BACHELOR THESIS

Perceived Sound Quality of Dynamic Range Reduced and Loudness Normalized Popular Music

Jakob Lalér

Bachelor of Arts
Audio Engineering

Luleå University of Technology
Department of Business, Administration, Technology and Social Sciences

Perceived sound quality of dynamic range reduced and loudness normalized popular music

Lalér Jakob

ABSTRACT

The lack of a standardized method for controlling perceived loudness within the music industry has been a contributory cause to the level increases that emerged in popular music at the beginning of the 1990s. As a consequence, discussions about what constitutes sound quality have been raised. This paper investigates to what extent dynamic range reduction affects perceived sound quality of popular music when loudness normalized in accordance with ITU-R BS. 1770-2. The results show that perceived sound quality was not affected by as much as -9 dB of average gain reduction.

TABLE OF CONTENTS

ABSTRACT......2

INTRODUCTION......4
Aim, Objectives and Limitations......4

Background......4
"Dynamic Range"......5
Dynamic Processing Tools......5
Multiband Compressors......5
Look-ahead Limiters......6
Compression and Hearing......6
EBU R-128......7
The Research Conducted......7
The Algorithm......7
Loudness Range......8
Integrated Loudness......8

METHOD......8
Participants......10
Equipment......10
Stimuli......10
Signal Processing......11
Quality Assessment......12

RESULTS......12
All genres......13
Rock......13
Pop......14
Country......14
Electronic......15

DISCUSSION......15
Reliability and Validity......16

FUTURE WORK......16

ACKNOWLEDGEMENTS......16

REFERENCES......17

APPENDIX......18

1. INTRODUCTION

Compression is used in the music industry to change the level of a signal in order to make it better suited for a specific situation. This might be to control the natural dynamic changes of an instrument at the tracking session in order to make it fit later in the mix. Compression can also be used to increase the loudness of a signal and the development of software dynamic processing tools has undermined the so called “loudness war” [1].

In broadcasting, the loudness war escalated when commercial interests discovered a loop hole in the frequency modulation technique used for radio and television. QPPM meters (quasi-peak program meters) generally has an integration time of 10 ms, making them unresponsive to shorter transients than this. Therefore, a headroom was necessary in order not to exceed the maximum level for broadcasting. A *Permitted Maximum Level* was set to -9 dBFS which is 30 kHz in frequency deviation allowing a 20 kHz of headroom up to 50 kHz which is the standard maximum deviation in many countries. In order to gain loudness, this headroom of 4.4 dB was sacrificed with the use of dynamic processing, i.e. lowering the peaks and raising the RMS level. 50 kHz was equated with -9 dBFS, leaving no headroom for the signal at all. This meant that a program which transmitted -9 dBFS as 30 kHz was lower in perceived loudness than a program with a realign permitted maximum level of -9 dBFS as 50 kHz [2].

The problem has been well known for a long time and has led to a standard, initialized by the ITU-R and adopted and developed by the EBU. The ITU-R standard is being implemented in the EU, the USA as well as other countries. This standard seems to be one step in the right direction, removing any loudness advantage and instead favoring wide dynamic material. In the film industry, a uniform way of trimming monitors before mixing has created a theatre standard so the problem with loudness differences will perhaps be most prominent within the music industry in the nearby future.

Today there are websites that are aimed at "fighting" the loudness war in music [3, 4, 5] and on forums, discussions about the bad quality of hypercompressed music is ongoing. Hypercompressed music (in this paper defined as music with audible compression artifacts such as modified transients, lack of inherent musical dynamics and occasionally clipping) is believed to remove dynamics, create musical clutter and reduce excitement and emotion in the music [6]. It is also a risk of further degradation in sound quality when hypercompressed music is coded through various data rate reduction processes. This could potentially cause intermodulation distortion and listening fatigue [7].

There are definitely situations where music benefits from heavy dynamic processing, aesthetically and functionally (e.g. style of music and place of reproduction). With music software such as iTunes and Spotify offering level control for all tracks, part of the music industry is moving closer to a way of eliminating loudness advantages but the problem still prevails.

1.1 Aim, Objectives and Limitations

This paper accounts for a listening test, investigating the relationship between perceived sound quality and dynamic range reduced popular music using participants with different listening experience. It is focused on specific genres and limited to one dynamic processing tool, using one setting in order to make the result as reliable as possible. Sound quality in this study is equal to basic audio quality which is defined as the overall quality of the mix, assessed by each individual listener (see section 3.5).

2. BACKGROUND

There have been discussions about the degradation of sound quality in popular music productions due to the extensive use of dynamic reduction tools [8, 6]. Evidence for decreasing dynamics in popular music are available from websites such as Unofficial Dynamic Range Database and there has been AES workshops discussing the loudness war and its effect on sound quality [7]. It has been proposed that the ongoing trend in declining sales figures for the major music labels is not only due to piracy but also that the quality of the distributed music is worse now than before as a result of hypercompression. In the music industry, the increasing levels have been so immense that professionals and listeners alike have complained about the poor sound quality [8, 6, 5]. The problem seems to be a nonexistent uniform way of monitoring loudness.

2.1 “Dynamic Range”

“Dynamic range” is indeed a expression of ambiguity. The reason seems to be that the word is used to define multiple things. One explanation is *“the difference between the loudest and softest passages of the body of the music*..”. [8]. This could be interpreted as the inherent dynamic changes of a musical movement. However, the term can also be used to define the headroom of electric signals for a recording medium. Another interpretation of the word could be the crest factor (a signal´s peak amplitude divided by its RMS value). Furthermore, dynamic range could mistakingly be equated with loudness perception.

With this fact in mind, it is quite clear that the term per se must be defined in order to avoid misinterpretation. Since this paper is primarily concerned with loudness range as described in EBU – Tech 3342 and crest factor, defined as the difference between a signal´s peak amplitude and its RMS value (or put differently, the peak amplitude divided by the RMS value) [9], these two terms shall be separately used henceforth with the addition of dynamic variation which expresses macrodynamic changes without a specific time window of analysis (in the headline of this paper, dynamic range is analogous to crest factor).

During the most intense era of the loudness war (c:a 1990-2005), there seems to be evidence for preserved loudness range in music. This holds true despite the fact that the use of dynamic processing tools, which were developed and more extensively used during this time, will result in a decreasing crest factor and loudness range. In addition, this is contrary to the fact that the overall crest factor decreased during the same time [10].

The explanation seems to be related to the music processed and its musical arrangement. In [10], three different genres are exemplified and their RMS gain is affected differently. Here, it is clear that the pop/rock song is not as sensitive to dynamic processing as the other two music examples (opera and jazz) and therefore more dynamic processing can be applied to this example (pop/rock) without affecting the difference between the softest and loudest parts in the song as much. The larger this difference is (softest to loudest part) on a shorter time scale, the more can the music be compressed without affecting the loudness range.

