3 = 4)$$

Where:

$$\begin{aligned} P(\alpha_1 = 0) &= P(x_1 = 0 \wedge x_2 = 0) \\ &= 0.5^4 \end{aligned}$$

and

$$\begin{aligned} P(\alpha_2 = 2 \mid \alpha_1 = 0) &= P(x_4 = 0) \\ &= 0.5^2 \end{aligned}$$

and

$$P(\alpha_3 = 4 \mid \alpha_1 = 0 \wedge \alpha_2 = 2) = 1$$

That gives a probability of consistency of:

$$\begin{aligned} P(c) &= 3! \cdot 0.5^4 \cdot 0.5^2 \cdot 1 \\ &= 6 \cdot 0.5^6 \\ &\approx 0.094 \end{aligned}$$

D Introduction and Questionnaire

D.1 Initial Introduction

At first we would like to thank you very much for participating in our experiment. Before starting we would like you to read this introduction and fill out the questionnaire. If you have any questions, please do not hesitate to ask us. Also note that you are not obliged to finish the experiment. You can leave at any time during the experiment should you so wish.

The sound levels you are exposed to are in no way hazardous.

This experiment investigates the perception of music quality in noisy environments. Since it is your *opinion* that is of interest, there is never a right or wrong answer, just answer the best you can to the questions asked throughout the experiment. Feel free to use as much time as you like, before giving your answer.

The experiment consists of two parts, with a break in between.

In the cabin, a computer program will guide you through the test.

In the first part we would like you to focus on the music signal itself and try to ignore the noise. You will be presented to a reference sound and two soundsamples called "A" and "B". Your task is to choose if sample A or B sounds most identical to the reference signal.

We would like you to try to disregard the noise on the signal, and *focus on the naturalness* of the signal, and pick the one of the two that sounds most like the reference.

There will be 4 different pieces of music presented 3 times each. Make, sure that the headphones are placed correctly for left - and right ear.

After the first part, there will be a break. In this break, we will inform you of your task in the second part.

Finally, we will ask you to leave your phone in the waiting area while you participate in the experiment.

D.2 Introduction: Part 2

In this part there will be no reference signal. Your task will be to pick sample, where the music experience is the most degraded by the noise.

The music will be presented automatically *and only once*, so try to concentrate as best as you can. Press the "A" button, if you feel that the music experience is degraded the most by the noise in the first piece of music, or press the "B" if it is the second piece.

The music will be identical to the 4 pieces in the first part of the test, but each will now be presented 6 times.

You are welcome to ask any question about the task.

Make, once again, sure that the headphones are placed correctly for left - and right ear.

D.3 Questionnaire for the Listening Experiment

Thank you very much for participating in our experiment. We would like you to answer following questionnaire. The answers will be treated as strictly confidential information.

Personal Information

Full name:	____	
Gender:	◯ Male	◯ Female
Age:	____________	

Initial health questions

1. Do you have some hearing disorder?
◯ Yes ◯ No

- If yes, which kind?

  ____

2. Do you have a cold or are you suffering from hay fever?
◯ Yes ◯ No ◯ To some extend

3. Have you been exposed to high levels of sound, such as a concert, within the last 48 hours?
◯ Yes ◯ No

Background information

1. How many hours a day do you in average listen to music:

(a) On headphones?

0-1 hour	2-3 hours	4-5 hours	6-7 hours	8-9 hours	More than 9 hours
◯	◯	◯	◯	◯	◯

(b) On loudspeakers?

