universität
wien

MASTERARBEIT

Titel der Masterarbeit

"Loudness and auditory masking for Mobile TV"

Verfasser

Ing. Martijn Cornelis Sack

Angestrebter akademischer Grad

Diplom-Ingenieur

(Dipl.-Ing.)

Wien, 2011

Studienkennzahl lt. Studienblatt	A 066935
Studien lt. Studienblatt	Master studium Medien Informatik
Betreuer	Ao. Univ. Prof. Dr. Helmut Hlavacs

universität wien

Selbständigkeitserklärung

Hereby I declare that I wrote the Master Thesis with the title

Loudness and auditory masking for Mobile TV

by myself and no one else except the sources that are mentioned in the Bibliography.

Vienna, 7-10-2011

Signature

universität wien

Acknowledgement

The thesis that lies before you would not exist if my parents would not have inspired and supported me by taking the challenge to move to Vienna. Papa 57/6 en mama, bedankt!

Also a big inspirer who made a lot possible for this thesis is Helmut Hlavacs. He enabled me, together with Shelley Buchinger, to go to China so I could present a paper [1] on a conference. This meant a lot to me. Karin Anna Hummel must be mentioned as well for contributing her knowledge in mobile computing. For the rest I would like to thank my CACMTV colleagues, Tijnemans!, Evilein, Horsie, Claudia, people who took part in experiments, ESN, IEEE BMS, Hollandse avond Wenen, The Finnish, Wilhelm, Nelly and Kaja. You and all the others made my student life in Vienna unforgettable!!!

Abstract

In the past decades the consumption of multimedia content has been increased dramatically. First by portable cassette and compact disc players, later digital files and now live TV and radio streaming on the internet. The available content is quite diverse including classical concerts, audio books, sports matches, home made videos and so on. The sources of this content are also diverse and so are the environments of content consumption. The wish to consume every type of content on every location is clearly there, but is the Quality of Experience (QoE) also as high as the wish is clear? The answer is no. Environmental elements are influencing the QoE. This influence is often perceived as annoying and disturbing. Examples of environmental influences are sunlight, a shaking seat in a bus or train and loud noise. The scope of this thesis will be the audio domain.

This thesis focuses on tackling the problem that users of mobile multimedia are having with going from one environment to the other while consuming multimedia content. In attempt to increase the Quality of Experience, several experiments with test persons are done. The description, setup and results of these experiments are presented in this thesis.

This thesis does not only present a method to increase the QoE of mobile consumption of audio, but also an implementation suggestion of how this technology could be brought to everyday situations. The application suggestion is supported with advices for content producers, broadcasters and mobile device vendors.
This paper concludes with a summary of booked successes during this research.

universität wien

In den letzten Jahrzehnten hat sich die Produktion von multimedialen Inhalten dramatisch erhöht. Wurden zunächst Schallplatten, tragbarere Kassetten und später Compact Discs erzeugt, wird heute der Großteil multimedialer Inhalte auf digitalen Files, oder live über Digital-TV oder Internet-Streams gespeichert bzw. übertragen. Multimediale Inhalte zeichnen sich durch eine große Bandbreite an Inhalten aus, dazu gehören etwa klassische Konzerte, Hörbücher, Sportübertragungen, selbst produzierte Videos und vieles mehr. Auch die Quellen und Konsumenten unterscheiden sich oft beträchtlich. Dabei zeichnet sich der Wunsch der Konsumenten ab, Inhalte überall und jederzeit konsumieren zu können. Tatsächlich ist die mögliche Qualität etwa von mobilen Konsumationsmöglichkeiten noch beschränkt. Bei mobilem Konsum beeinflusst die Umwelt die erlebte Qualität der Inhalte entscheidend, welche oft als störend oder ärgerlich empfunden wird. Beispiele umweltbedingter Störungen sind etwa Sonnenlicht, Rütteln in Bus oder Bahn, und laute Hintergrundgeräusche. Diese Arbeit beschäftigt sich mit letzterem. Konkret werden Probleme untersucht, die auftreten wenn sich Konsumenten mobiler Inhalte von einer Umgebung zu anderen bewegen. Im Rahmen der Arbeit wurden zahlreiche Experimente mit Testpersonen durchgeführt. Der Aufbau und die Resulate dieser Experimente werden in der Arbeit beschrieben. Über rein technologische Maßnahmen um die subjektive Qualität mobiler Konsumation zu steigern hinaus wird in dieser Arbeit auch beschrieben, wie die von mir entwickelten Techniken real umgesetzt werden können. Dabei werden auch Umsetzungsprozesse für Prodzenten, Broadcaster, und Hardware-Hersteller angegeben. Die Arbeit wird mit einer Zusammenfassung der beschriebenen Innovationen abgeschlossen.

Contents

1 Introduction

1.1 Motivation

During the master study that I am finishing at this moment of writing, I had to attend the course Praktikum aus der Medieninformatik. During the first lecture of this course there was a representative of the research project CACMTV, Shelley Buchinger. She explained that this project has the goal to increase the Quality of Experience (QoE) regarding consuming video content on mobile devices. It was soon very clear that the majority of the researchers in this project were aiming for increasing the QoE of only video in the content. No researcher was doing any research on audio within the CACMTV project. As a professional broadcast engineer I heard very often the saying 'without proper audio there is no proper video'. I cannot say anything else that I fully agree on this. Supported by my daily annoyances of improperly mastered audio for mobile consumption I thought that this might be a good opportunity to tackle this problem.

This was the starting point of doing research with the aim to increase the QoE on audio for mobile content consumption. After finishing the course and doing some first experiments, I got the invitation to join the CACMTV group for doing more investigation on this topic. I did this successfully and I got even a paper published on this topic and presented a part of my results in Shanghai, China on the IEEE conference '2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting'.

1.2 IEEE International Symposium 2010

This master thesis is an additive to a paper which was presented by the author of this thesis on the IEEE conference '2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting'[1]. The conference brought a lot of inspiration to continue doing research on the topics that were brought up during that conference. Topics of the conference:

- Multimedia systems and services
- Transmission and networking
- Multimedia processing
- Multimedia Quality: Performance Evaluation
- Multimedia devices

Key researchers and key decision makers from all over the world were attending the conference e.g.:

- Dr. Peter Siebert - Executive Director of the DVB Project Office
- Prof. Ulrich Reimers – Chair of DVB Technical model, Professor at Technical University of Braunschweig, Germany
- Mark Richer - President, ATSC, USA
- Craig Todd - CTO, Dolby Laboratories, USA
- Keiichi Kubota - Director General, NHK Science and Technology Research Laboratories, Japan

Since the conference did not have many attendees, the atmosphere was very informal and thus easy to communicate with everybody.

[1] http://www.ieee-bmsb2010.org

1.3 Added value of this thesis

There are several reasons for writing a master thesis about this topic. The main reason is that there never was a proper opportunity to present fundamental theories and audio processing methods that were used in the frame of this research. The second reason is that all loudness questions regarding television and radio broadcasting seem to be answered in the past months with publications from the European Broadcast Union (EBU). However, there is still a long way to go for mobile content consumption.

By publishing research results and advices, this thesis will contribute to recent creation processes in developing new standards and recommendations. New developments in this field will help content providers to increase the QoE for mobile content users. This will finally increase the QoE in mobile content consuming. This master thesis contains also a concept of a real life application.

This application concept should help the community to find a way to apply the results of the loudness and auditory masking for mobile TV research.

2 Audio dynamics basics

Since the majority of this master thesis is about dynamical behavior and treatment of audio, a brief introduction is on its place. Describing audio properties of productions, transmissions and recordings can be done with many parameters e.g. total harmonic distortion (THD), frequency response, dynamic ranges, etc. This master thesis sticks mainly with the last one.

Cultivation of dynamic behavior in audio has in general two main purposes: artistic purposes and shaping audio signals for transmission and reception. Producers of content, e.g. moviemakers, orchestra's or radio stations have the freedom of producing their content the way they like. These producers are taking context, public, production, transmission and publication medium into account. A moviemaker produces for a cinema, an orchestra for a concert hall and a radio station for a radio reception in a car. Each and every context, media or public is having its own constraints. Producers must adapt to these constraints to reach a preferable end result: a high Quality of Experience (QoE) for the end users. Most of these productions are not only available for the original context, media or public, but can be heard or watched in other contexts as well. Transformation and optimization of content for usage in new contexts is an important and challenging task, which must be done very carefully to keep the aimed QoE at a high level.

This chapter will explain the basics about dynamics of audio related to their production, transmission and end user context.

2.1 Dynamic ranges

A dynamic range describes the window in which a signal can or is allowed to exist. In the context of this thesis, this is an audio signal. The used unit for describing

the values related to a dynamical range is decibel or dB. A dynamic range can be described in the analog electrical and digital quantized domain. The dynamic range is presented in a logarithmic scale.

$$20 \log_{10} \left(\frac{U_{max}}{U_{min}} \right) = dynamic\, range\ [dB] \tag{2.1}$$

Formula (2.1) describes how to calculate a dynamical range in the analog electrical domain. U_{max} and U_{min} are values in [Volt]. They describe the maximum and minimum electrical value in which a signal can exist or is allowed to exist. The outcome of the formula is in [dB].

$$20 \log_{10} \left(2^{bit\, depth} \sqrt{\frac{3}{2}} \right) = dynamic\, range\ [dB] \tag{2.2}$$

Formula (2.2) describes how to calculate a dynamical range in the digital quantized domain. Normally the theoretical noise floor depends on quantization noise. The term *bit depth* expresses the word-length (e.g. 8, 16, 24 bit) of the digital signal to be describe. According to this formula a Compact Disc, which has a bit depth of 16 bits, has as medium a dynamic range of approximately 98 dB for RMS noise floor.

The ways to determine dynamic ranges presented above can be used to determine signal, storage or transmission media limitations.

2.2 History of audio dynamics in broadcasting

Broadcasters are bound by the limitations of the different media that they are using to produce and/or publicize their productions. In order to do this properly they have to use instruments to measure used signals. Over the years many different broadcasting regulations and even more metering methods have been published for analog as well as for digital productions. Most of these solutions have the same goal; never to exceed the upper or lower limit of a medium.

The following example will explain why regulation and good metering of dynamic ranges are trivial for proper broadcasting.

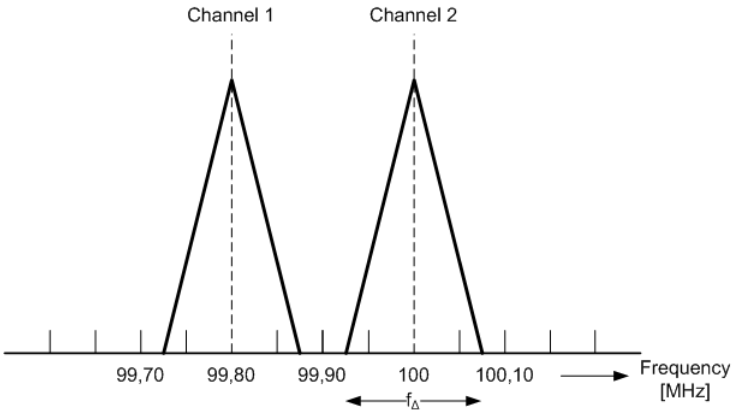


Figure 2.1: Frequency spacing of two FM radio stations

Figure 2.1 shows a typical spacing structure of the frequency spectrum used in modern radio broadcasting. Shown here are two users of the spectrum. In this example there are two FM radio stations: Channel 1 and Channel 2. According to regulations the space in the spectrum between two base frequencies of transmitters is 200 kHz. Each channel has a base frequency. The base frequency of Channel 1 is 99.80 MHz and of Channel 2 it is 100 MHz. By modulating an information signal on this base frequency, side bands will appear. These side bands are taking extra space in the spectrum. Regular frequency modulation (FM) used for broadcasting produces a side band under and above the base frequency. These side bands are allowed to become 75 kHz wide. All together the whole station uses 150 kHz, $f_\Delta = 150\,\text{kHz}$.

It is very important that these rules are maintained. When for example the $f_\triangle$ of Channel 1 becomes 275 kHz instead of 150 kHz, it will interfere the transmission of Channel 2. This interference can have several causes. The most common cause is when a transmitter gets an input signal with a too big dynamic range. That is why a proper transmitter has always a built-in limiter to avoid this.

This limiter will produce a distorted signal while it is limiting. It is the task of a broadcast engineer that the limiter in a transmitter will not be used.

The solution for broadcast engineers for avoiding trouble is proper metering of the used signals. In the past almost every broadcast company created its own way of dealing with this problem. Even until today there are many different methods of measurement and metering scales, not only for analog production environments, but also for digital production environments. In Europe there is an organization, European Broadcast Union (EBU), which represents most of the public broadcasters in Europe. A task of this organization is to unify broadcasting techniques on a pan-European scale. This means that all European public broadcasters should use the same broadcasting standards and methods.

Recommendations for analogue & digital audio level	Alignment Level (AL) -9 dB (35%)	Nominal Level (PML)[a] 0 dB (100%)
ITU-R BS.645-2 Transmission Level (international)	0 dBu[b]	+9 dBu
ARD HFBL-K Studio Level (national)	-3 dB (adaptation)	+6 dBu (adaptation)
US Reference Level (national)	-	+4 dBu (adaptation)
EBU digital Transmission & Studio Level (international)	-18 dBFS	-9 dBFS[c]

Table 2.1: Audio levels in studio and transmission environments [2]

a. PML = Permitted Maximum Level
b. 0 dBu = 0.775 V rms (sine wave) = 1.1 V peak
c. dBFS = Clipping Level (FS = Full Scale)

When calibrating or using a signal meant for broadcasting it is necessary to let the signal stay within its boundaries. To determine whether the signal is within the tolerated ranges, the signal must be metered and aligned. Table 2.1 shows different interpretations of the ideal way of aligning signals for as well the analog and digital domain. When a broadcast environment needs to be calibrated, the environment

needs to be aligned. This is done by sending a (mostly) 1 kHz signal though the broadcast environment and adjust the level of the signal to the alignment level of the used audiometer. The described alignment methods, except the US Reference Level, have one thing in common, the 9 dB headroom. The headroom is the space between alignment level and the nominal level. The reason of this 9 dB headroom is to deal with peaks in the signal. The EBU digital Transmission & Studio Level has a 9 dB negative offset, this is due to inter-sample peaks that cannot be measured but still can exist. For more details of audio metering in broadcasting read [2]. This document describes in detail how levels can be measured for different goals. These topics are out of the scope of this thesis.