Loudness range, as defined by the EBU – Tech 3342, has a set time for the analysis window (see section 2.4). It has been proved that the length of the analyzing time window used when defining the variation in the RMS levels of a compressed song, will affect the outcome [10]. This means that different songs, compressed with the same settings, can result in different decrease of LRA values.

2.2 Dynamic Processing Tools

There is evidence for increasing loudness in popular music since the 1990´s [10, 3, 5]. Since the overall RMS value has increased, the peak-to-RMS value has decreased [11]. A simple RMS value corresponds well to human loudness perception (see section 2.4) so it is clear that the perceived loudness of popular music has increased too. From the middle of the 1980´s until 2005, the average level in popular music has increased by 5 dB and during the same time, the crest factor (or more specifically, a simplified version of the crest factor) has decreased by 3 dB. It is clear that this is due to the excessive use of dynamic compression in modern music productions. It is also clear that peak clipping is more common during this era than during previous decades which is further evidence for the influence of dynamic processing [10].

2.2.1 Multiband Compressors

The development of multiband compressors has made it possible to increase the loudness of a song to a greater extent than before. These tools offer the engineer control over separate frequency bands with controls for each band so that they can be manipulated without affecting the whole signal. Therefore they might be used in a similar way to an equalizer. One advantage with a multiband compressor is that the whole mix is not affected by a loud sound in one band [12]. This means that the signal can be compressed more precisely (unlike a single-band compressor that processes the whole signal with the same settings) and therefore avoiding pumping artifacts to a greater extent. The use of multiband compressors can however, lead to degradation in sound quality. Katz explicitly states that such compressors are ”*...potentially the most deadly audio process...ever invented*” inasmuch as it ”*helps (to) fuel the loudness race*” [8].

Another potential problem with all conventional multiband compressors, is the non-linear response of the signal level at the band boundaries. This is due to the fact that the energy of the signal at the boundaries is not fully recognized by the overlapping frequency bands which results in lower compression values.

2.2.2 Look-ahead Limiters

A common dynamic processing tool utilized in mastering houses, which has facilitated the increasing levels in popular music, is the look-ahead limiter. It includes a delay line in the audio path but not in the side chain path which enables the limiter to predict the waveform and instantly react to sudden transients [8]. This form of limiting is efficient for controlling peaks because of its infinitely small reaction time but with this function comes the risk of changing the overall sound of the mix. With a look-ahead function, transients can be severely changed with the risk of altering the final mix.

2.3 Compression and Hearing

Loudness perception is a subjective phenomenon and corresponds to an individual listeners´ preferences and physical abilities. Thus, loudness may seem difficult to quantify since every listener has a unique interpretation of loudness. However, as already stated, the average amount of acoustical energy that a listener is exposed to over a fixed period of time, corresponds well to average human loudness perception [13]. Moreover, short and sudden transients does not have a significant effect on the overall loudness of a song [14].

Why do engineers, musicians, producers and typical listeners favor louder mixes? There are some good explanations to this. Physically, the Fletcher-Munson curves prove that human hearing is non-linear. This means that a song reproduced at a lower level, tends to sound less good simply because it does not excite as many hair cells as a louder version of the same song would have done and therefore the frequency ranges are not heard in as great detail. Besides, a loudness increase has a larger effect on the human psyche than the same loudness decrease which probably motivates an increase in loudness than vice versa [15]. As a consequence, two songs with different loudness will be perceived differently and a typical listener will generally favor the louder of the two.

So far, the argument has been concerned with loudness differences between materials. What happens when a listener hears two versions of the same song reproduced with the same loudness? Reduction of peaks by dynamic processing tools will alter the overall sound in a mix. It corresponds to a decrease in crest factor, resulting in less dynamic variations which will change much of the original arrangement of the music. This will yield a more static mix which might be regarded as less natural [6].

There is also the aspect of coloration from dynamic processing. Peak clipping for instance, causes harmonic distortion and it is common for hypercompressed music to be mastered close to full scale. This was observed by the author when performing measurements of a total of 60 commercially released songs (see appendix). It is also believed that distortion causes listening fatigue [14] and it has already been stated that hypercompressed music removes natural dynamic variations which will (to some extent) affect the neural activity (i.e. create a static impression of the music).

Stone et al performed a two-talker separation task with twenty-four normal hearing listeners using different amounts of multichannel compression [16]. It was concluded in this test that heavy multiband compression reduces speech intelligibility. In addition, heavy compression demands higher cognitive effort by the listener in order to discriminate mixed signals since reduced temporal and spectral contrasts (amplitude variations over time and amplitude variations across frequencies) minimize information to the auditory system. This test also proves that compression prior to summing signals increases the ability to differentiate the signals. This is most likely due to cross modulation at the gain reduction stage of the compressor.

In this test it was also found that compressed speech with the same RMS value as uncompressed speech yielded higher loudness values. This suggests that a simple RMS value is not sufficient in predicting loudness and as far as the author is aware, a loudness test with compressed and uncompressed loudness normalized material according to the EBU - Tech 3342 specification has hitherto not yet been performed.

2.4 EBU R-128

In order to avoid level inconsistencies at the consumers end of the distribution chain for broadcasting material, the ITU-R released the document BS. 1770-1 in 2006-07 which presented an algorithm to measure audio program loudness and true-peak audio level. This recommendation was later developed by the EBU and defined as EBU R-128.

2.4.1 The Research Conducted

In the first phase of the ITU-R research, Soulodre developed a model for a listening test on monophonic material [17]. Results from listening tests based on this model were accumulated on five different sites with a total of 97 subjects. The program material consisted of sound effects, music, sporting events, news broadcasts, television and movie dramas and advertisements from actual radio and television broadcasts. The stimuli also contained speech segments in different languages. A total of 48 test sequences were played with two level offsets so that the subjects most likely would have attenuated and increased the stimuli against the reference signal.

In addition to these tests, ten loudness meters from seven companies were included in a subsequent test using the same material as in the previous tests. In addition to these ten loudness meters, a frequency weighting curve, referred to as Revised Low-frequency B-curve (RLB), and a RMS measurement were included. These two loudness measurement models performed well in a previous evaluation by Soloudre [13]. The stimuli were processed by each of the loudness meters and by subtracting the loudness of the reference signal from the rating of each of the test sequences, the loudness meters were normalized. The results from the second test proves that the meters performed quite differently. The Leq (RLB) meter proposed by the Soulodre, performed best overall and the meter based on RMS measurement also showed good results.