0-1 hour	2-3 hours	4-5 hours	6-7 hours	8-9 hours	More than 9 hours
◯	◯	◯	◯	◯	◯

2. When do you like listening to music? *(You may mark more than one answer)*

- ◯ When you work
- ◯ When you relax
- ◯ In the evenings
- ◯ When you are physically active
- ◯ When traveling from one place to another, eg. on your way to work
- ◯ Never
- ◯ Other
  *Please state when:*____________

3. How do you normally get to - and from school or work?

- ◯ Bus
- ◯ Train
- ◯ Car
- ◯ Bike
- ◯ Walk
- ◯ Other
  *Please state what:*____________

4. To what extend do you find the background noise annoying when traveling?

Not at all	A little	Some	A lot	Very much	Don't know
◯	◯	◯	◯	◯	◯

Thank you for answering our questions.
Kind regards,
Group 1060 at the Acoustics Master Program
AAU, Spring 2010

E
Setup of Listening Test

The room used to conduct the listening test was the audiometry cabin B in the acoustics department of Aalborg University, Frederik Bajers vej 7. A sketch of the room and the setup used is depicted on Figure E.1.

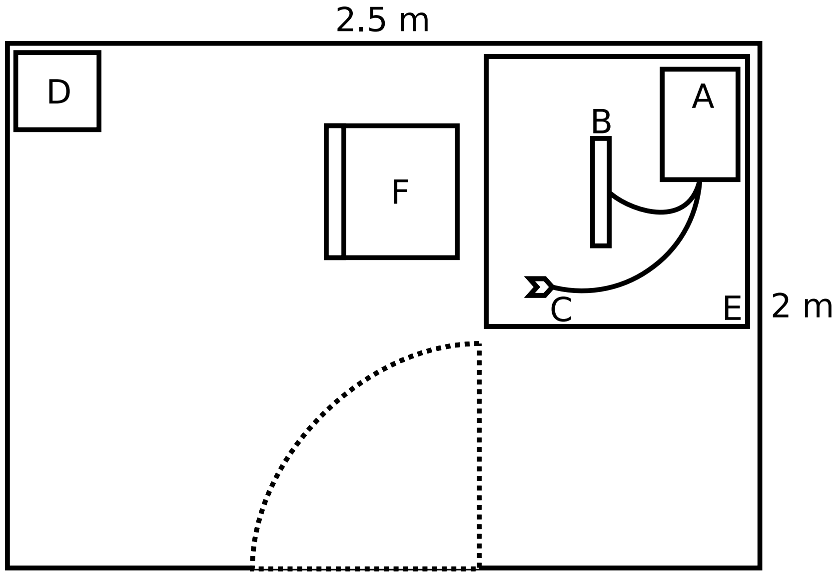


Figure E.1: Setup used for listening test.

Equipment used is found in Table E.1.

Pressure Field or Diffuse Field

The largest dimension of the room is roughly 2.5 meters. With respect to room modes, wavelength equal to double the rooms dimension can be a problem. In this case, the first room mode is around:

Item	Name	Type	Aau No.
A	Lenovo laptop	3000 N200	N/A
B	PC flat-screen	LG flatron L1710S	57417
C	Bluetooth mouse	Notebook mouse 5000	N/A
D	Sub woofer	Dali SWA8	61413
E	Table	N/A	N/A
F	Chair	N/A	N/A
G	Soundcard	Creative soundblaster	2157-25
H	Headphones	DT990 PRO	2036-5

Table E.1: List of items used for the listening experiment. Item G was mounted underneath the table. Item H was placed on the table.

$$\frac{\lambda}{2} = 2.5m \Rightarrow f = 68.6Hz$$

where:

λ is the wavelength $[m]$
f is the frequency $[Hz]$

Since the main energy is focused around 32 Hz, less than half the first room mode, we choose to look at it from a pressure chamber point of view, meaning that the pressure is identical in all of the cabin. To make sure that this was the case, the pressure field were measured in three points around 0.2 meters apart, were the test subject would sit. All measurements were within 1 dB, and these measurements also were the base for calibrating the subwoofer.

Interface

The notebook was closed and placed on the table, out of the way of the test subject. The external soundcard was mounted underneath the table, out of sight of the test subject.

The test subject interacted directly with the monitor, mouse and headphones, via the GUI.

The subject was instructed to put on the headphones, sit down comfortably and follow the on screen instruction.

Bibliography

[Bey] *DT990 PRO datasheet.*

[Dav88] H. A. David. *The Method of Paired Comparisons.* Charles Griffing & Company LTD, London, Oxford University Press, New York, 1st edition, 1988.