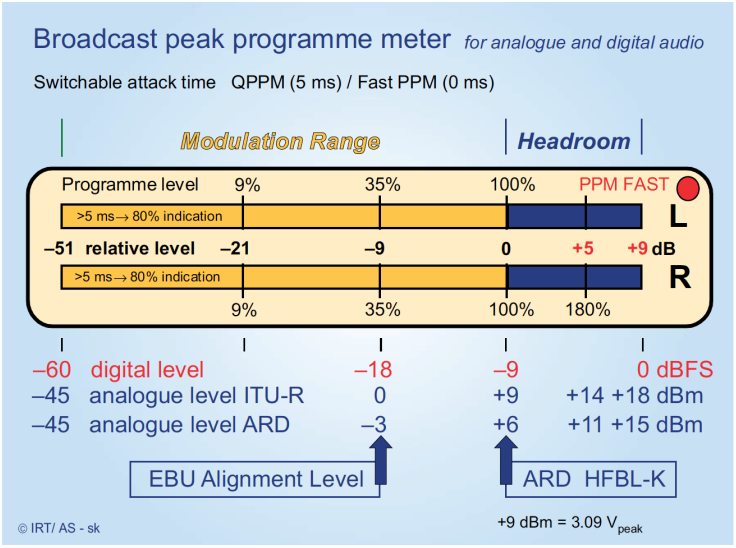


Figure 2.2: Recommended broadcast peak-programme meter [2]

Figure 2.2 gives an overview of how peak-programme meters can be designed. Every meter has a modulation range which defines the borders of the space that the broadcaster can use for its programs, respectively 42 dB, from -51 dB till -9 dB on the relative level scale. The broadcaster aligns a 1 kHz sine wave to this point. There is 9 dB headroom for peaks in the program signal, -9 dB till 0 dB on the relative level scale.

The signal may never exceed the permitted maximum level, 0 dB on the relative level scale to avoid interference with other broadcasters or a clipping signal.

2.3 Typical desired dynamic footprints

In the previous section the dynamic range of FM radio broadcasting is used to explain the constraints of a medium. There is also another reason to maintain a dynamic range; this is from the point of view of how a content user perceives content. The user experience of content depends strongly on the context where the content user is consuming this content. Every context consists of an environment and a certain user behavior. This section will introduce the reader to typical preferred dynamic ranges vs. environments.

Audio in multimedia content can be presented in various formats and by various media, e.g. a movie in a cinema over high-end sound-system, a television production in a living room over small loudspeakers and music on an iPod through small cheap ear-phones at a bus stop. Every environment has different types of noise. A cinema is a well protected, almost noise free, static environment. However, a typical user of an iPod has nomadic behavior, e.g. walking in the street or sitting in a bus, and therefore an iPod user is exposed to constantly changing and sometimes-strong environmental noises. Every user context needs a specific treatment concerning dynamic ranges.

Figure 2.3 shows different environments vs. typical dynamic footprints and has a relative scale in dB. A dynamic footprint defines three parts; headroom, preferred average area and noise floor. The preferred area describes the dynamic range between the lowest audio level and the top of the desired average audio level. Below the preferred area, the noise floor described. The noise floor describes the level of environmental noise. The headroom is the area above the preferred average area. In the headroom area there is room for peaks in the audio signal. It is recommended to take these 'dynamic guidelines' in consideration while producing or transforming content for a certain environment.

Cinema A cinema is an acoustically closed area. This means that influences of environmental noise are reduced as much as possible. On the other end, it is also possible to produce a lot of loud sounds burst without causing noise around the cinema. A cinema has the goal to give a very true reflection of how a moviemaker meant his movie to be. The noise floor is reduced to the level of -37 dB. This gives a dynamic range of 37 dB for the average sound. Peaks can be played back at +24 dB, which is a very high level. The total dynamic range is 61 dB.

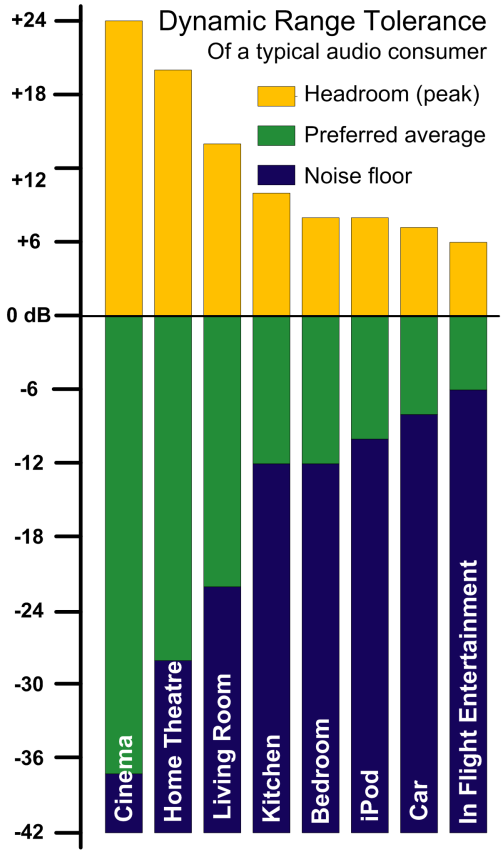


Figure 2.3: Dynamic Range Tolerance for consumers under different conditions [3]

Home-theatre A home theatre is a lesser noise-protected area than a cinema. A home cinema is situated in a person's home. Although a home cinema can be optimized for playing back very silent and loud movie scenes, it is still a house with neighbors and people who are living their lives. The noise floor is with -28 dB less extremely low as in a cinema, the headroom is still big with 20 dB. The dynamic range is 48 dB.

Living-room In a living room are people living their lives. This means e.g. talking, relaxing, working and of course watching TV. A big difference between a cinema, home theatre and living room is that the living room is a multi purpose area and not an area that is meant for watching movies only. Environmental noise is therefore often a disturbing factor and the playback equipment is not always high-end either. Both limitations make it necessary to reduce the average sound range to 22 dB and the headroom to 14 dB. A living room has a preferred total dynamic range of 36 dB.

Kitchen A kitchen is an area in the house where is a lot going on. There are people talking, eating and there is of course being cooked. This brings a lot of noise, which must be overcome. For this reason the noise floor is at -12dB. Another limitation is the equipment. A kitchen with state-of-the-art a TV-set is rare. It is also very annoying to turn the volume of the TV constantly higher and lower while cooking. That's why the headroom is reduced to 10 dB. The total dynamic range is then 22 dB.

Bedroom When people are laying in a bedroom, most of the time it is quiet. The issue is not that there is a lot of environmental noise, but it is needed to keep the dynamic range low. It is not desired that loud explosions be played in full intensity. Therefore the headroom is set at 8 dB and the noise floor at -12 dB, in total a dynamic range of 20 dB.

iPod An iPod is a mobile device that can be used at any place and at any time. The user of the iPod is faced to all types of environmental noise. This has an influence on the expected noise floor and an optimal dynamic range. To stay safe, a dynamic range of 18 dB is preferred. The preferred dynamic range

consists of a preferred average of 10 dB and a headroom of 8 dB.

Car A consumer of audio in a car has to deal with noise while driving. This noise can vary from a quiet background noise while driving slow and loud while driving fast on a motorway. Most of the in-car entertainment systems are able to play at a high volume, by this feature environmental noise can be overcome. The noise floor is at -8 dB while the headroom is 8 dB. These together gives a dynamic range of 16 dB.

In-flight-entertainment The dynamic footprint with the smallest preferred dynamic range is in-flight-entertainment. A content consumer has to deal with a lot of environmental noise of the airplane and people in the surrounding. Another limitation is the, often poor, quality of the provided headphones by the airliner. The noise floor is at -6 dB and the headroom is only 6 dB. This means that the difference between the quietest and loudest sound presented to the listener is only 12 dB.

Summary The differences between a cinema and an in-flight-entertainment system are the most extreme. The biggest difference is that a cinema is created to watch movies and in an airplane watching a movie is a feature. This is very clear when the dynamic footprints are compared. The total dynamic range of a cinema should be 61 dB where a listener in an airplane has the optimal experience when the total dynamic range is 12 dB. This is a difference of 49 dB, which is a lot. Comparing these environments shows that optimizing the audio according to the site of consumption is needed.

An example of using content in the wrong environment is a badly transformed movie with a dynamic footprint for cinema scaled to a video format for iPod without taking care of desired dynamic footprints in the audio. Normally a movie consists of very wide dynamic range. Take for example soft parts in a dialog between actors and very loud parts during action scenes. Figure 2.3 shows that the noise floor of an iPod is much higher than the noise floor inside a cinema. This means that the described environmental noise in the iPod context will mask a big part of the sound with the cinema profile. However, when a loud action scene is presented

with the dynamics of the cinema to a content consumer using an iPod, the audio level is too high and can even cause hearing damage. The solution is to decrease the preferred average area and headroom to make the content suitable for iPods.

3 Loudness

Brittanica Encyclopedia: *Loudness, in acoustics, attribute of sound that determines the intensity of auditory sensation produced.* [4]

In the beginning of the 1980's the first, legal, commercial radio stations were appearing next to European public radio stations. These stations earn money by selling advertisement slots. This roughly means; more listeners; higher revenues. A typical method to keep and to attract new listeners is to make the radio station sound louder. Since commercial broadcasters were not bound to follow EBU recommendations, these recommendations are more or less strict guidelines for public broadcasters. Commercial broadcasters do not have to keep the 9 dB headroom free for peaks as long as they stay within their space of the frequency spectrum. The commercial broadcasters were shifting their alignment level a couple of dB's up into the headroom space to make the radio station sound louder. This was the beginning of the so called 'loudness war', louder is better. Undermining the creative purpose of maintaining a dynamic range to make money took not only over audio processing for radio, but commercial TV took over this habit followed by public broadcasters lacking EBU regulations.

A new way of treating loudness in broadcasting was introduced. This chapter introduces to reader to the term loudness.

3.1 Introduction on loudness interpretation

Loudness is in general a subjective expression for how loud a sound sounds. Each and every person has another perception on about how loud a sound is. Various factors can influence this. Examples are age, hearing capabilities, listening environment

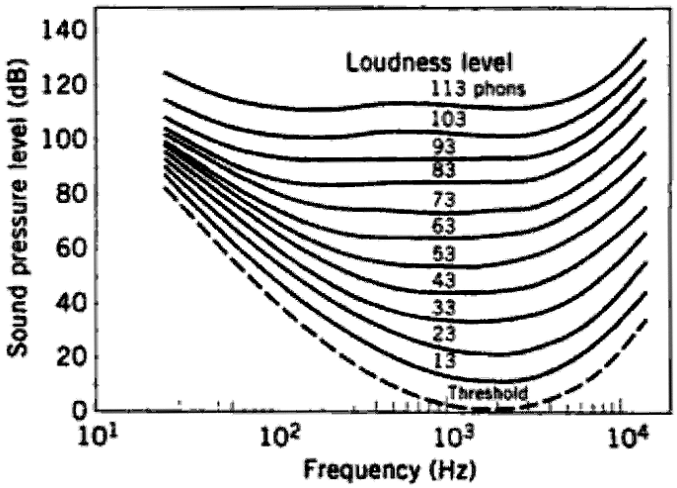


Figure 3.1: Equal-loudness contours for pure tones presented through earphones [5]

and the perceived sound itself. It is known that the hearing curve of a person is not linear in frequency vs. sensitivity. The louder the presented sound pressure level (SPL) gets, the more straight the curve will become. Figure 3.1 shows the relation, loudness level in phones, between a single frequency tone and various sound pressure levels. For example a 100 Hz tone with a SPL of 40 dB is perceived equally loud as a 1 kHz tone with a SPL of 20 dB.

As mentioned before loudness is a subjective expression of an auditorial sensation. In order to do objective measurements, a loudness measurement model and a unit are needed. The ITU-R BS.1770-1 Recommendation [6] provides an algorithm to do this. This algorithm translates an integrated audio signal with a certain level in dB on a nominal full scale into a loudness level in LKFS. The LKFS unit expresses values in decibel (dB). The used algorithm can be applied for mono, stereo and multi channel audio signals.

$$z_i = \frac{1}{T} \int_0^T y_i^2 dt \tag{3.1}$$

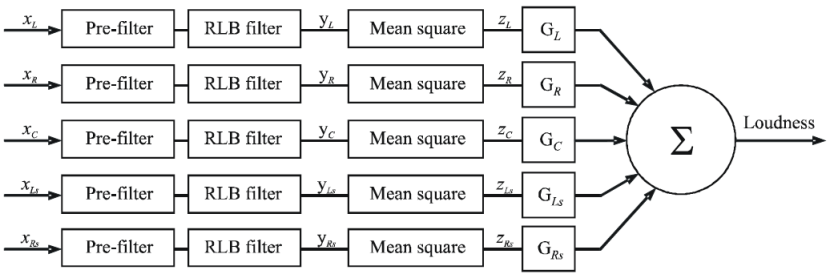


Figure 3.2: Block diagram of multichannel loudness algorithm [6]

This thesis introduces the BS.1770-1 recommendation in a nutshell. Figure 3.2 shows the signal flow path of the algorithm.

The signal flow path shows x_N as input and a loudness level as output with four processing stages in between. The first two stages are applying a pre-filter that represents the response of the acoustic effects of a head. The exact parameters of these filters can be found in [6]. The third stage is a mean-square measurement in the interval T as 3.1 [6]. $(i = L, R, C, Ls, Rs, N)$.