2.4.2 The Algorithm

For multichannel use, the algorithm includes a pre-filter to take into account diffraction and other acoustical phenomena of the head [17]. This filter could be described as a high shelving filter. Adjacent to this stage, a second order high pass filter is applied and the signal becomes `K-weighted´. The idea with this weighting is to give a signal with relevant spectral content to the listener. Thereafter a processing block computes the average energy over time. This is followed by a gain factor depending on which channel is processed. The rear channels are increased 1.5 dB since they are perceived as louder, supposedly as a consequence of what has been proposed to be a cognitive phenomenon [7]. The LFE-channel is excluded from the measurement and the number of channels can be changed to include fewer if needed. The algorithm is illustrated below:

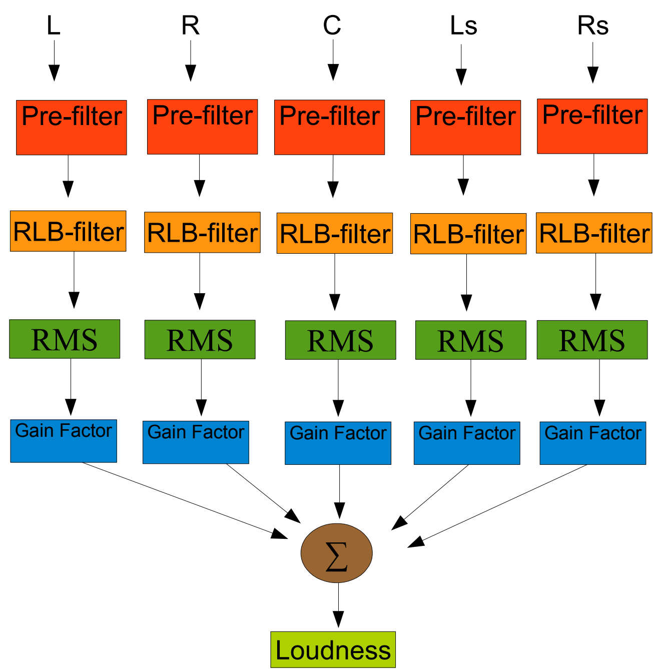

Figure 1. Multichannel loudness algorithm after ITU-R BS. 1770-2 [19].

By utilizing this method, the loudness of typical program material can be controlled in an effective way. However, it was discovered that segments of lower sound affected the loudness result considerably. Therefore, the EBU developed the algorithm by introducing a gate to accommodate for this. This gate is only applied for measuring integrated loudness and loudness range. With the more refined algorithm, the loudness measurement is more reliable, excluding fade outs and longer segments of low level sound from the measurement. Furthermore, the EBU added a `Target Level´ to the measurement which is -23 LUFS [2].

2.4.3 Loudness Range

Loudness Range (LRA) computes the variation of loudness over time. The LRA value is not analogous to a RMS value, peak value or the difference between them. It is based on the loudness algorithm presented in ITU-R BS.1770 and is specified in LU (loudness units) where 1 LU is equivalent to 1 dB. Furthermore, the algorithm is akin to the measure of statistical dispersion in descriptive statistics, thus very loud sounds and segments of low level sound do not affect the LRA value [18].

The LRA value is quantified by applying different gating thresholds to the measurement. An absolute threshold of -70 LUFS is exceeded only when a signal is considered present. Thus, the measurement starts when the signal level goes beyond this value whilst ignoring any low level noise below it. A different relative threshold gate is set -20 LU below the absolute-gated loudness level which is the highest level measured after the absolute threshold has been exceeded. A sliding analysis-window of 3 seconds is employed in this scheme. The relative threshold is therefore not fixed but varying, depending on the loudness of the present signal. An overlapping window is used in order to make the measurement robust against shorter segments of audio and a minimum of 2 seconds (66%) is required.

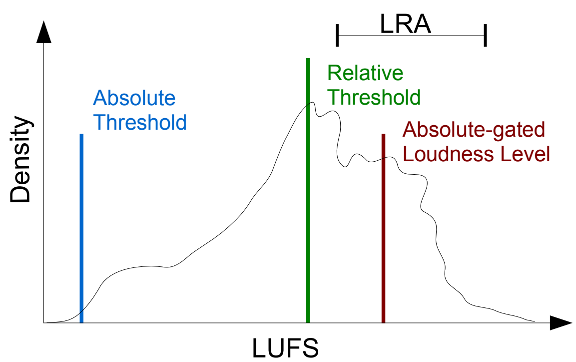

Figure 2. LRA gating scheme after EBU – Tech 3342 [18].

If a program with steady loudness which includes low level background sound is measured without the gates, the LRA value would be misleadingly high. With the gates however, such material can be measured correctly. The cascaded gates leaves out low level parts and the spread of the loudness levels between the 10th and 95th percentile above the relative threshold is the LRA value. It should be stated that the EBU does not specify a LRA value but recommends LRA measurement to see if dynamic range reduction is needed in order to make the signal fit to a specific distribution platform (podcast, TV, radio, film etc.).

2.4.4 Integrated Loudness

Loudness Range is a supplement to the measurement of Integrated Loudness level and they are somewhat different. Both measurement uses the same algorithm but for Integrated Loudness, the relative threshold is set to -10 LU relative to the absolut gated loudness level. Moreover, the analysis windows (specified as *gating blocks* by the ITU) are 400 ms and overlapping by 75% (300 ms) [19].

3. METHOD

Initially, the author evaluated the possibility to manipulate the LRA values of the stimuli with dynamic processing. In order to decide the amount of dynamic processing to be applied to each stimuli, LRA measurement of 60 randomly selected albums were conducted from justiceforaudio.org [5]. These albums were compiled from three categories listed as “Bad Quality Audio”, “Only Mildy Affected” and “High-Quality Audio” (see Appendix for additional information). The categories are based on measurement using the TT Dynamic Range Meter software.

The songs were coded with 320 kbps Ogg Vorbis. Prior to these measurements however, the author evaluated what possible side effects the Ogg Vorbis algorithm could have on the result. Since the largest difference observed was +/-0.2 LRA between Ogg Vorbis and wav-files, the former format was deemed sufficient for the measurements. This conclusion is also in agreement with the findings in [20].

However, the songs measured ranged from 0.9 to 18.2 LU and the difference in loudness units between categories listed as “bad affected quality audio” and “mildly affected quality audio” in [5] was not explicit enough to be used as a guideline for the construction of the stimuli. Furthermore, the LRA values for two of the stimuli was only 1.3 LU (uncompressed). In addition, as seen in the diagrams below, small variations in LU due to dynamic processing on a stimulus with a low LRA value (uncompressed) proved not reliable. In figure 7 and 8, the LRA values increased notwithstanding that the gain reduction increased. Therefore this method was abandoned.