[ea00] Søren Krarup Olesen et al. An improved mls measurement system for aquiring room impulse responses. *Nordsig 2000*, pages 117–120, 2000.

[EZS57] G. Flottorp E. Zwicker and S. S. Stevens. Critical band width in loudness summation. *Journal of the Acoustic Society of America*, volume 29:548–557, 1957.

[Fle40] Harvey Fletcher. Auditory patterns. *Reviews of modern physics*, volume 12:47–66, 1940.

[FM33] Harvey Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation. *Journal of the Acoustic Society of America*, Volume 5:82–108, 1933.

[FM37] Harvey Fletcher and W. A. Munson. Relation between loudness and masking. *Journal of the Acoustic Society of America*, Volume 9:1–10, 1937.

[FS24] Harvey Fletcher and J. C. Steinberg. Loudness of a complex sound. *Physics*, Rev. 24:306–, 1924.

[FZ90] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: Facts and models.* Springer, 1st edition, 1990.

[FZ07] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: Facts and models.* Springer, 3rd edition, 2007.

[HC02] Yuan-Hao Huang and Tzi-Dar Chiueh. A new audio coding scheme using forward masking model and perceptually weighted vector quantization. *IEEE transactions on speech and audio processing*, Volume 10:no. 5, July 2002.

[ISO75] *DS/ISO 532 Fremgangsmåde til beregning af hørestyrke*, 1st edition, 1975.

[Lon05] Marshall Long. *Architectural Acoustics.* Academic Press, 1st edition, 2005.

[MG96] Brian C. J. Moore and Brian R. Glasberg. A revision of zwicker's loudness model. *Acta acustica*, 82:335–345, 1996.

[MG02] Brian C. J. Moore and Brian R. Glasberg. A model of loudness applicable to time-varying sounds. *Journal of the Acoustic Society of America*, 50, no. 5:331–342, 2002.

[Mil] Miljøvejledning for transportydelser.

[Mø06] Aage R. Møller. *Hearing.* Academic Press, 2nd edition, 2006.

[Moo95] Brian C. J. Moore. *Hearing.* Academic Press, 2nd edition, 1995.

[Moo03] Brian C. J. Moore. *An introduction to the psychology of hearing.* Academic Press, 5th edition, 2003.

[SSSN37] John Volkman Stanley Smith Stevens and Edwin Newman. A scale for the measuremnt of the psychological magnitude of pitch. *Journal of the acoustical society of america*, Volume 8:185–190, 1937.

[Ste25] J. C. Steinberg. The loudness of a sound and its physical stimulus. *Physics*, Rev. 26:507–, 1925.

[Van94] John Vanderkooy. Aspects of mls measuring systems. *Journal of audio engineering society*, 42 (4):219–231, 1994.

[WS04] Florian Wickelmaier and Christian Schmid. A matlab function to estimate choice model parameters from paired-comparison data. *Behaviour Research Methods, Instruments & Computers 36(1), 29-40*, 2004.

[ZÖ2] Udo Zölzer, editor. *DAFX Digital audio effects.* Wiley, 1st edition, 2002. ISBN 0 471 49078 4.

[ZE03] Karin Zimmer and Wolfgang Ellermeier. Deriving ratio-scale measures og sound quality from preference judgements. *Noise Control Eng.*, 51 (4):210–215, 2003.

[ZS65] Eberhard Zwicker and Bertram Scharf. A model of loudness summation. *Psychological Review*, 72, no. 1:3–26, 1965.

[Zwi58] E. Zwicker. Über psychologische und metodische grundlagen der lautheit. *Akustische Beihefte*, Vol. 8, Heft 1:237–258, 1958.

[Zwi60] Eberhard Zwicker. Ein verfahren zur berechnung der lautstärke. *Acustica*, Volume 10:304–308, 1960.

[Zwi61] Eberhard Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *Journal of the Acoustic Society of America*, Volume 33:248, 1961.