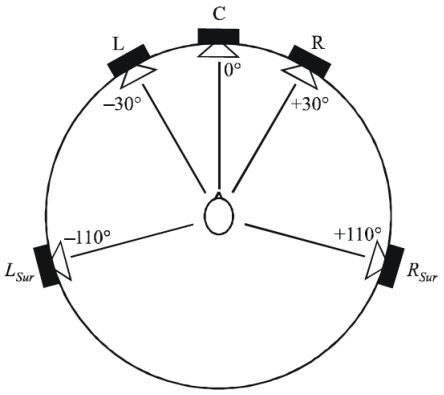


Figure 3.3: Loudspeaker configuration in surround setting [6]

Stage four is to sum the measured channels. Since sound sources in front of a person are perceived less loud than sound sources behind a person, signals for surround speakers must be weighted differently. Table 3.1 shows the weights and equation (3.2) [6] shows the final processing stage. Figure 3.3shows the loudspeaker setup in

a 5.x surround setting. The five stands for the number of normal loudspeakers and the x for the number of sub-woofers. the number of sub-woofers is irrelevant for calculating loudness values.

$$Loudness = -0.691 + 10 \log_{10} \sum_{i}^{N} G_i z_i \qquad \text{LKFS} \tag{3.2}$$

Channel	Weighting, G_i
Left (G_L)	1.0 (0 dB)
Right (G_R)	1.0 (0 dB)
Centre (G_C)	1.0 (0 dB)
Left Surround (G_{Ls})	1.41 (~ +1.5 dB)
Right Surround (G_{Rs})	1.41 (~ +1.5 dB)

Table 3.1: Weightings for the individual audio channels [6]

3.2 Loudness treatment in broadcasting

Broadcasters can mainly be divided into two groups, radio and television. They have both a different way and art of broadcasting. With television broadcasting the focus lies on video plus 'a bit of audio', radio brings information over audio only. The old fashioned, low quality approach of audio treatment in television broadcasting is converging towards the higher level of radio broadcasters. An important factor for this change is that end users are having more and more excellent audio equipment next to their high quality flat screens. Figure 3.4 – left shows an equipment rack of a broadcaster. This broadcaster uses "Orban Optimod" hardware for audio processing. What is shown is that national radio and television stations are processed by the same hardware. This means that at least this company sees no differences in the urge of maintaining a high audio quality. Figure 3.4 – right shows monitoring and metering equipment for audio and video.

In this setup it is possible for a broadcast engineer to monitor content at several stages in a broadcast chain.

Figure 3.4: Audio processing and monitoring in a broadcast environment

An increasing amount of people is annoyed by commercial breaks that are very loud in comparison with the content in between these breaks. The same counts for enormous loudness differences between sudden loud action-scenes in movies and quiet dialogs. Also the difference in loudness between TV channels while zapping is too big according to these complaints. To deal with this problem television broadcasters are uniting to find standards for loudness treatment. The biggest and most active taskforce is hosted by the European Broadcast Union (EBU) and is called PLOUD. This group represents broadcasters, content producers, researchers and equipment manufacturers. The group was founded in 2008.

One of the main goals of this group is to produce a recommendation that defines a solution in proper loudness handling. In August 2010 the PLOUD group published

the EBU R128 recommendation [7] together with a metering specification: EBU Tech 3341 [8] and a loudness range descriptor: EBU Tech 3342 [6]. After publishing [7], the EBU released a logo that can be used for branding equipment. This tells broadcasters who are buying new, metering, equipment that a device is EBU R128 compliant. The logo is presented Figure 3.5.

Figure 3.5: EBU R128 logo for branding on equipment and productions

The difference between the conventional approach and the new approach of metering audio levels, ITU-R BS.1770-1 and EBU R128 recommendation, is that the new method expresses itself in the Loudness K-weighted Full Scale (LKFS) unit and not the signal amplitude in dBFS. Note that ITU-R BS.1770-1 uses the LKFS unit and EBU R128 LUFS (Loudness Unit Full Scale) for absolute values. The values must be interpreted equally [9]. EBU R128 uses LU (Loudness Unit) for relative loudness values. Levels are expressed in how consumers experience content and not how, for example, good or bad a transmitter can stay within a dynamic range of telecom standards for ideal transmission. User experience is key in this new approach.

3.3 EBU R128 in a nutshell

The EBU R128 recommendation states that the following descriptors should be used to characterize an audio signal [9]:

- Program Loudness
- Loudness Range
- Maximum True Peak Level

What the PLOUD group is trying to reach with the EBU R128 is that all programs are normalized to the same loudness level. This will bring generally the following features:

- All TV channels are equally loud
- No sudden loudness changes when a commercial break starts
- Better programme peak treatment

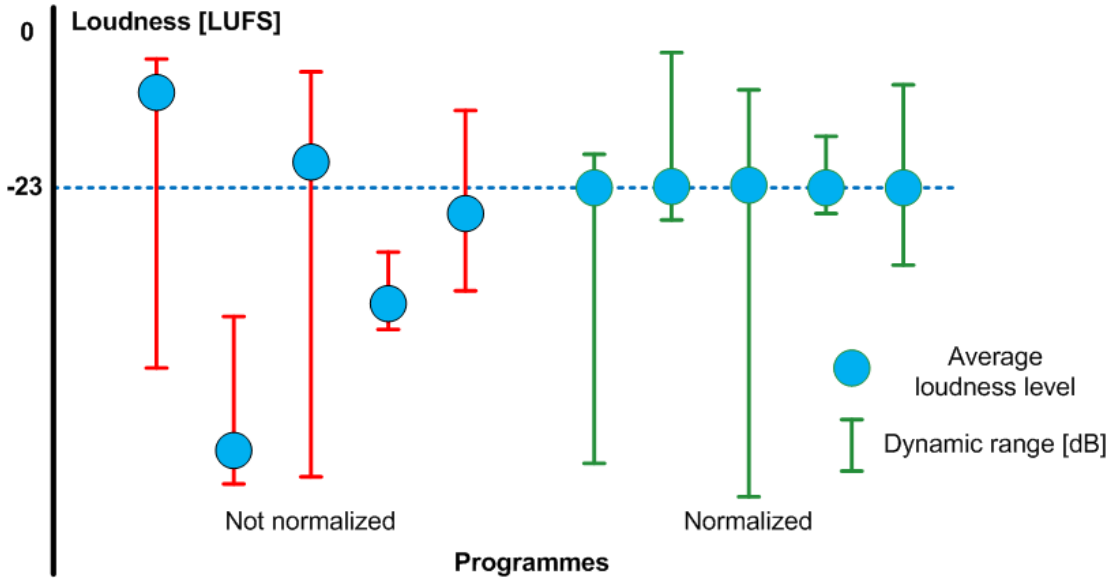


Figure 3.6: Loudness normalization

These features are reached by implementing the ITU-R BS.1770-1 recommendation for loudness measurement and set the average loudness level of a programme to a recommended value. The EBU R128 recommendation states that this value is −23 LUFS with a maximum deviation of ±1 LU. This means that the average loudness of a programme must be −23 LUFS. Figure 3.6 shows result of loudness normalization. This seems easy to implement, but there some exceptions where have been taken care of in the R128 recommendation.

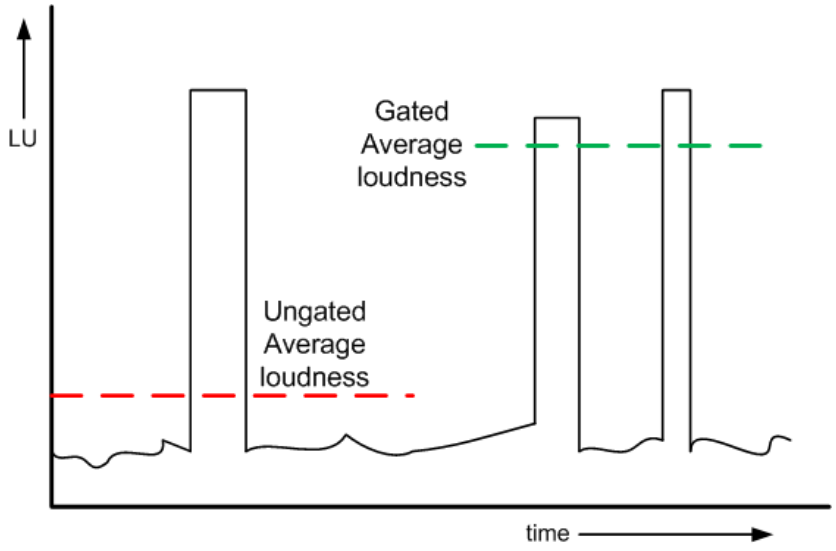


Figure 3.7: Difference between gated and ungated measurement

The R128 recommendation describes foreground and background audio. When there is for example a tennis match with a silent audience, there is a low average loudness level. However if there is a cheering audience when a tennis player gained some points, the loudness level increases rapidly for a short time until the game goes on and the audience will be quiet again. The total average loudness level of the program will be very low. This has the consequence that to reach the recommended average of −23 LUFS the gain will be increased so much that the loudness level will be unacceptably high during the parts where the audience is cheering. The loudness range will be very big in this case. To solve this problem,

the R128 recommendation implements a measurement gate. This means that the measurement for the program loudness pauses when the loudness level is 8 LU lower than the aimed average for longer than 400 ms. Figure 3.7 shows in a graph the difference in average loudness levels that are gated and not-gated.

The gating option is recommended to apply with recorded programs. However during, for example, live broadcasts it is not possible to know what the loudness average of the total program will become. For this reason different integration times are implemented in the EBU R128 recommendation. Table 3.2 lists these integration times. It is recommended for a live production to use the *momentary* and *short-term* modes.

Momentary	400ms (ungated)
Short-term	3s (ungated)
Integrated	start/stop (integrated M, gated)

Table 3.2: Defined integration times in EBU R128 [7]

When a sound engineer is working on a live production, the sound engineer can use loudness meters that are "EBU Mode" enabled to correct the levels according to the R128 recommendation. Practical usage solutions will be published soon by the EBU PLOUD group. When a broadcaster wants to use content from an archive, the broadcaster should upgrade the stored content towards the R128 recommendation before use.

3.4 Loudness treatment in mobile applications

At this moment for loudness treatment good solutions are brought up and manufactorers are producing equipment that helps to implement the R128 recommendation in broadcast chains. This is of course a very good development, but the status for mobile content is different.

The content that can be watched on a mobile device can be divided in a way as showed in Figure 3.8. The last, media files, can be found in two groups: professionally and user-generated. An example of professionally generated media is a movie from the iTunes store and user generated media could a video clip made on a mobile phone and consumed from e.g. YouTube.

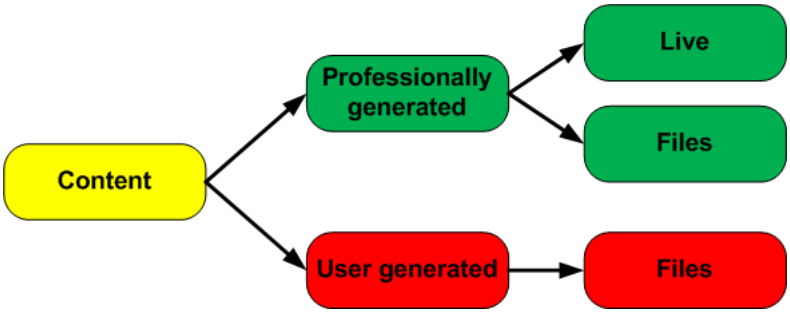


Figure 3.8: Types of content sources

It is logical that professionals should process the audio in professionally generated content for mobile consumption. However, this is not the case for user-generated content. This can only be done at the consumer side but no implementation has been found in a survey. It is however possible that online archives for user generated content, e.g. YouTube, pre processes the content before making it available for mobile devices.

After a survey on a conference for broadcasters, International Broadcast Convention 2010, it turned out that mobile content is not pre processed for mobile usage at all. In the best case EBU members will apply the R128 recommendation. What can be said is that for mobile content usage there is no recommendation or broadcaster that cares about audio treatment for mobile content usage at all.

4 Increasing QoE by changing dynamic range

In Chapter 3 it is explained that increasing the power density of an audio signal, especially on upper side of the spectrum, does increase the generated sensation of perceiving sound. The EBU R128 recommendation describes a certain loudness average and loudness distribution for broadcast programmes based on the ITU-R BS.1770-1 [6] loudness measurement method. What Figure 2.3 shows is that every environment has a specific preferred dynamic range. In the context of mobile content consumption the iPod profile seems most relevant. But since an iPod can be used everywhere, it is most unlikely that only one dynamic range is enough. In this chapter the impact of decreasing dynamic ranges on the Quality of Experience (QoE) is presented. The impact is measured with an experiment. Also methods to alter dynamic ranges are presented.

4.1 Methods to alter audio dynamics

Altering dynamic ranges of audio is a way of signal processing. There are several methods to do this. The majority of changes in dynamical characteristics are done by compressing signals into a different dynamic range. The two most important parameters having an effect in this process are the amplitude and frequency spectrum of a signal. The amplitude of a signal represents of course the dynamical behavior, but since an audio signal contains, most of the times, several tones with its own frequencies and dynamic behavior. As can be seen in Figure 3.1 and can be read in [6] the loudness curve is not straight but has a certain characteristic. This is why also the frequency spectrum of a signal has been taken into account

with compression of signals. Figure 4.1 shows a high-end single band compressor. This compressor is used at many professional locations, e.g. recording and radio studio's. A special feature of this compressor is that reaching a certain audio level at the input cannot only trigger the compression function, but this compressor has also so called side-chain inputs. A side-chain input listens at certain properties in a signal; e.g. start compressing when a signal reaches a level at a certain frequency.

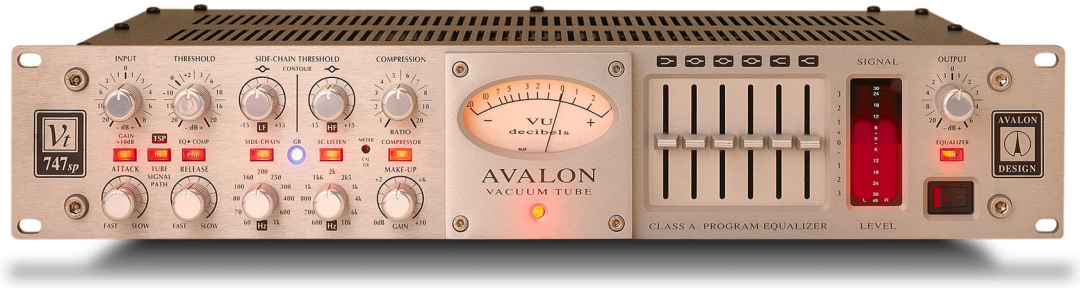


Figure 4.1: Single band compressor in hardware

4.1.1 Single band compressor

What a compressor basically does is decreasing the dynamic range of a signal. A single band compressor has generally the following parameters:

- Threshold
- Make-up gain
- Ratio
- Attack time
- Release time

Figure 4.2 shows what a compressor does with a signal in the following example. For example: a compressor has on its input an audio signal with a variable amplitude

from -90 dB till 0 dB. This means that the signal has a dynamic range of 90 dB. Given is that the headroom is 30 dB. When for some reason the dynamic range must be decreased till 75 dB and the headroom may be decreased to do this, a compressor can be used. Table 4.1 shows the parameter settings to be set on the compressor.