Instead, arithmetical gain reduction values were determined for each dynamic range reduction category. These were decided based on the fact that typical dynamic range reduction at a pre-mastering house rarely exceeds ≈ 2 dB [21]. This level is regarded as subtle compression in this paper (subtle compression was left out in order to minimize the length of the listening test) and consequently the following values were decided for each genre, where AGR corresponds to Average Gain Reduction, PGR to Peak Gain Reduction and LRA to Loudness Range:

	None (N) = 0 dB
	Moderate (M) ≈ 3-4 dB
	Strong (S) ≈ 6 dB
	Extreme (E) ≈ 9 dB

Figure 3. Dynamic reduction categories

	AGR	PGR	LRA
	0 dB)	-	1.8 LU
	≈ 3-4 dB	-8.5 dB	1.3 LU
	≈ 6 dB	-10.4 dB	1.2 LU
	≈ 9 dB	-12.7 dB	0.9 LU

Figure 4. Rock music

	AGR	PGR	LRA
	0 dB	-	1.3 LU
	≈ 3-4 dB	-8.5 dB	1.0 LU
	≈ 6 dB	-9.7 dB	1.1 LU
	≈ 9 dB	-13.7 dB	0.9 LU

Figure 5. Pop music

	AGR	PGR	LRA
	0 dB	-	1.5 LU
	≈ 3-4 dB	-5.7 dB	0.6 LU
	≈ 6 dB	10.0 dB	0.9 LU
	≈ 9 dB	-12.8 dB	0.9 LU

Figure 6. Country music

	AGR	PGR	LRA
	0 dB	-	1.3 LU
	≈ 3-4 dB	-9.3 dB	0.8 LU
	≈ 6 dB	11.6 dB	0.7 LU

	≈ 9 dB	-15.0 dB	0.5 LU

Figure 7. Electronic music

3.1 Participants

A total of 15 subjects (4 women and 11 men) in the ages of 22 to 31 years participated in the test. These were all students at the school of music in Piteå except for one subject who worked in Piteå, outside the school. Four subjects were enrolled in music education, four where first and third year sound engineer students and six studied media programs. The participants were chosen to represent "typical listeners", defined in this paper as listeners that listen to music at least a few hours every week. The group stated an average of 20 hours of listening to music every week. Eleven subjects claimed that sound quality is important when listening to music and ten subjects reported that they listens to music more often through loudspeakers than through headphones. All subjects reported normal hearing.

3.2 Equipment

The following equipment was utilized in the listening test and the making of the stimuli:

- MacBook Pro 15 inch, 2 GHz Intel Core i7, 8GB 1333 MHz DDR3 RAM, Mac OS X Lion 10.7.3
- Behringer BCF2000
- Fireface 800 interface
- AKG K271 MKII
- KRK VXT6 monitors
- Waves L2 Ultramaximizer Peak Limiter Plugin version 7.0.0.21 Build 4733
- Pro Tools 10.1
- VisLM 1.0
- HP Intel Core i5
- STEP version 1.08a
- Motu UltraLite-mk3

3.3 Stimuli

A progression from verse to chorus in popular music often includes a natural loudness change due to conventional arrangement practices. This holds true for most songs. Excessive use of dynamic tools tend to flatten out these macrodynamic changes and therefore the stimuli in this test initially included a verse and chorus to investigate the impact of compression on a verse to chorus transition. Due to the length of the stimuli however, this was later changed to include only one chorus per stimulus.

The stimuli consisted of 18 second-excerpts from four songs of four different genres. These were all high quality multi-track recordings, recorded in 44.1 kHz/24 bit, except the rock song which was recorded at 96 kHz/24 bit and re-sampled to 44.1 kHz in Pro Tools prior to mixing. One project was provided by the author, another by the artist *Himlakropp* and two additional projects from two separate tracking sessions.

Before the test, subjects were instructed to change the playback level of a test signal consisting of a mastered popular song to a level that they were comfortable with. Once the test started, the subjects did not change the playback level. An account of these levels can be found in the appendix.

The stimuli were chosen to represent common genres of commercially distributed music. They were assessed after their arrangements and instrumentation so that the result of the compression applied would affect the overall sound. For example, all of the songs included drums in the rhythm section and electric guitars and electric bas accept for the electronic music. The songs chosen were not slow in tempo or with static dynamic variation.

3.4 Signal Processing

The mixes were done without any effects and dynamic processing on separate channels except for the electronic music which had light side-chain compression and reverb on some instruments. This was done to eliminate any possible audible affect that such processing might introduce. Thus, the stimuli went through only one compression stage, applied on the final mix bus. The main purpose when creating the mixes were to create them as clear and even as possible and as a consequence, some minor level changes were done to the vocals and some instruments in order to avoid harsh frequencies as this could have affected the test results.

The release time of the L2 Ultramaximizer Peak Limiter Plugin (L2) was set to a maximum of 1.000 and the AGR function (automatic gain release) was not engaged since pumping artifacts was experienced to a greater extent when this function was engaged. This is most likely due to the very short attack time in combination with a fast release time. The short release time used, did however introduce some minor distortion artifacts for the strong and extreme categories.

The AGR values were subjectively assessed from the L2 gain reduction meter since no exact average level of reduction could be displayed. However, these levels were carefully monitored several times in order to make the values as exact as possible. The PGR values were observed from the peak hold function on the L2. The LRA and the integrated loudness value were obtained from a VisLM loudness meter and each mix, with and without dynamic processing applied, were normalized to EBU´s target level of -23 LUFS before re-recording of the final mixes in Pro Tools.

As can be seen on the wave forms (top channel = Left, bottom channel = Right) for gain reduction categories *none* and *extreme*, the transients were changed to a great extent in all four songs, removing much of the impact of the rhythm instruments. In the rock song, the transients from the snare drum, the guitars and the lead vocal were radically reduced. This is even more apparent in the pop song which is very much based on a rhythm section including acoustic drums and sampled marimbas.

The first and third beat in the country song is emphasizing by the drums as well as the electric bas and the small loudness increase on these beats is an important part of the arrangement. It can be seen that much of the impact is lost after extreme compression. In the electronic song, the crest factor has also changed even though the uncompressed version does not have as prominent transients as the other songs.

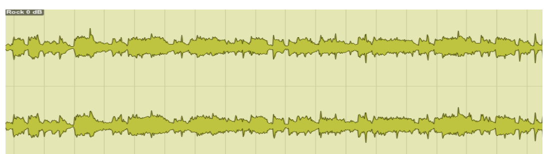

Figure 8. Rock 0 dB Gain Reduction

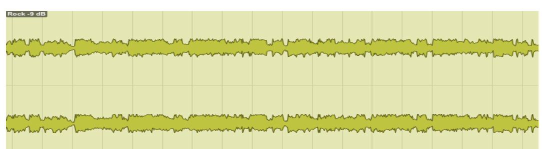

Figure 8.1. Rock -9 dB Gain Reduction

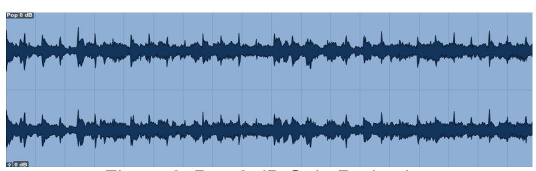

Figure 9. Pop 0 dB Gain Reduction

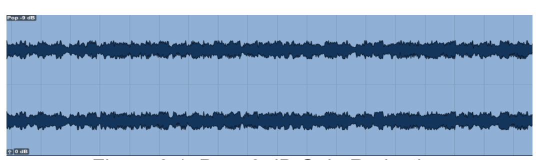

Figure 9.1. Pop -9 dB Gain Reduction

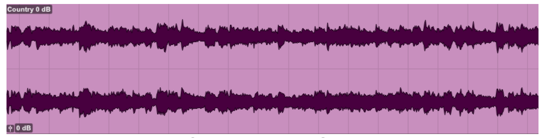
Figure 10. Country 0 dB Gain Reduction

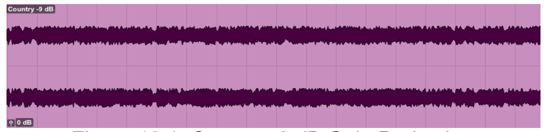
Figure 10.1. Country -9 dB Gain Reduction

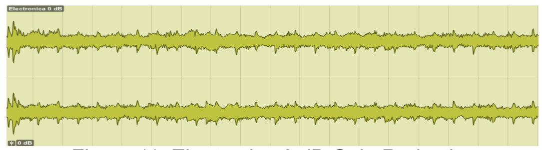
Figure 11. Electronica 0 dB Gain Reduction

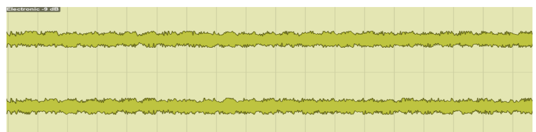
Figure 11.1 Electronica -9 dB Gain Reduction

3.5 Quality Assessment

The test was focused on basic audio quality and therefore the subjects rated the overall audio quality of the different music excerpts (stimuli). The subjects were instructed to listen to the overall sound (not specified further than this) and was told to form an opinion based on their own preference. As a consequence, specific attributes were not asked for.

In order to access the perceived basic audio quality of the stimuli, a two-forced choice comparison test was performed in STEP. The software randomized all stimuli so that the signals (genres) were associated with conditions A and B on the evaluation panel in an unsystematic way making sure no logical pattern of association could be observed.

4. RESULTS

The results for the listening test are listed below. The diagrams show mean values with a 95 % confidence interval (CI) where the lower part (negative) of the scales correlates to the rightmost dynamic reduction category and the top part (positive) to the leftmost. If the top or bottom marker of the CI is above or below 0, the result is significant, proving that one of the two dynamic range reduction categories in the comparison is preferred.

A low CI, where the top and bottom markers for each comparison are close to the mean marker, indicates that the participants are more in agreement with their opinions. The comparison of the dynamic range reduction categories is indicated under each CI in a consecutive order (i.e. N-M is a comparison of None and Moderate, N-S is a comparison of None and Strong etc.).

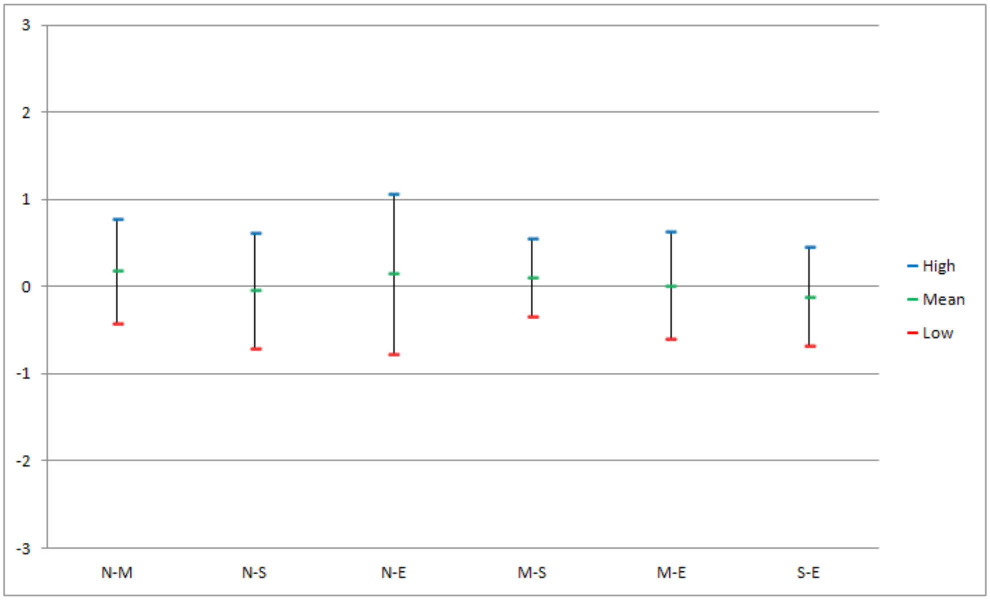

Figure 12. Perceived quality for all signals

The result for all signals combined show a consistent mean value centered around 0 indicating that none of the dynamic range reduction categories were perceived as better when compared to each other. The difference in perceived quality between the None and Extreme categories show the greatest CI and the Moderate and Strong comparison the lowest.