Parameter	Value
Threshold	-30 dB
Make-up gain	15 dB
Ratio	1:2
Attack time	fast
Release time	fast

Table 4.1: Compressor parameter settings in example

The *threshold* for the compressor lies here at -30 dB. This means that when the input level is higher than this threshold, the compressor decreases the steepness of the increasing slope by half when a ratio of 1:2 is used. Without make-up amplification the maximum output is -15 dB. But when the maximum dynamical space must be used, the compressed signal must be amplified to reach the level of 0 dB again. The *attack time* is set fast to let the compressor immediately attack on the signal when the threshold has been exceeded. The *release time* has been set to slow to let signals be as much unharmed as possible. The settings of the attack and release time are depending on the application of the compressor.

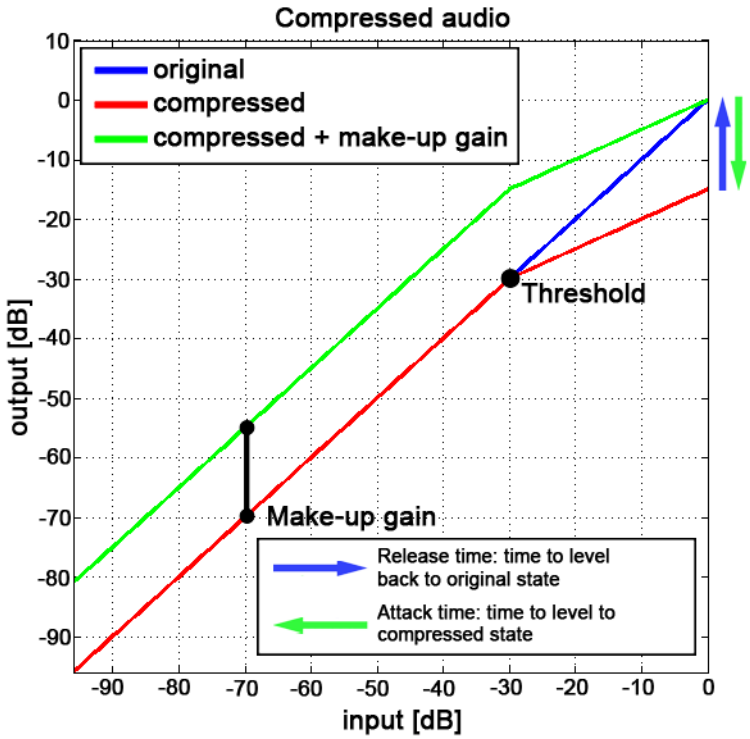


Figure 4.2: Single band compressor behavior

4.1.2 Multi band compressor

A multi band compressor does basically the same as a single band compressor; the difference is that a multi band compressor can process on defined frequency bands. A multi band compressor can be very useful when for example the lower frequencies are very dominant in a signal and must be flattened. It can also be used to meet the R128 recommendation since the loudness curve is not linear. Figure 4.3 shows a software version of a simple multiband compressor. Three different frequency bands are specified with each their own compression settings.

4.2 Experiment with decreasing headroom

Since the given iPod dynamic profile [3] specifies a dynamic range for just a single device, it is not given for which environment this is valid. It can be vital for the QoE that an audio signal must be presented with the right intensity to be perceived well. Normally increasing the volume on the mobile device does this, but the question is whether this is necessary or not. By increasing the volume the whole signal gets more intense, from the average signal level, but also the peaks in the headroom. A consumer may find it needed to increase the volume level that much that hearing damage can occur.

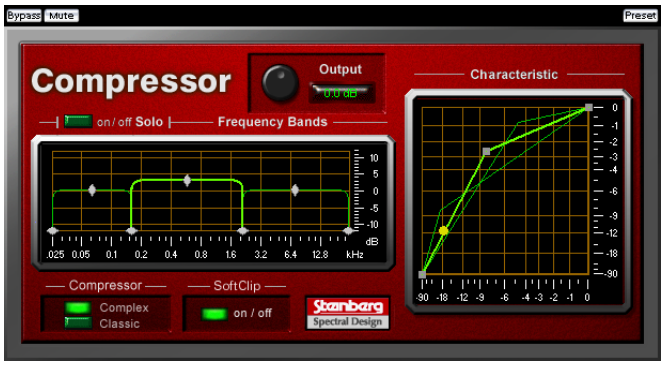


Figure 4.3: Software version of a multi band compressor

Figure 4.4 shows two signals, one without correction for iPod and one with correction for iPod. The red area represents the environmental noise, a noise masker, with a dynamic profile in between. What can be derived from this figure is that the majority of the signal without correction is swallowed up by the environmental noise. Only some of the signal peaks are loud enough to overcome the environmental noise. The signal that has been processed for the iPod has become loud enough to lift the whole headroom and a big part of the average of the signal above the noise masker.

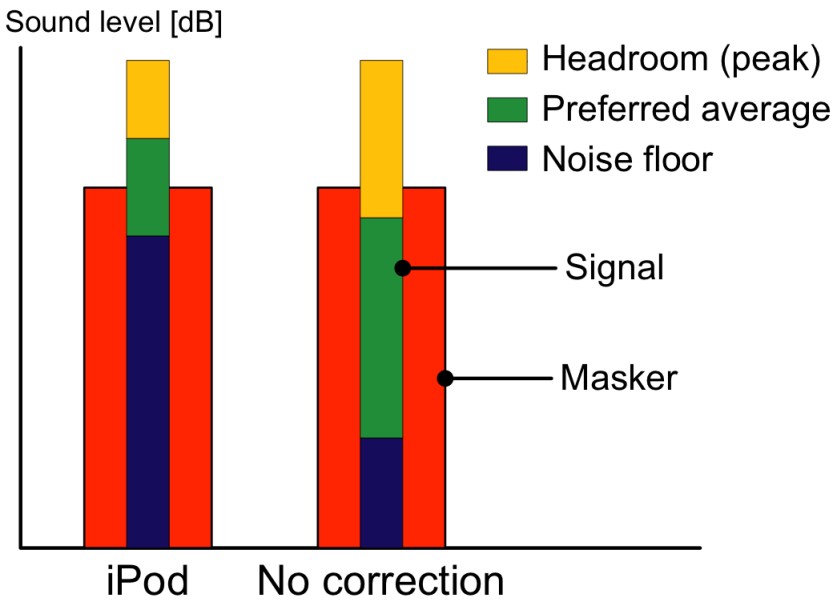


Figure 4.4: Decreasing dynamic range vs. auditory masking

What the impact on the Quality of Experience (QoE) of this way of processing is in relation to environmental noise has been measured with an experiment. This experiment description and results are described in this section.

4.2.1 Experimental design

The design of the experiment is based on the ITU-R BT.500-10 recommendation [10]. During this experiment test subjects were exposed to environmental noise and test videos with processed and unprocessed audio. The content was presented on an iPod with the standard Apple iPod earphones. Environmental noise was presented on closed headphones, which are covering the ears plus earphones completely.

The test content that has been used were captured television programs. The content is listed in Table 4.2.

The content has been presented in three variants, unprocessed, processed for a

living room and processed for an iPod. The used dynamic profiles are shown in Figure 4.5 and Table 4.4, which are derived from [3].

Content type	Name of programme
Television show	James May's 20th Century
Music video	David Gilmour live at Gdansk - High Hopes
Sports	Australian Open - Tennis

Table 4.2: Used content in experiment

What can be observed is that the dynamic range for the iPod is half the size as the dynamic range for a living room. Expected is that the score of the processed material for iPod will be higher than the unprocessed content and the content that has been processed for a living room. Assumed is that broadcast material is normally produced for consumption in a living room, non the less, all content has been reprocessed for this experiment.

Noise	processing type	TV show	music	sport
No noise	No processing	-28	-27	-27
	Living room	-15	-14	-14
	iPod	-9	-8	-10
Street noise	No processing	-30	-24	-30
	Living room	-14	-14	-15
	iPod	-8	-8	-8
Bus ride	No processing	-28	-16	-28
	Living room	-15	-15	-16
	iPod	-9	-10	-11

Table 4.3: RMS levels [dB] of content in the experiment

The processing of the content has been done with software. First the audio was separated from the video into different files. The audio files were loaded into *Steinberg WaveLab*[1] and processed with software compressors which are plug-ins for this software until the desired values were met. After the processing, the video and audio files were combined. The exact RMS levels of the processed content per noise type are given in Table 4.3.

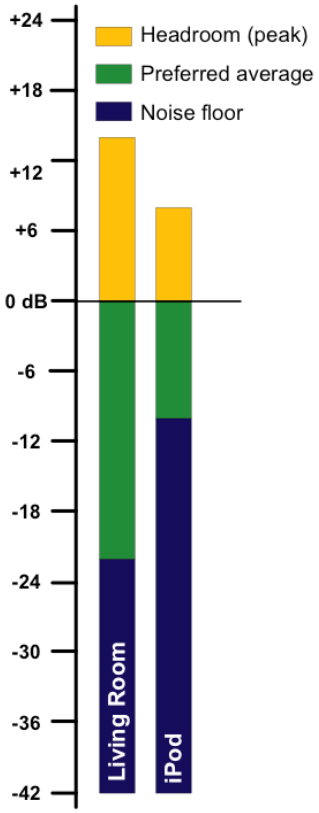


Figure 4.5: Dynamic profiles used for experiment [3]

Pre-recorded noise recordings were used in this experiment. For recordings a stereophonic microphone (Rode NT-4 x-y) and a Tascam audio interface were used to ensure high quality recordings. Figure 4.6 documents our noise registration procedure.[1] Important to mention is that the noise recordings were normalized.

[1] *http://www.steinberg.net*

This means that on a full scale (dBFS) the peak of the noise lies on 0 dB. The test person could not change the volume setting of the environmental noise. The environmental noise that has been used is listed in Table 4.5. Expected is that the impact of the noise *type at a bus stop* has a higher impact on the presented content than noise type *inside a moving bus*.

	Living room	iPod
Noise floor	-22 dB	-10 dB
Preferred average range	22 dB	10 dB
Headroom	14 dB	8 dB
Dynamic range	36 dB	18 dB

Table 4.4: Dynamic profiles used for experiment [3]

The experiment consisted of three parts: a training period and two periods where test subjects had to rate the content. The training period was necessary to let test subjects getting used to the experiment, equipment and sound level. During the test period the test users were listening to content without environmental noise. Test subjects were asked to set the volume level of the test content to their ideal volume level. This was done for two main reasons. First, the test person could never be exposed to dangerous sound pressure levels during the experiment and second, differences in hearing capabilities between test persons have been eliminated this way. The test content was dance music processed according to the iPod profile. After the training period the test subject was told to never change the volume setting again. The training period lasted for about 30-45 seconds.

Figure 4.6: Recording environmental noise

For rating the content the test subjects had to wear a special glove, which measured the position of the fingers. When the hand was open, the test subjects were rating maximal and closed had represented a minimal score. A maximum score means that the influence of the environmental noise on the content consuming experience was minimal. A minimum score means that the experience level was at such a low rate that presented content could not be perceived at all. For more details about this way of rating read [11].

Noise type	RMS Sound level in dBFS
No noise	no value
Inside a moving bus	-24 dBFS
At a bus stop	-19.5 dBFS

Table 4.5: Used types of environmental noise

Before the experiment took place, the test subjects had to chose two out of three content types as described in Table 4.2. After the test subject chose a content type, the rating started. Figure 4.7 shows the order of content presentation together with environmental noise.

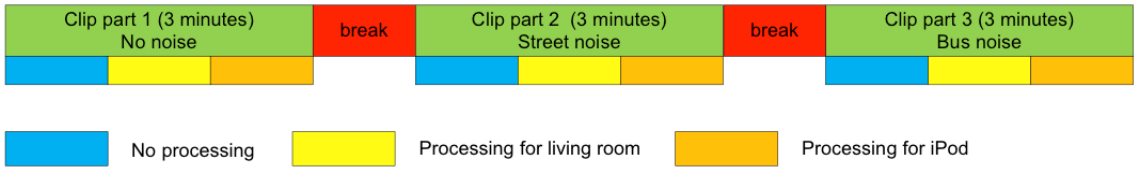


Figure 4.7: Presentation order of content during the experiment

4.2.2 Experiment results

The experiment generates of every test subject a file with the position of the test subject's fingers. The measurement was continiously over time like presented in [11]. For the television show content 14 average rating have been collected, for the sport content 13 and for the music show again 14. For the statistical software R has been used for evaluation[2]. The results of the collected rated content averages were divided into three groups: no noise, street noise, bus ride noise. These groups were evaluated with an ANOVA test. For this test a null hypothesis was formulated:

[2]http://www.r-project.org

Noise type	Result	p-value
No noise	The correction method has no effect on the quality rating	0.6407
Street noise	A difference between the correction method was heard	2.242e-08
Bus ride noise	A difference between the correction method was heard	9.76e-14

Table 4.6: Results of ANOVA test on correction type vs. environmental noise [1]

"The correction types (none, living room, iPod) do not influence the user ratings". The results of this evaluation are given in Table 4.6.

Furthermore, it has been investigated by another ANOVA test whether the result depended on the content type and no indication for a dependency could be found. Hence, it cannot be stated that the content type influenced the result. This result implies that the scores of different content types can be taken together for further analysis. [11]

The behaviour of the Mean Opinion Scores by the test persons can be studied in Figure 4.8. It depicts the average MOS values and the related confidence intervals over all content types for each background noise and processing type. When observing the results obtained for no background noise it can be noticed that quality ratings are highly independent from the correction method that has been applied. Hence, the headroom decrease does not affect perceived quality in optimal conditions for the selected type of content. During a "Bus Ride" instead the quality is seriously deteriorated due to background noise.