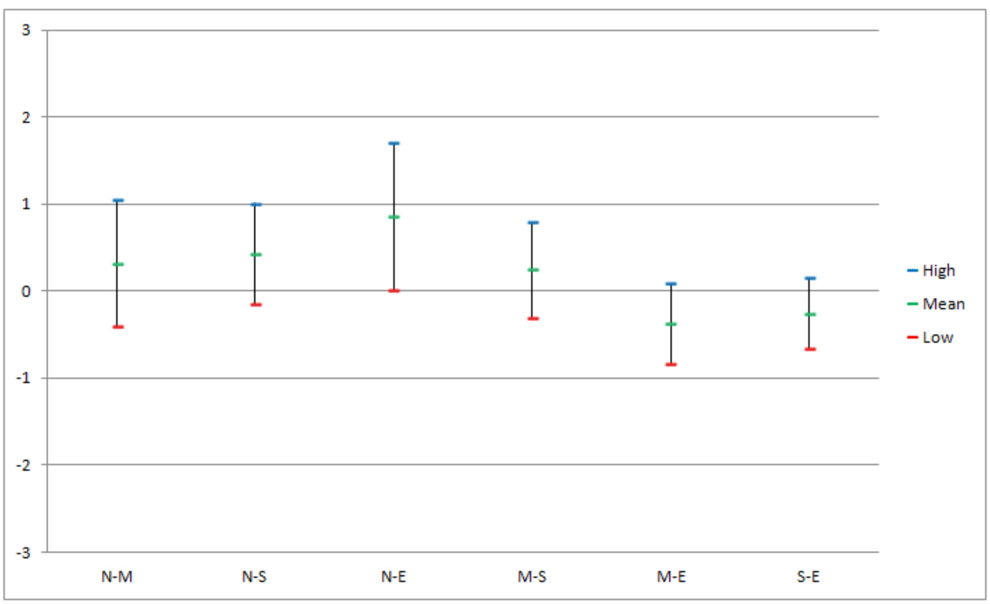

Figure 13. Perceived quality for Rock

For the rock music, the result show that the CIs are spread on both side of 0. The conclusion is therefore that a difference between the categories is not evident. When analyzing the result of the None and Extreme comparison, the red marker is very close to the positive side of the scale. This shows a tendency to rate less compression as being better. It should be mentioned however, that such a conclusion is not reliable since the CI are based on approximations.

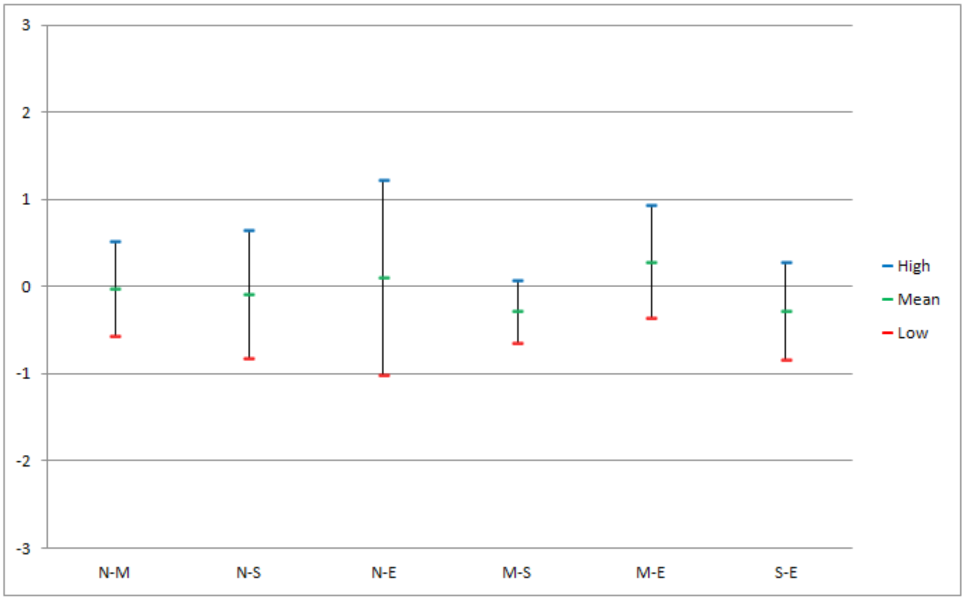

Figure 14. Perceived quality for Pop

For the pop music, the CIs are placed on both sides of 0 in all of the comparisons and for that reason, no significant difference between the categories is found for this genre.

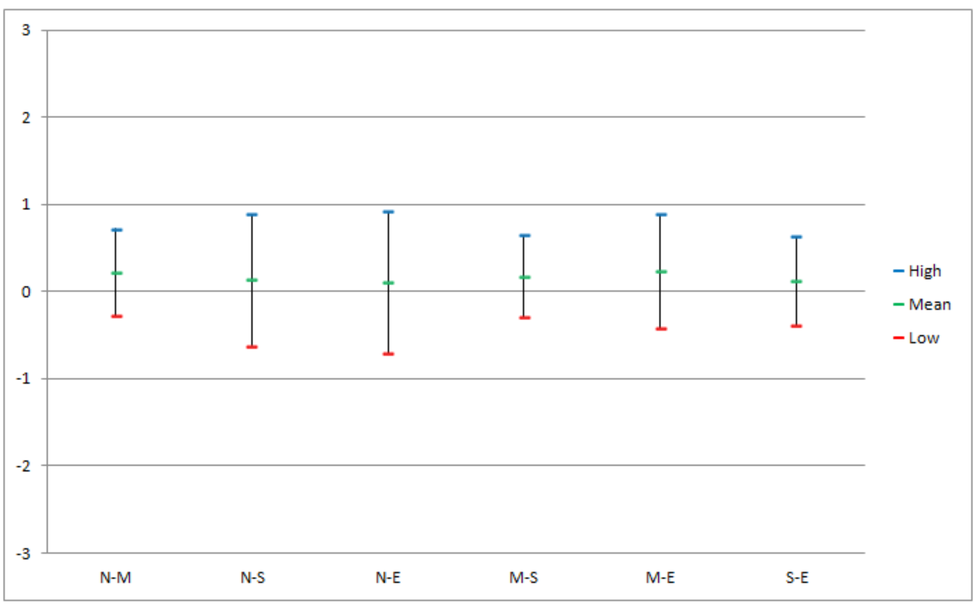

Figure 15. Perceived quality for Country

For the country music, the mean value is again centered around 0, giving the same result as the previous two genres.

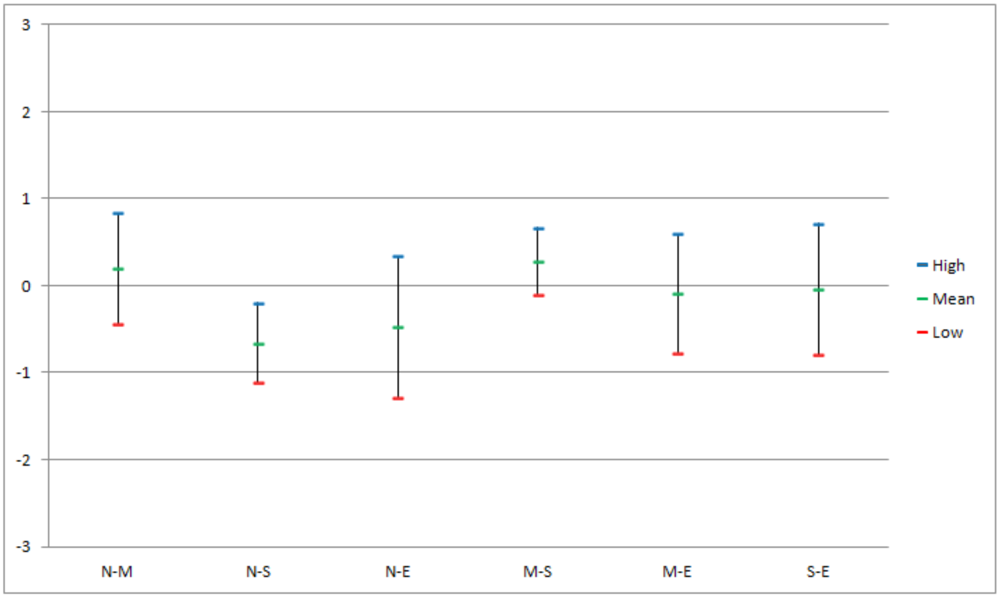

Figure 16. Perceived quality for Electronic

In contrast to the other genres, the result for the electronic music shows that there is a significant difference in the comparison of the None and Strong versions. In the other five comparisons for this genre, the CIs are overlapping 0.

5. DISCUSSION

A total of four different genres of popular music were compressed using four gain reduction categories prior to loudness normalization to -23 LUFS. The results in this test show that popular music with extreme compression does not correlate with a decrease in perceived sound quality. The mean values are close to 0 and this is true for every genre, indicating that the listeners could not clearly hear a difference between the dynamic range reduction categories. In addition, the CI is quite small in many of the comparisons (for every genre), demonstrating some minor evidence for consistency amongst the listeners.

Before the test, it was hypothesized however, that listeners might prefer more dynamic range reduction. This was based on the idea that the subjects in this test actively listened to music during the most intense part of the loudness war era and therefore might be more cognitively trained to listen to highly compressed music. Furthermore, during this period portable music players increased dramatically amongst this generation. These music players are often equipped with poor isolated earbuds and used in noisy environments and therefore the music necessitates a low crest factor in order to be fully heard.

As mentioned in section 2.3, a divergence in loudness perception was found between compressed and uncompressed speech with the same RMS value in [16]. This might raise some questions about the results in this test and whether small differences in loudness perception between the stimuli have been perceived by the subjects. A few subjects reported that they perceived small variations in loudness between the stimuli and as pointed out earlier in this paper, a louder version of the same song is most likely perceived as being "better ". However, the reason behind this observation, which a few of the listeners hesitantly pointed out, is most likely due to the change of the level of each instrument in the mix. The compression increased the Side-signal (or put differently, decreased the Mid-signal), especially for the higher gain reduction categories which created a somewhat different mix. This was evident in all genres but especially noticeable in the electronic and pop music where pads and synthesizers were heard more clearly.

When looking at the result from all signals combined, it is clear that the differences observed were small for all categories. The mean values (represented by the green horizontal lines) and the CIs (red and blue) show that the difference in perceived sound quality between the categories, are in all cases not distinguishable. As a consequence, it could be concluded that dynamic reduction of this magnitude does not result in a degradation of sound quality.

As seen in figure 13, the results show that the participants did not favour any specific version of the rock song. The comparison involving the uncompressed version is closer to the positive side of the scale but since the CIs are still overlapping 0, any conclusions based on this fact can not be made. In figure 14, the results from the pop song is shown. In this case, higher gain reduction seems to be associated with higher quality but since the CIs are spread on both sides of 0 in all cases, an analysis is hard to make in this example too.

This also holds true for the country music as can be seen on the mean values in figure 15. For this genre, it is difficult to find any preference at all for a specific version. Before the test was conducted, it was hypothesized that for the electronic music, higher gain reduction would receive higher scores since the genre is very much associated with high compression and dynamic range reduction is part of the aesthetic of the genre. This idea might be true seeing that the result of the comparison of the None and Strong versions did have a significance. Perhaps this is due to the change of the sound that the limiter introduced. The CIs are however, spread on both sides of 0 in most cases.

5.1 Reliability and Validity

With this result in mind, there are a few details in this study that should be addressed. Each of the songs used in the study are said to belong to different genres but today such a division is not always easy. Many songs can easily be placed in more than one genre. The fact that all signals were mixed together on to the mix bus prior to dynamic reduction should raise some concern as to the result of the overall sound. The reason for this is because the frequency content of the signals interacted before compression which may have an effect on the perceived quality of the stimuli. Another possible mastering solution, utilized in [22], is to compress the mix in stems prior to the mix bus.

Furthermore, the choice of compression method may affect the transparency of the final result. The L2 is a brickwall limiter with a look-ahead function and it can reduce the crest factor to a great extent without introducing distortion and other audible artifacts. However, it was noticed that the overall sound of the stimuli were modified to some extent by the L2. The affects ranged from minor pumping artifacts to distortion when used with more extreme settings. In addition, the functionality is quite limited and more detailed settings could be an advantage in a test related to dynamic range reduction and sound quality since this would perhaps give more reliable data.

The study has however, been limited to one well known limiter with few settings to exclude as many parameters as possible that might otherwise make the result difficult to analyze. The music was choosen to represent specific genres of popular music and the songs were different in style, arrangement and instrumentation and this might explain the result to a greater extent. Furthermore, the participants were choosen to represent listeners with different listening experience. Thus, the study has been controlled in its form and focused towards specific music, involving only a specified group of listeners.

This leads to the conclusion that the results in this test does not prove that loudness normalized popular music is perceived as "better" after being dynamic range reduced with as much as -9 dB. It can not be concluded however, that the opposite is true. Having said that, if a final mix is aimed at a specific distribution platform involving further processing stages such as compression, limiting and/or data rate reduction, there is compelling evidence that some headroom should be left in the final mix [7].

6. FUTURE WORK

It would be particularly interesting to see a study utilizing compression in stems prior to summing on the final mix bus since this would explain the influence of the compression better. Such a study might also include more compressors/limiters. A study that included heavier compression could contribute further by giving results that explain how much compression is needed before most listeners would deem the quality inadequate.

7. ACKNOWLEDGEMENTS

The author would like to thank Jonas Ekeroot for his support of the work, Thomas Eberger for contributing with valuable information and Himlakropp for his music. A special thanks to Ida Bengtsson for additional support and help with changes of the final draft.

8. REFERENCES

1. B. Katz. "Integrated Approach to Metering, Monitoring, and Leveling Practices Part 1: Two-Channel Metering". *J. Audio Eng*. Soc., Vol 48, No 9, 2000, Florida.

2. F. Camerer. On the way to Loudness nirvana - audio levelling with EBU R 128. *EBU Technical Review*. 2010.

3. “Unofficial Dynamic Range Database”. Retrieved 03.01.2012 from: http://www.dr.loudness-war.info/

4. Pleasurize Music Foundation. Retrieved 20.02.2012 from http://www.dynamicrange.de/

5. Justice For Audio. Retrieved 20.02.2012 from http://www.justiceforaudio.org/forum/

6. E. Vickers, ”The Loudness War: Background, Speculation and Recommendations”. Presented at the *129th AES Convention*. November 2010, San Fransisco.