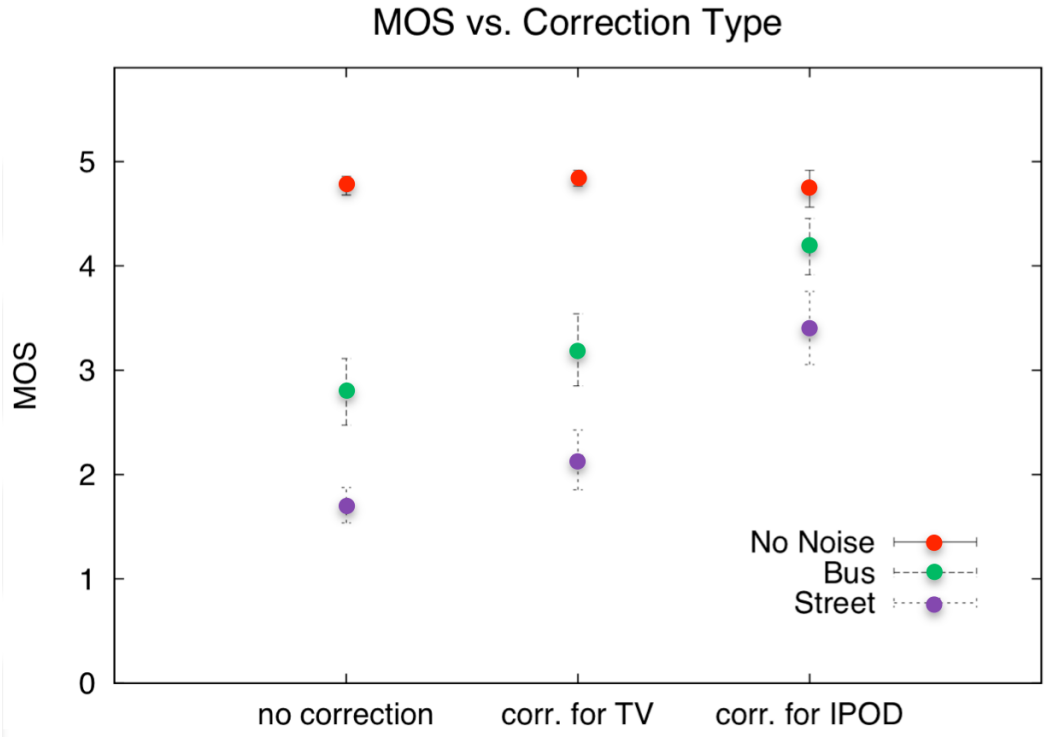


Figure 4.8: Rating averages of dynamic profiles vs. MOS score [11]

The mean MOS decreases from 4.8 to 2.8. The effect is even more accentuated for "Street Noise" where the MOS has decreased down to a value of 1.7. These values are far from being acceptable. In both cases the correction method "Living room" helped to increase the MOS but not sufficiently enough as to provide acceptable quality. When adopting the correction method "iPod" the MOS could be sufficiently increased for both types of background noise. In case of "Street Noise" the MOS has increased to 3.4 and in case of "Bus Ride" a value of 4.2 could be reached. To summarize, it can be stated that the approach of decreasing the headroom reveals to be very effective for avoiding quality loss caused by environmental noise. [11]

5 Changing loudness and adapting environmental noise properties

In the past chapters it has been made clear that the influence of non-processed content for a certain environment on the quality of experience is really big. The old approach of measuring levels has a negative influence on how users perceive audio in different environments. Really big annoyance factors for (mobile) content consumers are big sudden changes in audio levels and environmental noise. When signal levels were measured and put into a preferred dynamic range, the characteristics of the human ear were not taken into account. This could lead to a bigger difference in perceived levels than measured levels. To tackle this problem, ITU-R came up with a new method [6] of measuring audio signals. The outcome of this measurement was a level in LKFS, still on the logarithmic scale. This unit was changed in a later moment, on request by the EBU, to the Loudness Unit Full Scale (LUFS). Since there were a lot of complaints of content consumers regarding differences in loudness levels while watching TV, the EBU started research to solve this problem. The European Broadcast Union (EBU) founded the P/Loud group, which had the following tasks:

- Find an average loudness level for signals that would fit in the desired level by consumers, but this level should also fit within technical requirements.
- Develop a new workflow in production environments for live and recorded content (archives).
- Let manufacturers develop new loudness meters.
- And the crucial task, let broadcasters, producers and device vendors adapt this new way of production.

The EBU released a set of documents, which contain the EBU R128 recommendation. The first three tasks are covered by this recommendation since the members of the P/Loud group were broadcasters, meter manufacturers and production houses. This leads to meeting the fourth task; a wide adaption potential of the R128 recommendation. Now the part of a constant loudness range is covered, this range and level should be adapted to the environment where the content consumer is currently in. Research shows [3] that every type of environment requires different dynamic range values. The main reason that a different dynamic range is needed is due to environmental noise. Environmental noise can mask presented content, in other words: consumers cannot hear the presented content the way they want it. Noise can mask the content partially or totally. Figure 4.4 shows this. To meet the required dynamic footprint, the content needs to be processed. The easiest way to do this is by using a compressor. A compressor decreases the dynamic range into a desired footprint. This method is widely used to do this. But since the hearing and content characteristics are not linear behavior, are single (frequency) band compressor does not often meet the desired result. That is why there are multi band compressors. These compressors can process multiple frequency bands at the same time. When the end result meets the desired loudness range this might still not be the desired end result to meet the optimal QoE. This is the starting point of the experiments and results that are presented in this chapter.

5.1 Introduction to an equalizer

Spectral properties of a signal can be altered with an equalizer. What an equalizer does is nothing but changing the amplitude of a signal at given spans of the spectrum.

An equalizer is an array of filters. These filters are having the following parameters:

- Operating frequency
- Quality factor (steepness)
- Amplification
- Band pass / Low pass / High pass

Figure 5.1 shows two different examples of frequency response characteristics that an equalizer can have. The properties of these characteristics can be studied in Table 5.1.

Equalizer characteristic 1 (eq1)

Span #	Frequency span [Hz]	A_{start} [dB]	A_{stop} [dB]
1	1 Hz -10 Hz	+10 dB	+ 10 dB
2	10 Hz -100 Hz	+10 dB	0 dB
3	100 Hz -1 KHz	0 dB	0 dB
4	1 KHz - 10 KHz	0 dB	-10 dB
5	10 KHz - 20 KHz	-10 dB	-10 dB

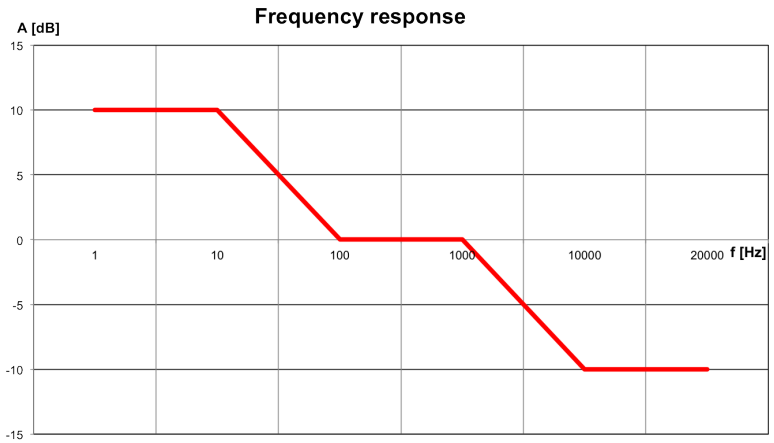
Equalizer characteristic 2 (eq2)

Span #	Frequency span [Hz]	A_{start} [dB]	A_{stop} [dB]
1	1 Hz -10 Hz	+0 dB	+ 10 dB
2	10 Hz -100 Hz	+10 dB	0 dB
3	100 Hz -500 KHz	0 dB	0 dB
4	500 KHz - 1 KHz	0 dB	-8 dB
5	1 KHz - 11.5 KHz	-8 dB	+10 dB
6	11.5 KHz - 20 KHz	+10 dB	0 dB

Table 5.1: Equalizer characteristics

The properties of Table 5.1 and Figure 5.1 is a frequency span in [Hz] and a signal attenuation A in [dB]. It means when a signal on the input of Equalizer 1 (eq1) reaches the output, the signal is changed.

Equalizer characteristic 1 (eq1)



Equalizer characteristic 2 (eq2)

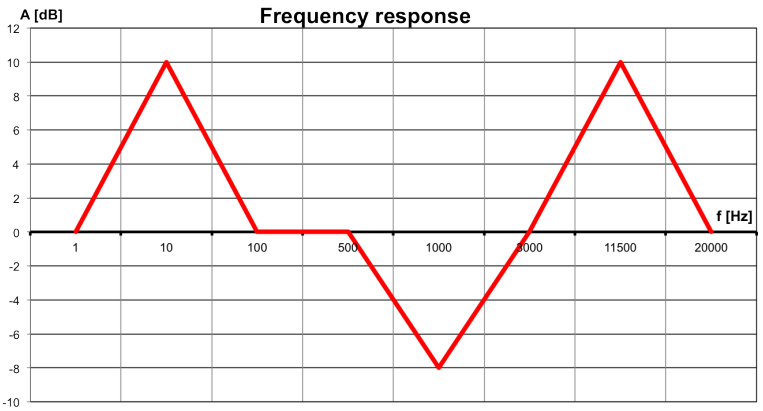


Figure 5.1: Equalizer characteristics Frequency [Hz] vs. Attenuation [dB]

The first three span numbers of (eq1) are described here:

1 In the frequency span from 1 Hz till 10 Hz, the input signal will be attenuated with +10 dB.

2 In the frequency span from 10 Hz till 100 Hz, the signal will be attenuated at the start of the span with +10 dB an at the end of the span, the signal will not be attenuated. Within the span there is a slope downwards with a steepness of -10 dB/decade. This means that every decade (e.g. 10 Hz - 100 Hz, 100 Hz - 1 KHz) the attenuation rate will decrease with 10 dB. In this frequency span the attenuation is +5 dB at 55 Hz.

3 In the frequency span from 100 Hz till 1 KHz, the input signal will remain the same. The attenuation is 0 dB in the whole frequency span.

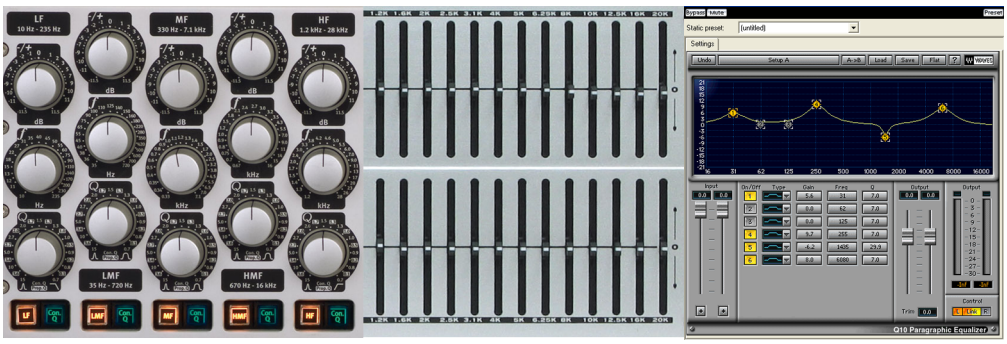


Figure 5.2: Hard and software equalizers

Figure 5.3 shows a schematic representation of an equalizer with a single filter. It is usual that an equalizer has several of these filter circuits in parallel to make it possible to alter several frequency spans simultaneously. First is determined what filter type must be applied. For filters that are applied in the frequency domain, the operating frequency is the value where a *hpf* or *lpf* has its main focus or where the center of the frequency span of a *bpf* is active. The steepness of the filters is defined by a so-called quality factor, or Q-factor. Equation (5.1) shows how to calculate a Q-factor. f_0 is the resonance frequency of the filter cirquit and Δf is the frequency space between the two points where the amplitude has been filtered

for 50% by a bpf. These points are 6 dB lower in intensity than f_0. The smaller the Q-factor is, the steeper the filter slope is.

$$Q = \frac{f_0}{\triangle\ f} \tag{5.1}$$

The intensity of the filters influence on the total signal can be adjusted with an amplifier.

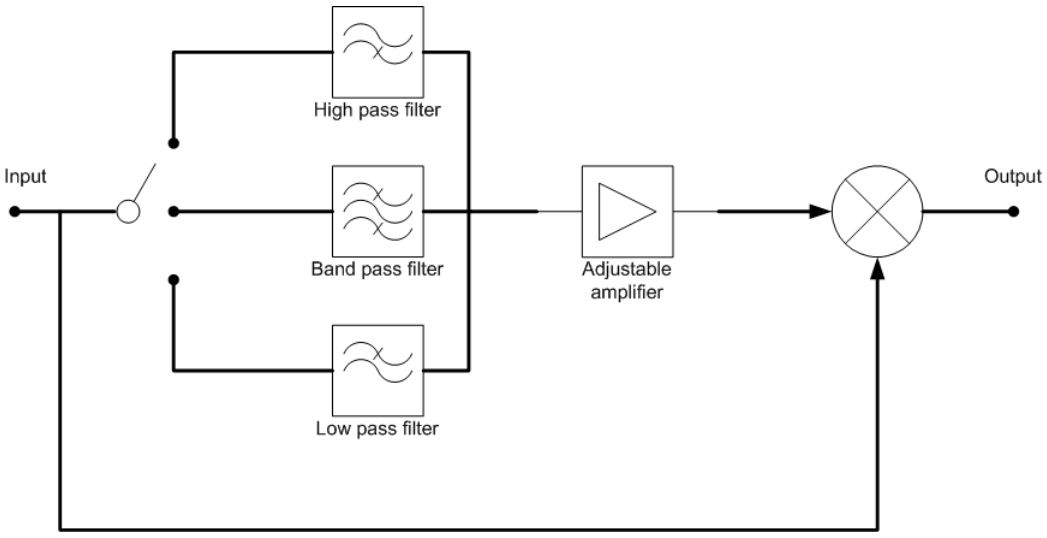


Figure 5.3: Schematic of an equalizer with a single filter

Implementations of equalizers can be done in hardware and software. There are two main types of equalizers: parametrical and graphical equalizers. The parametric version has all possible equalizer parameters available to adjust and the graphic version has static frequency and Q-factor settings, only the amplification setting can be altered. Implementations in software can be done in e.g. Digital Signal Processors (DSP), plug-ins for media production software and as software in an mp3 player. In the experiment that is mentioned in this chapter, software plug-ins for a media production platform were used. Figure 5.2 shows three versions of equalizers. Displayed are a five-band parametric equalizer and a thirteen-band graphic equalizer in hardware and a software equalizer. The software equalizer in

this Figure 5.2 is a six-band parametric equalizer.

5.2 Adapting environmental noise properties

Auditory masking means that parts of the spectrum are masked by possibly unwanted tones (noise). When this happens, wanted content cannot be perceived in a desired way. To eliminate this, an equalizer can be used to overrule unwanted noise maskers when these can be described. To determine how to set the parameters of an equalizer environmental noise must be analyzed. For analyzing environmental noise, several noise types were recorded. The same noise recordings were used in earlier described experiments. To analyze noise Steinberg Wavelab was used. This software has the option to measure and map spectral properties over time done with Fast Fourier Transformations (FFT). From this analysis several sources of noise could be identified. In Figure 5.4 and 5.5 the FFT plots of the bus ride noise and street noise are displayed.

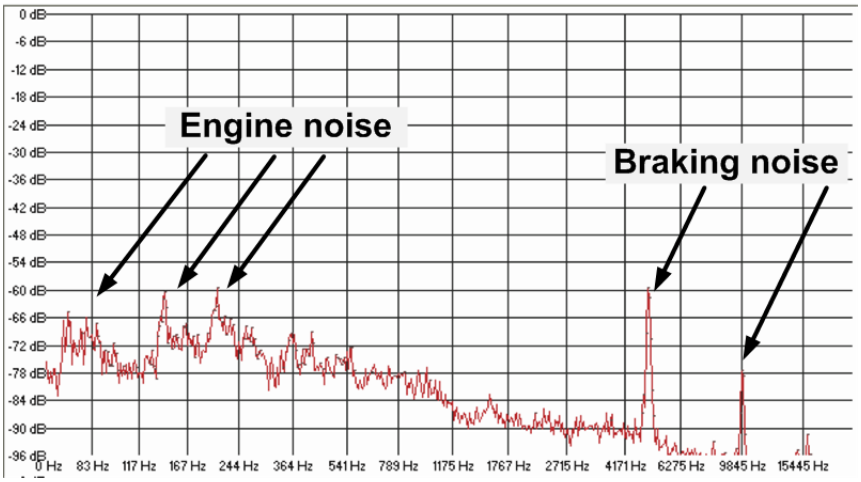


Figure 5.4: FFT plot of bus ride noise

The FFT plots of the bus ride are showing the noise sources that are produced in a bus very well. Two noise sources can be distinguished easily: the engine noise and braking noise. The engine noise is a constant noise that is shifting up and down in frequency during the ride. During the ride the bus is using its brakes to slow down, this braking produces a plosive noise burst in the higher regions of the spectrum. Also the engine noise affected, because while braking the throttle is released.

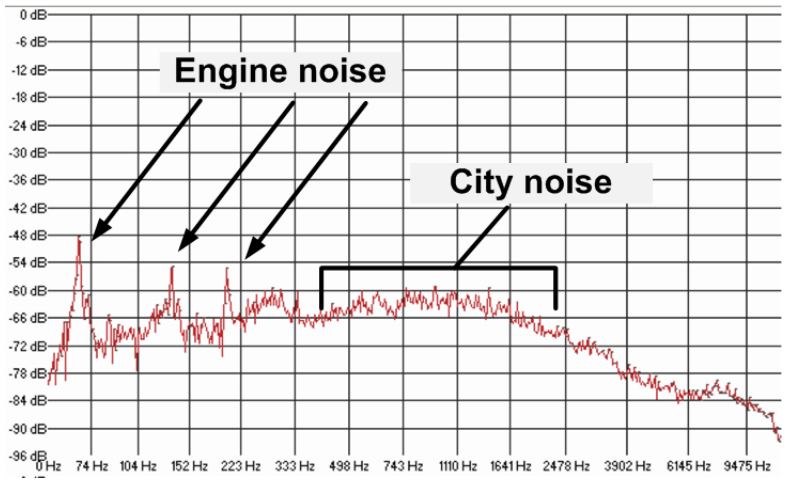


Figure 5.5: FFT plot of street noise

The noise at the street consists mainly of plosive noise bursts of passing vehicles and a continuous noise generated by the city. The noise of the vehicles has a very typical shape, three peaks in the spectrum and the city noise has a very broad continuous character. When the noise plots are compared, it can be observed that the levels of the noise at the street are higher then the levels of the noise in the bus.

Since the Quality of Experience (QoE) is decreasing by auditory masking, caused by environmental noise, overruling the noise masker is one probability. To overlap the environmental noise masker, an equalizer can be used to follow the spectral characteristics of this noise.

Figure 5.6 shows environmental noise from a recording together with the frequency response, in blue, of an equalizer that is set to follow the noise pattern. By doing this, the influence of the environmental noise should be decreased and should increase the QoE.

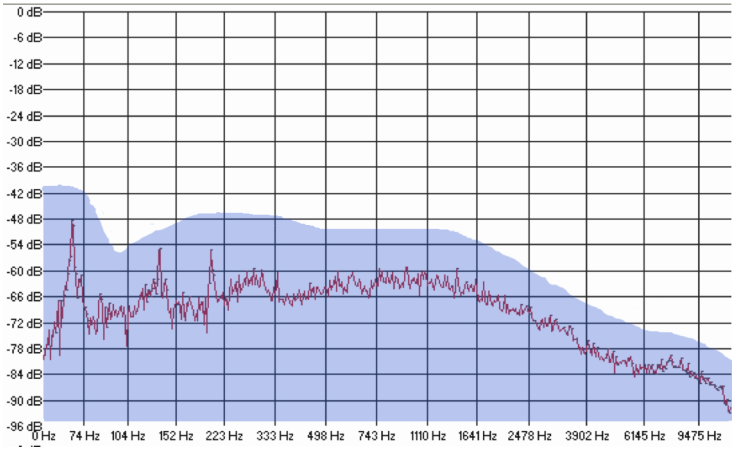


Figure 5.6: FFT plot street noise including frequency response of correction

5.3 Experiment

5.3.1 Experiment 1 - influence of environmental noise

First the influence of environmental noise on the overal perception of audio from a mobile device is tested in an experiment. During the experiments users had to listen to three noise types with a short break in between. During the experiment, test users were wearing two pairs of headphones. One pair of small ear phones were used for the content (pop music) presentation while a pair of big headphones on top of the small ear phones presented the environmental noise. While presenting the noise and the content, the sound pressure level (SPL) of the environmental noise is leveled to the same SPL as it originally was. The adaption time to the environmental noise and volume setting was between 30 and 45 seconds. 16 Test users were asked to adjust the volume setting of the iPod during the test until the

desired volume level was reached in respect to the environmental noise. Then the volume setting was noted and later, another noise type was presented. [1]

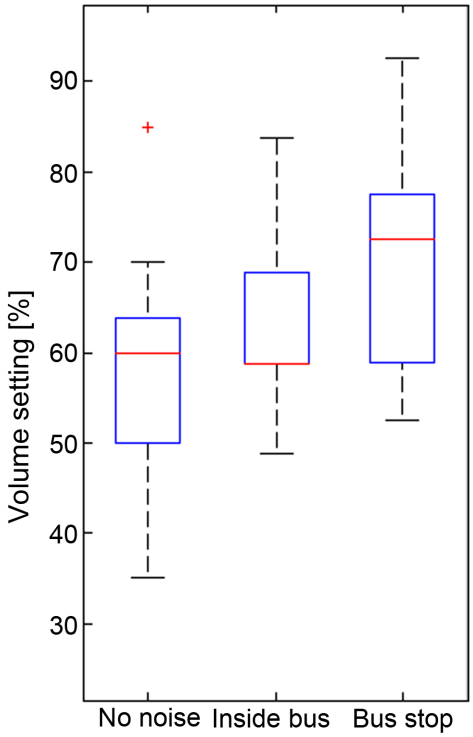


Figure 5.7: Location vs. Preferred volume setting

In Figure 5.7 the volume settings performed by all test persons is represented by a box plot. It can be observed that it is increased from 60% to 70% in average when some noise (in the bus) occurs and up to 75% when a typical bus stop noise is considered. Since desired loudness increases are considerable (up to 85% in average) it can be deduced that environmental noise has high impact on the users satisfaction.[1]

5.3.2 Auditory masking prevention

Environmental noise in general is not constant. Instead, it occurs in peaks. It might be a waste of effort to continuously decrease the headroom. Even if it seems that the perceived quality is not affected by decreasing the headroom there might be different content such as classical symphonies that would benefit from a larger headroom. Therefore, more sophisticated techniques to enhance the quality of experience of multimedia subject to environmental noise have been studied by performing a third subjective experiment.[1]

5.3.3 Methodology

Since the shape of the noise is highly relevant for the purpose of this experiment, the first step consists in analysing the recorded environmental noise. Once the shape of the sounds have been understood the content clips will be altered according to to the noise characteristics. For example multiband compression and equalization is applied to the portions of the content that will be subject to noise peaks. Then the content will be presented to test users while being exposed to environmental noise again by using two pairs of headphones as described in Chapter 4.[1]

5.3.4 Noise analysis

Fast Fourier Transformation (FFT) plots have been used to analyse the recorded noise samples. Two major noise types could be identified: plosive noise bursts and continuous noise. Both types can have a broad band and a small band characteristic. For example passing cars, car horns and car brakes are plosive bursts while continuous noise is created for example by engine noise inside a bus, talking people and city noise. For each of these two noise types one representative clip has been selected: “Street Noise” containing plosive noise bursts and “Bus Ride” for continuous noise.

Figures 5.4 and Figure 5.5 show FFT plots of “Bus Ride Noise” and “Street Ride” noise. The FFT analysis windows has a size of 262144 samples, with a sample

rate of 96 KHz of the recording, this means the plots are displaying a period of 2.731 seconds.

The plots are representing noises of different types. Figure 5.5 shows the shape of the noise of a passing car containing peaks at around 65 Hz, 130 Hz and 200 Hz. The window reaching from 330 Hz to approximately 2 KHz contains a constant background noise produced by the city. The whole noise profile has a broad band characteristic. Figure 5.4 shows a "cleaner" noise. The peaks around 60 Hz, 140 Hz and 200 Hz are produced by the engine of an urban bus. The peaks around 4800 Hz and 10 KHz are produced by its brakes. Between the two noise sources the sequence seems rather clean. In between the identified frequencies no environmental noise should affect the quality of the multimedia content presented by a mobile player. This could be the reason for lower volume setting increase for "Inside Bus" than "Bus Stop" in Paragraph 5.3.1 and the higher MOS decrease noticed for "Street Noise" in respect to "Bus Ride" presented in Paragraph 4.8.[1]

5.3.5 Content selection

To increase the QoE in both considered noise environments, two different approaches have been followed. Three filter types have been used for processing the content: linear dynamical compression (LC), multiband dynamical compression (MB) and equalizers (EQ). In Section Paragraph 4.8 it has been shown that QoE increases significantly when the headroom is decreased with a linear compressor. Hence, first the headroom is decreased and then frequency related filters are applied to the content. Again, test users had the opportunity to select two content types out of three categories (TV show, music, sport). The selected TV show contains an audio that is composed by the voice of a narrator and music. The music type consists of an excerpt of an opera – sport is represented by a Formula 1 race with a narrator and race car engine noises. With this choice most common content and audio types are covered.[1]

5.3.6 Content processing

In Section Paragraph 5.3.4 “Street Noise” and “Bus Ride” have been analyzed. It has been noticed that in “Street Noise” noise explosions occur during a frequency interval reaching from 65 Hz to 330 Hz. Therefore the content will processed as follows: (1) interval that contains all observed explosions [65 Hz, 500 Hz] and (2) the part not containing sound explosions [330 Hz, 2 KHz]. For “Bus Ride” the noise peak has been identified to be in the interval [4.9 KHz, 5.1 KHz]. Furthermore, the EQ positions 70, 140, 280, 600 Hz have been identified to be critical. In Table 5.2 the procession type is listed for each reason.[1]

Chunk	Street noise	Bus ride
no	no processing	no processing
14	LC to -14 dB RMS	LC to -14 dB RMS
	EQ 65 Hz - 500 Hz	EQ 70, 140, 280, 600 Hz
	MB 330 Hz - 2 KHz	MB 4.9 KHz - 5.1 KHz
9	LC to -9 dB RMS	LC to -9 dB RMS
	MB 65 Hz - 500 Hz	EQ 70, 140, 280, 600 Hz
	EQ 330 Hz - 2 KHz	MB 4.9 KHz - 5.1 KHz
14a	LC to -14 dB RMS	LC to -14 dB RMS
	MB 65 Hz - 500 Hz	MB 70 Hz - 600 Hz
	MB 330 Hz - 2 KHz	MB 4.9 KHz - 5.1 KHz
14b	LC to -14 dB RMS	LC to -14 dB RMS
	MB 65 Hz - 500 Hz	EQ 70, 140, 280, 600 Hz
	EQ 330 Hz - 2 KHz	MB 4.9 KHz - 5.1 KHz

Table 5.2: Processing properties for street and bus ride noise

5.3.7 Test procedure

Similar to previous experiments the test users were asked to determine the maximum SPL they wanted to be exposed to. The test clips were presented in the same way and order as the previous experiment. The only difference consists in the fact that presentations were always subject to background noise. The length of one clip was approximately 10 minutes.[1]

Filter	Band pass
Frequency range	start - end @ 3 dB
Q factor at single frequency	5
Attenuation	+6 dB

Table 5.3: Equalizing properties

Frequency range	start - end @ -3 dB
Ratio @ 0 dB	1:3
Threshold @ 0 dB	-14 dB
Make up gain	+4.66 dB
Knee	soft

Table 5.4: Compressor properties

5.3.8 Test results

For determining whether there are differences between the groups we carried out several versions of analysis of variance. For comparing each group with each other we chose a TukeyHSD and a pairwise t-test using the adjustment of Holm, both being robust against aggregating type-I errors. We also created the following complex contrasts: compare "no correction" with all other groups, compare 14dB

with the other enhancement groups, and compare 9dB with the EQ/MB versions of 14db. These comparisons were computed using both standard ANOVA and the method from Scheffe. We carried out the analysis for the three video types individually, and also for all videos together. For each video type, we computed results for both street and bus noise.

As general result the most significant differences are between no-enhancement, and any other group. These were found by all tests homogeneuously. For sport and music there was a also some weak difference found between 14dB and other enhancement types. In the all-videos case this difference is found strongly, i.e., when looking at all videos no matter what type, for both street and bus the -14 dB enhancement (chunk 14) is definitely rated worse than all other enhancement types. On the other hand, no difference whatsoever can be found between the -9 dB enhancement and the EQ/MB versions of -14 dB.