7. F. Rumsey, ”If it´s LOUD, does that mean it´s bad?” *J. Audio Eng. Soc*., Vol 56, No. 6, 2008, June

8. B. Katz, *Mastering Audio: The Art and the Science*. Oxford, UK: Focal Press, 2007.

9. E Vickers. "Metrics for Quantifying Loudness and Dynamics". September 14, 2010.

10. E Deruty. "'Dynamic Range' & The Loudness War". Published in *SOS*, September, 2011.

11. E Deruty. Elements of Popular Music Analysis. January, 2012. Retrieved 27.03.2012 from: http://1-1-1-1.net/IDS/?p=747

12. E. Vickers. “The Non-flat and Continually Changing Frequency Response of Multiband Compressors”. AES Convention Paper from the *129th AES Convention*. 2010, San Francisco.

13. G. A. Soulodre & S. G. Norcross. “Objective Measures of Loudness”. Presented at the *115th AES Convention*. 2003, New York.

14. S. E. Rogers. “The loudness war is fought with (and over) compression”. Workshop at the *131st AES Convention*. 2011, New York.

15. B. Katz. ”Loudness War”. *127th AES Convention*. 2009, New York. Retrieved 15.03.2012 from http://www.youtube.com/watch?v=XCd6MHlo_iA&feature=related

16. M. Stone et. al. “Multichannel Fast-Acting Dynamic Range Compression Hinders Performance by Young, Normal-Hearing Listeners in a Two-Talker Separation Task”. *J. Audio Eng. Soc*., Vol 57, No 7/8, 2009, Cambridge.

17. ITU-R BS.1770-1. “Algorithms to measure audio programme loudness and true-peak audio level”

18. EBU - Tech 3342. “Loudness Range: A measure to supplement loudness normalisation in accordance with EBU R 128”.

19. ITU-R BS. 1770-2. “Algorithms to measure audio programme loudness and true-peak audio level”

20. E. Skovenborg & T. Lund. “Loudness Descriptors to Characterize Wide Loudness-Range Material”. Presented at the *127th AES Convention*. 2009, New York.

21. T. Eberger, Stockholm Mastering (personal communication, March, 2012).

22. W Campbell et. al. “The effect of dynamic range compression on the psychoacoustic quality and loudness of commercial music”. *Internoise 2010*. 2010, Lisbon.

9. APPENDIX

LRA Measurement

Measured LRA Values for 60 randomly selected songs from [5] are listed below. The categories are defined below and discussed briefly in section 3.

- Bad Quality Audio: "*clipped/overcompressed/limited albums where the recording suffer as a result*"
- Only Mildy Affected: "*albums which are brickwalled but have little or no clipping and not pushed to the extent of distortion / crackling / destroying the audio. These albums likely sound fine, however in most cases could have sounded much better.*"
- Bad-Quality Audio: "*albums which maintain maximum audio quality (relevant to the genre) that showcase how good audio can be made with modern tools.*"

$LRA_{hi} = LRA\ highest\ value$
$LRA_{lo} = LRA\ lowest\ value$

"Bad Quality Audio":

- Florence and the Machine – Ceremonials (2011)

  Only if for a night: 10.3
  Shake it out: 8.8
  Never let me down: 8.9
  All this and heaven too: 5.0

- White Stripes - De Stijl (2000)

  You´re Pretty Good Looking: 1.1
  Apple Blossom: 3.5
  White Moon: 13.7
  As Ugly As I Seem: 3.4

- TV On The Radio – Dear Science (2008)

  Halfway Home: 3.7
  Stork & Owl: 2.7
  Family Tree: 9.8
  Lover´s Day: 5.1

- John Mayer - Continuum (2006)

  Belief: 4.9
  The Heart of Life: 6.2
  Bold As Love: 9.7
  I´m Gonna Find Another You: 8.1

- Audioslave - Audioslave (2000)

  Show Me How To Live: 5.5
  Gasoline: 8.2
  Set It Off: 14.7
  Hypnotize: 6.9

$LRA_{hi} = 14.7$
$LRA_{lo} = 1.1$

“Only Mildy Affected”:

- Soundgarden - Telephantasm (2010)

  Hunted Down: 6.6
  Rusty Cage: 2.8
  Spoonman: 3.0
  Black Rain: 5.7

- Billy Joel - Turnstiles (2010)

  Say Goodbye To Hollywood: 4.7
  New York State Of Mind: 12.5
  Prelude/Angry Young Man: 4.7
  Miami 2017 (I´ve Seen The Lights Go Out On Broadway): 18.2

- Rolling Stones - Exile on Main St. (2010 Bonus Tracks)

  Rip This Joint: 1.7
  Loving Cup: 7.7
  Plundered My Soul: 2.3
  Casino Boogie: 4.8

- Eric Clapton - Reptile (2001)

  Got You On My Mind: 3.0
  Find Myself: 3.9
  Second Nature: 3.0
  Superman Inside: 0.9

- Cream - Royal Albert Hall (2005)

  I´m So Glad: 5.5
  Sleepy Time Time: 2.8
  Deserted Cities Of The Heart: 2.2
  Crossroads: 5.1

$LRA_{hi} = 18.2$
$LRA_{lo} = 0.9$

“High-Quality Audio”:

- Marina and the Diamonds – The Family Jewels (2010)

  I Am Not A Robot: 7.2
  Mowgli´s Road: 4.5
  Hollywood: 4.5
  Numb: 7.7

- Porcupine Tree - The Incident (2009)

The Blind House: 14.7
The Incident: 8.0
Degree Zero Of Liberty: 11.3
Circle Of Manias: 4.3

- Bruce Springsteen - The Wild, The Innocent And The E Street Shuffle (1994)

  The E Street Shuffle: 6.3
  Kitty´s Back: 5.2
  Incident On 57th Street: 12.0
  Rosalita (Come Out Tonight): 4.7

- Depeche Mode - Music For The Masses (1987)

  Strangelove: 12.9
  I Want You Now: 17.3
  Nothing: 4.6
  Agent Orange: 9.8

- U2 - The Joshua Tree (2007 Remaster)

  Where The Streets Have No Name: 12.8
  I Still Haven´t Found What I´m Looking For: 3.8
  Exit: 16.6
  Running To Stand Still: 14.1

$$LRA_{hi}=17.3$$
$$LRA_{lo}=3.8$$

Reproduction Levels

The reproduction level for each subject is listed underneath where gain values correspond with the volume on the Motu interface. The following abbreviations are used:

s = sound engineer student
m = music student
o = other listener

1. -17 dB (s)
2. -16 dB (o)
3. -16 dB (o)
4. -17dB (o)
5. -17 dB (s)
6. -17 dB (s)
7. -16 dB (s)
8. -11 dB (o)
9. -12 dB (o)
10. -13 (m)
11. -16 dB (m)
12. -8dB (m)
13. -16 dB (o)
14. -12 dB (o)
15. -8 dB (m)