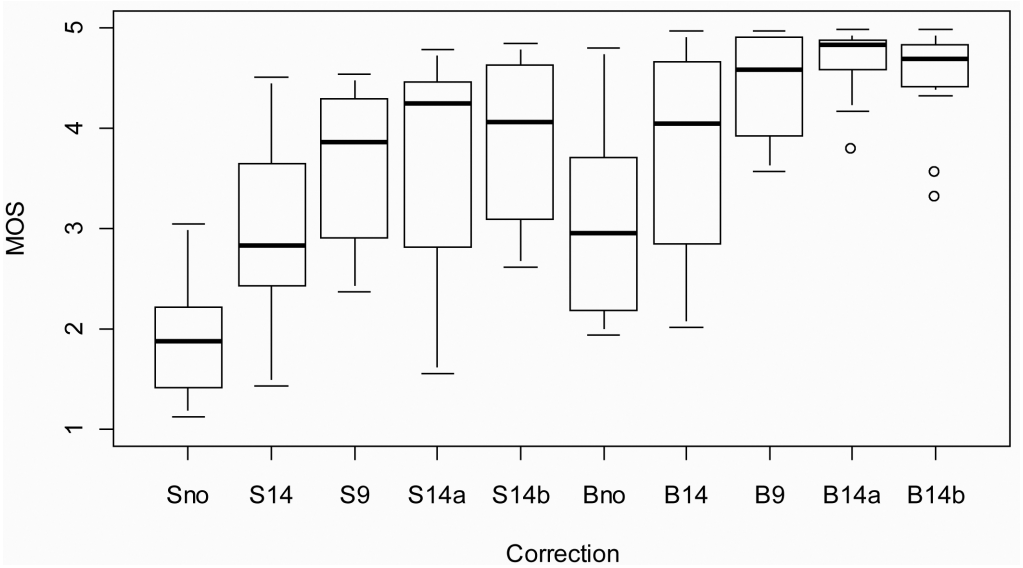


Figure 5.8: Correction methods vs. MOS-score

For the complex contrasts we find a split decision. The ANOVA always found a strong difference between “no-enhancement vs. all others”, and between “-14,dB and the other enhancements”. When comparing -9 dB to the EQ/MB versions of

-14 dB, always a weak indication was found. The Scheffe-based test only found differences in the first case, but not in the second and third case.

In Figure 5.3.8 the MOS values for all five processing types (no, 14, 9, 14a, 14b) for sports content with "street noise" (S) and "bus ride" (B) are presented from left to right. It can be observed that in both cases of no correction (Sno, Bno) the MOS is significantly lower than the other scores. Decreasing the headroom to -14 dB (B14, S14) increases the MOS significantly but still the result can be improved. For the other three methods indicated by 9,14a and 14b the resulting MOS can hardly be distinguished.

Overall we can conclude that there is definitely a difference between no-enhancement and using any enhancement, but also (weaker) between -14 dB and the other enhancement methods. However, it is reasonable to assume that there is no difference between -9 dB, and the EQ/MB versions of -14 dB.[1]

5.4 Summary of experiment results

In Chapter 4 and 5 has been proven that the introduction of environmental noise has a negative influence of the Quality of Experience while listening to audio from a mobile device. Figure 5.7 is showing this. What can be observed in Figure 4.8 is that there are solutions against this unwanted influence of environmental noise. The related experiment proofs that decreasing the dynamic range of audio does increase the QoE. In Paragraph 4.2.2 is presented that the introduction of environmental noise will press the QoE. However, this negative influence can almost be made undone. Figure 4.8 tells that when a minor correction is applied, processing audio to the dynamic footprint for TV, the QoE increases significantly. A stronger correction, applying the dynamic footprint for iPods, the QoE reaches from the MOS rating "poor and fair" to the rating "fair and good".

When analyzing noise components separately, different noise sources and noise types can be observed. When a prepared masker for a particular noise, bus-ride noise or street noise, suppresses the detected noise sources the QoE increases.

Figure 5.3.8 shows a set of boxplots of ten correction methods on three noise types. Correction method Sno and Bno means that no corrections are applied on the content. Correction methods starting with S are representing correction methods for street noise and correction methods starting with B are representing correction methods for bus-ride noise. The correction methods are described in Table 5.2.

The plot in Figure 5.3.8 shows that the extremer the correction method is, the higher the QoE reaches. The correction methods presented in Paragraph 5.3 perform so well that for content without correction methods, the rating can be increased from "poor" to "good" with introduced street noise and from "fair" to "very good" with introduced bus-ride noise.

The overall results of the results are showing that environmental noise can be overcome by applying the right audio processing.

6 Implementation suggestion

In the next paragraphs an implementation suggestion of the described technology of this thesis is presented. In short; the previously described audio correnction technology to increase the QoE in case of environmental noise is working, but as with a lot of new technologies, a technology must find its way towards the real world to become available for daily life use.

It is unlikely that users of mobile multi-media devices will alter settings of mobile devices continuously to set parameters of the environment where they are for optimal audio enhancing. Therefore the alterations of these settings must be done automatically. Automatic detection of the user context is supposed to be possible with localization capabilities (e.g. GPS, Wi-Fi finger printing, 3G / GSM Base station triangles etc.) that are build into modern smart phones. With a constant available internet connection, context detection mechanisms and context databases are reachable by end users.

In this Chapter the word "context" is used often. In the previous part, the word context was used in a more technical sense (e.g. a certain frequency span of a disturbing factor at a location), but in this part the word context means where the mobile content consumer is and which disturbances the consumer could be facing.

6.1 Path from content provider towards end user

In the Figure 6.1 the process of context gathering is showed. This chapter describes this process step-by-step.

To be able to describe the context of a mobile device, an exact location of the mobile device must be known. To do this, certain technologies can be used which

are available in common for modern smart phones.

The used localization technologies in this schematic are GPS, Wi-Fi (or WLAN) finger printing and GSM. Wi-Fi and GSM are not meant as localization technologies for themselves, but since there are databases available with the exact locations of their base stations, a location can be derived.

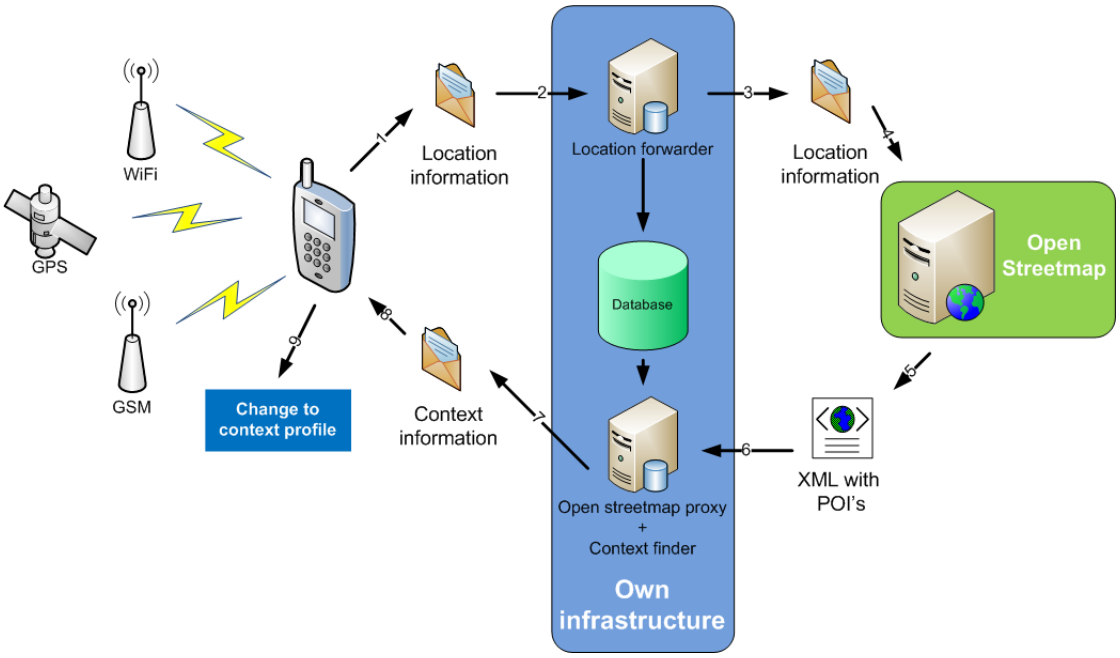


Figure 6.1: Context gathering process

By measuring the signal strength of each of available base stations at a certain point in the landscape, where the mobile device is, a rough estimation of the current location can be made. How this works is not a topic in this thesis.

The process description step by step:

1. Mobile phone wraps the estimated location in a message.
2. The location message will be sent to a location forwarder. This location forwarder will store this location in a local database. This database will put

all sent locations in a sequence to determine later the route of the mobile phone user.

3. The collected location information will be wrapped in an anonymous message that contains just the location and the size of a rectangle.

4. The location message with location information will be sent to *Open Street Map*[1]. Open Street Map is an online database with location related enitities. Not only map information is provided, but also rich information like bus stops, railway stations, traffic lights etc. Information from this database can be extracted by entering coordinates of a rectangle.

5. An XML file will be parsed from open street map as a reply.

6. The XML file from step 5 will be received by an Open Street Map message collector. This collector should work as a proxy and as an information filter. The collector will determine from the location history and new received the context where the mobile user is. Chapter 10 describes how the context detection works.

7. The context proxy will send a simple message to the mobile device with the context description of the user. Examples of the context are tram ride, bus stop, bus ride, walking etc.

8. The mobile device receives the message and switches to the recommended preset.

9. The end user is happy with the new settings.

6.2 Concept of environment detection

Detecting or measuring a location is not enough to determine the context of where a person is when the context could be a moving vehicle where a person travels with or in or when that person walks. Of course that same person can also sit while

[1] http://www.openstreetmap.org

waiting for a line bus. To cover the possibility of moving vehicles, a track record must me made. Figure 6.2 shows an example.

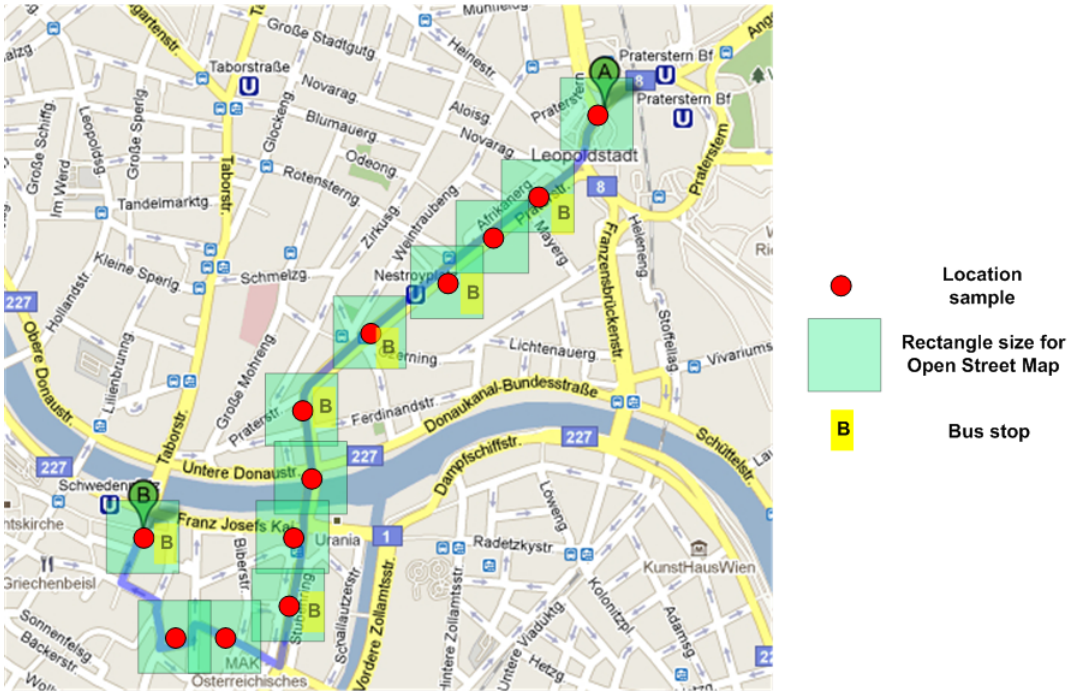


Figure 6.2: Example of a track record of a user

In the example of Figure 6.2, the track record of a user has been plotted with a blue line. The user started at point A and ended at point B. During the trip the user sits in an autobus. By tracking and interpreting this record, this can be derived from Open Street Map data. Every red dot on the map, the smart phone of the user sends his location, number of GPS satellites, number of GSM base stations and speed to the server as shown in Figure 6.3. The server stores this data and retrieves context information from Open Street Map. From the retrieved data it is extracted that on the route lies a bus route and that the speed of the person lies at 0 km/h around bus stations.

The number of GPS satellites and GSM base stations are of such number that it is very likely that the user is above the ground. For example in a metro station, no GPS satellites can be received and the number of GSM base stations is very low. Figure 6.3 shows a rough sketch in a diagram of this example. Not all samples are shown but a brief summary.

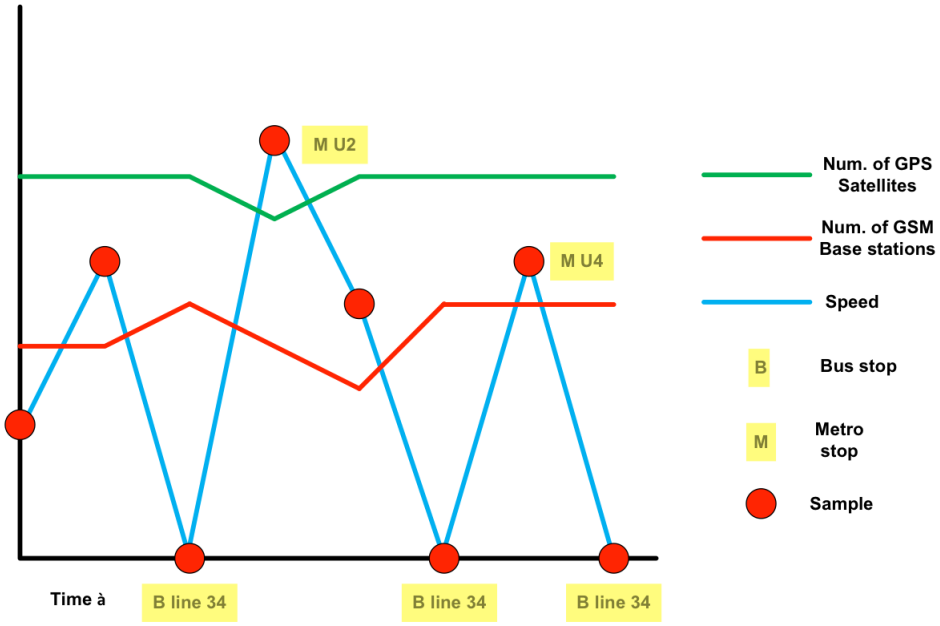


Figure 6.3: Sketch of a track record

What is presented in Figure 6.3 is the speed of the user, number of received GPS satellites and GSM base stations. In the yellow text boxes is a public transport entity shown when was around at that moment. The public transport entities were extracted from Open Street Map. In this case it is very likely that the user is sitting in a bus.

Other cases for other types of transportation can be derived easily.

In Figure 6.4 a decision diagram is shown in which can be seen how a used way of transportation can be derived. Of course this decision diagram can be refined and more parameters could be implemented, but this diagram is presented here to show the concept of context detection.

Figure 6.4: Context detection logic

But even that diagram is far from optimal; it works for the given example. The order of deriving the environment / mean of transportation from the Figure 6.3 by using Figure 6.4 would be:

1. Number of received GPS satellites and GSM/3G base stations: **HIGH** => user is outside and/or above the ground
2. Speed: **FAST** => User is being transported
3. User follows public transport route or track: **YES** => User uses public transport
4. At which public transport stops is the user holding? **BUS** => The user is inside a bus

With this method it is easy to determine the environment of a user by just using available contextual sensors and logical analysis.

7 Advice for content producers, broadcasters and mobile device vendors

In the previous chapters is described how different each environment has its own influence in the Quality of Experience of a user of mobile multimedia. In this case audio only, but it is very likely that this also counts for video. The following advices are addressed to three groups: content producers, broadcasters and mobile device vendors.

7.1 Advice for content producers

For the content producers lies the task to produce their generated content following the EBU R 128 recommendation. This recommendation will not be the last step in the process for loudness normalization for especially mobile media usage. This is a very important step for content producers in general to work according a supported measurement standard [6] and following a recommendation [9] that makes sense. This will not only help broadcasters to normalize all the content for television, but it will also make it a lot easier to convert media for different presentation devices. In the end the quality of productions will increase and that is what should count.

7.2 Advice for broadcasters

It would be ideal for broadcasters to have concent delivered for broadcasting and pre-processed archives processed according to the EBU R128 recommendation, but this is not the case. When broadcasters start to accept only incoming programs, commercials and station bumpers following [9], the work is half done.

Audiovisual archives can be devided into two major divisions:

- Content stored on old media. Commonly used media types are:
  - Film: e.g. super 8mm movie reels
  - Anaglog video: e.g. 1 inch Bosch b-format and 2 inch Ampex VR-8000 format
  - Digital video: Sony D2 (composite uncompressed), Sony Digital Betacam (compressed component video)
  - File based systems: e.g. Sony Professional - Blue ray / Memory Stick / SD-card. (XDCAM)
- Content stored in files on an automated common platform (e.g. harddrives, tape robots, etc).

The step when content is captured from old storage media and stored into files is called ingestion. During this ingestion step archives are restoring the content from for example old film reels having bad quality in order to upgrade it to the original state. Not only the quality of the content itself is important, but meta-data must be ingested into files as well (e.g. shot lists, production data, etc.). This is a time consuming step and media is often played and analyzed lots of times. During all these runs, or a final run, or even easier, when the content is finally a file, the loudness values of the content should be stored with the rest of the meta-data. It is not even needed to alter to content, but the loudness average would give broadcasters in the final broadcast step the opportunity to give this file an offset in the last processing stage.

When broadcasters are taking the two mentioned points into account, a big step forwards in loudness treatment can be made.

7.3 Advice for mobile device vendors

In the previous chapter is presented that the final processing stage for optimal loudness and auditory masking treatment should be done in the mobile device

itself. To do audio processing tasks without consuming too much energy from the battery, a low-power Digital Signal Processor (DSP) should be implemented in mobile devices. The DSP can do audio processing tasks which do not increase the load of the main CPU a lot. A CPU that does not have a high work load does not consume a lot more of energy than normal which result in a low impact on battery performance.

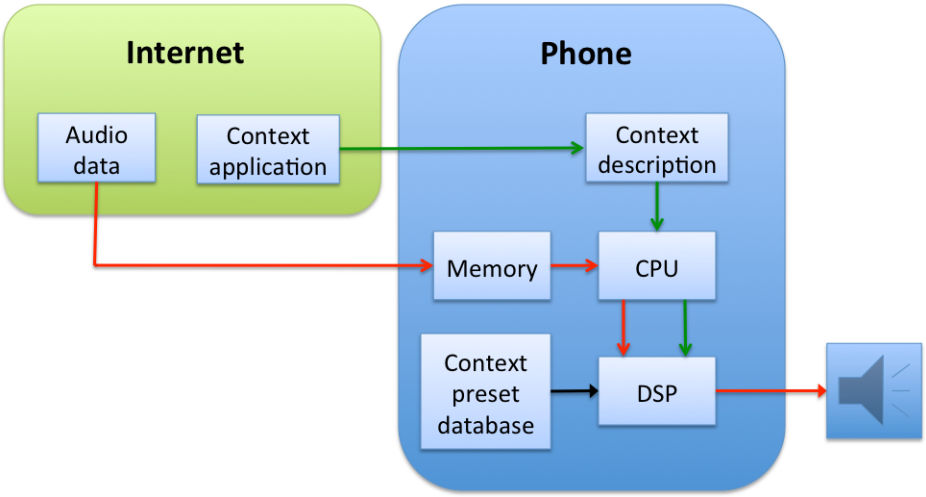


Figure 7.1: Schematic of phone and DSP

In Figure 7.1 shows a schematic of how a DSP should be implemented in this concept. The audio data comes in this example from the internet. The audio data comes via the memory of the phone in the CPU and parses into the DSP. The DSP loads from context descriptions from the internet into the DSP. The DSP loads from a context preset database the requested processing settings. The DSP combines the audio data with the processing settings and plays the enhanced audio. All components in the schematic are already present except for the context application and the DSP. The role for mobile phone vendors is to add a DSP for the audio processing. When a mobile phone vendor wants to add video processing in the future, a more capable DSP can be used as well.

8 Conclusion

This thesis starts with a survey on the history and recent industry status of handling loudness and dynamic range issues for television and radio. The outcome of the survey presents that for home environments there is a recommendation for broadcasters to optimize audio for TV sets. This optimization increases the QoE for people watching TV at home on a normal TV set. A survey on the International Broadcast Conference 2010 (IBC) in Amsterdam gave the result that the broadcast and content production industry is not working on recommendations and standards to optimize audio for mobile multimedia.

For users of multimedia content, the need for audio optimization is high. With an experiment it has been proved that introduction of environmental noise decreases the QoE dramatically while consuming audio at e.g. a bus stop or inside a bus.

Research shows that a multimedia usage environment or context, e.g. cinema, home or iPod, has an optimal dynamic footprint. A presented experiment in this thesis proofs that when environmental noise is introduced to a test person, who is listening to audio, the QoE decreases. However the QoE increases again when the dynamic range is decreased. The result is the best when a test person, who is listening to an iPod, listens to content that is meeting the preferred dynamic footprint of an iPod.

Another way to optimize audio content is to adapt properties of environmental noise into the processing. This thesis proofs that when recorded noise is analyzed for its properties, different noise sources can be recognized. By knowing with what kind of noise must be dealt with, a processing "preset" can be made. When this technique is applied to audio content, the QoE increases.

When the optimization technique of decreasing a dynamic range is combined with

the technique of adding noise properties to the optimization process, the QoE increases even more. An experiment, which is presented in this thesis, proofs this.

An application suggestion in this thesis should allow mobile multimedia devices with internet access to switch automatically to an optimal environmental setting. When this context detection technology is enabled and implemented, the user does not have to switch manually between optimization settings. A very big advantage of this application is that modern smartphones are able to enable this technology. Another advantage is that no external hardware is needed.

With the combination of the presented technology and the application suggestion should it be possible to deliver a higher Quality of Experience to mobile multimedia consumers that are now suffering from environmental influences. By having this new technology, the experience of mobile multimedia goes another step into the right direction, which is: Consuming content on a high quality whenever and wherever we want.

Curriculum Vitae

Personal information

Name	Martijn Cornelis Sack
Born	22. July 1981, Zwolle, The Netherlands
Nationality	Dutch
Email	martijn.sack@gmail.com

Education

1992 – 1997	MAVO Salland, Raalte, The Netherlands
1997 – 2003	MBO Electrical engineering for IT and telecom, Rens & Rens, Hilversum, The Netherlands
2003 – 2007	Bachelor study Media Engineering and Electrical engineering, Saxion University for Applied Sciences, Enschede, The Netherlands
2007	B2 and C1 German language course including Zentrale Mittelstufenprüfung Goethe Institut, Rotterdam, The Netherlands
2008 – onwards	Master study Media Computer Science. Universität Wien, Austria

universität wien

Employment

SS 2009 – SS 2011	Master thesis student, junior researcher University of Vienna, Multimedia and Distributed Systems, Faculty of Computer Science, Vienna, Austria www.cs.univie.ac.at
2007 – 2009	Web developer, process development MailStreet BV, Deventer, The Netherlands. www.mailstreet.nl
SS 2007	Graduation internship, Researcher interactive TV TNO ICT, Delft, The Netherlands. www.tno.nl
2002 – 2008	Owner, producer, director, engineer and developer. OMG Media Productions, Enschede, The Netherlands. www.omg-mediaproducties.nl
2001 – 2004	Graduation internship, network engineer, broadcast engineer and software developer Technicolor, Hilversum, The Netherlands. www.technicolor.com
2003 – 2006	Teacher and media engineer Saxion University for Applied Sciences Enschede, The Netherlands. www.saxion.nl

Spoken Languages

Dutch	Mother language
English	Speaking: Good, Writing: Good, Reading: Good
German	Speaking: Good, Writing: Good, Reading: Good

Conferences

Loudness and Auditory masking compensation for Mobile TV, Martijn C. Sack, Shelley Helmut Hlavacs, accepted for presentation at the IEEE International Symposium on Broadband Buchinger, Multimedia Systems and Broadcasting, 24 - 26 March 2010, Shanghai, China

Ambient Assisted Living Forum, PhD program, presentation and poster session, October 2009, Vienna, Austria

Publications

Slider or glove? Proposing an alternative quality rating methodology, Shelley Buchinger, W. Robitza, P. Hummelbrunner, M. Nezveda, M. Sack, Helmut Hlavacs, VPQM 2010 - Scottsdale Arizona

Project Experience

CACMTV The main intent of the project “Content Aware Coding for Mobile TV" is to close the research gap dealing with technical as well as dramaturgic aspects of mobile TV. The cooperation between two departments of the University of Vienna ensures the theoretical and methodological basis for the two necessary multi-disciplinarily. http://www.ani.univie.ac.at/~cacmtv/

AMASL Ambient Assisted Shared Living (AMASL) aims at creating a professional multimedia communication system for elderly people currently living alone in their homes with no one to talk to. The project aims at developing prototypical applications and installing them into actual households of elderly people and their relatives. http://www.amasl.at

GpENI Great Plains Environment for Network Innovation

- https://wiki.ittc.ku.edu/gpeni_wiki/index.php/

Research interests

- Ambient Assisted Living
- Broadcasting services
- Social networks
- Telecommunication
- (Mobile) Multimedia services

Teaching Experience

2003 – 2007	Teacher of course the “Video and audio engineering” at Saxion University of Applied Sciences, Enschede, The Netherlands
2009 – 2011	Tutor in the course “Distributed Computing” and tutor for bachelor students during thesis projects. University of Vienna, Multimedia and Distributed Systems, Vienna, Austria

Networks

- IEEE Broadcast and Multimedia Society (IEEE-BMS)
- European Broadcast Union (EBU), P/LOUD group, http://tech.ebu.ch/groups/ploud
- Telecommunication Cercle (TCC 08), TU Wien, Institute for Broadband Communication, Vienna, Austria www.ibk.tuwien.at/~tcc08

Bibliography

[1] M. C. Sack, S. Buchinger, W. Robitza, P. Hummelbrunner, M. Nezveda, and H. Hlavacs, "Loudness and auditory masking compensation for mobile tv," University of Vienna, Tech. Rep., 2010.

[2] G. Spikofski and S. Klar, "Levelling and loudness in radio and television broadcasting," *EBU TECHNICAL REVIEW January 2004*, 2004.

[3] T. Lund, "Level and distortion in digital broadcasting," *EBU TECHNICAL REVIEW – April 2007*, 2007.

[4] (2010, November). [Online]. Available: http://www.britannica.com/EBchecked/topic/348615/loudness

[5] M. J. Crocker and B. Scharf, *Handbook of Acoustics - Equal-loudness contours for pure tones presented through a pair of earphones.* Wiley-IEEE, 1998, no. 2, figure 91, p. 1184.

[6] ITU-R, "Recommendation itu-r bs.1770-1 - algorithms to measure audio programme loudness and true-peak audio level," ITU-R, Tech. Rep., 2006-2007.

[7] P. group, "Loudness metering: 'ebu mode' metering to supplement loudness normalisation in accordance with ebu r 128," EBU, Tech. Rep., August 2010.

[8] ——, "Loudness range: A descriptor to supplement loudness normalisation in accordance with ebu r 128," EBU, Tech. Rep., August 2010.

[9] ——, "Ebu – recommendation r 128," EBU, Tech. Rep. 1, 2010.

[10] I. R. Assembly, "Recommendation itu-r bt.500-10 - methodology for the subjective assessment of the quality of television pictures," ITU-R, Tech. Rep., 2000.

[11] S. Buchinger, W. Robitza, P. Hummelbrunner, M. Sack, and H. Hlavacs, "Silder or glove? proposing an alternative quality rating methodology," University of Vienna, Tech. Rep., 2010.

[12] S. Gelfand, *Hearing- An Introduction to Psychological and Physiological Acoustics*, 4th ed. INFRMA-HC, 2004.

[13] J. Hartung, B. Elpelt, and K.-H. Klösener, *Lehr- und Handbuch der angewandten Statistik.* München and Wien: Oldenbourg, 2002.

List of Figures

List of Tables