EFFECTS AND JUSTIFICATION OF LOUDNESS WAR IN COMMERCIAL MUSIC

Anssi Tenhunen
Master's Thesis
Music, Mind and Technology
Department of Music
22 April 2016
University of Jyväskylä

Tiedekunta – Faculty Humanities	Laitos – Department Music Department
Tekijä – Author ANSSI TENHUNEN	
Työn nimi – Title EFFECTS AND JUSTIFICATION OF LOUDNESS WAR IN COMMERCIAL MUSIC	
Oppiaine – Subject Music, Mind & Technology	Työn laji – Level Master's Thesis
Aika – Month and year APRIL 2016	Sivumäärä – Number of pages 26 or 35 with References and Appendices

Tiivistelmä – Abstract

Loudness has been proven and acknowledged to be one of the easiest ways to convey something feeling "better", and thus it has been frequently used technique in sound playback and recordings. However, with the invention of digital recording and playback mediums such as the CD, the limit of how loud sound can go has been met, but music is still being pushed to the limits and even further which can cause audible distortion, which often is viewed as unpleasant. This study tries to find out what is the threshold of acceptable loudness vs unpleasant loudness. A test was conducted where 25 test subjects were played 100 clips of volume matched clips of the same recording with different sort of tonal and dynamic range processing. Results showed that equalizer was perceived and disliked when compared to the unprocessed clip in linear fashion to the amount of processing, and participants either did not hear a difference or had more tolerance to limiters, but after a clear threshold point the results jumped to similar linear results. Thus using a limiter does give the mastering process and in that sense, the loudness war, a significant advantage when trying to achieve loudness until you reach the audible point of distortion. But as loudness normalization is becoming more common, the advantage gained from pushing the loudness further is nulled.

Äänekkyys (engl. loudness) on todistetusti ja tunnustetusti yksi helpoimmista tavoista välittää "paremmuuden" tunnetta, ja täten sitä on usein käytetty äänentoistossa ja äänitallenteissa. Kuitenkin digitaalisen tallennuksen ja toistoalustojen (kuten CD:n) myötä raja kuinka kovalle äänen voi laittaa on täyttynyt, mutta musiikkia edelleen työnnetään rajoja vasten ja tämä voi aiheuttaa äänenvääristymistä eli säröytymistä, joka usein katsotaan epämiellyttävänä. Tämä tutkimus yrittää selvittää, mitä on kynnys hyväksyttävän äänekkyyden ja epämiellyttävän äänekkyyden välillä. Testiin osallistui 25 koehenkilöä jotka kuuntelivat 100 äänekkyydeltään normalisoitua äänileikeparia erilaisilla tonaalisuuden ja dynamiikan käsittelyllä. Tulokset osoittivat, että taajuuskorjainkäsittely havaittiin ja karsastettiin verrattuna käsittelemättömään leikkeeseen lähes lineaarisesti käsittelyn määrän kasvaessa. Oallistujat joko eivät kuullut eroa tai heillä oli enemmän toleranssia dynamiikkaprosessoreille, mutta vastaukset hyppäsivät selkeästi vastaavanlaisiin lineaarisiin tuloksiin toleranssi-kynnysarvon ylittyessä. Täten dynamiikkaprosessorit antavat masterointiprosessille prosessissa, ja siten myös äänekkyyssodalle, merkittävän edun, kun yritetään saavuttaa äänekkyyttä ennen kuultavissa olevaa säröytymistä. Mutta koska äänekkyyden normalisointi on yleistymässä, dynaamisen alueen minimoimisella saatava etu mitätöityy.

Asiasanat – Keywords MUSIC, LOUDNESS, PERCEPTION, LOUDNESS NORMALIZATION	
Säilytyspaikka – Depository	
Muita tietoja – Additional information	

JYVÄSKYLÄN YLIOPISTO

ACKNOWLEDGEMENTS

This paper is dedicated to my father Lassi Tenhunen (*5.2.1955–17.7.2013*), my mother Irma Tenhunen who has helped me get through these hard times and pushed me to finish this paper, and Miia Jalava for inspiration in life. I would like to thank everyone who participated in my thesis research project, all of my colleagues in the MMT programme, my supervisors Marc Thompson and Geoff Luck. Major thanks to Sarah Faber for proofreading my thesis and Wylli Groove Federation for letting me use their music as the listening stimuli.

CONTENTS

1 INTRODUCTION

My thesis topic is "Effects and justification of loudness war in commercial music", and the focus topic is so-called "commercial music." The definition of "commercial music" used here is a piece of music that has been recorded and then distributed to the consumer market either in physical, digital, streaming, broadcast or similar format or method. "Loudness war" is a phenomenon in commercial music, in which recordings are slowly getting louder year by year by means of dynamic range compression.

Loudness is measured in decibels, a logarithmic scale that measures both peak and average sound levels of a recording, signal or sound pressure level. Loudness war is both relatively new and old phenomenon in the music industry, as loudness is proven to be one of the easiest ways to increase pleasantness in music (Shepherd, April 2014). But loudness is also a double edged sword; "Louder" is often viewed as a positive thing, but "too loud" is considered negative. In sound recordings, most of the loudness is often done in mastering, the final phase between the music making process and the physical duplication of the record.

In the heydays of vinyl records, the loudness maximum was set by the physical limitations of the vinyl format. The needle would jump out of the groove if the song was mastered too loud, thus the loudness only increased by mere 4 decibels in 40 years (Katz, 2002, p187-188). With the invention of CD, an entirely digital format, the medium wasn't the bottleneck anymore. As the maximum peak loudness was reached almost instantly at the dawn of CD, the next step was to increase the average loudness using different algorithms to maximize the perceived loudness of the recordings. Katz also mentioned (p 99-133) this was done mainly using limiters (dynamic range compression) and equalizers (tonal shaping). But this once again had drawbacks; when pushed too far, often audible distortion was added to the recordings, which was viewed as negative by more sophisticated listeners. With various optimized digital algorithms, CD loudness increased by 20 decibels in 20 years (Katz, p187), a dramatic increase.

There is one effective way to counter loudness war, and that is loudness normalization. When normal level measurement uses peak value as measurement, loudness normalization uses the perceived loudness as the measurement. For example you might have two songs: a dynamic jazz tune and a very compressed rock tune. Their peak value might be the same, but perceived loudness might be 10 decibels apart. What loudness normalization does is reduce the level of the louder piece by 10 decibels, so both sound equally loud

Similar effects were often encountered in television broadcasting, when commercials were much louder than the actual program. In March 2013, I attended an Audio Engineering Society seminar on broadcasting loudness, where the then new R128 broadcasting loudness standard (EBU R128, 2012; EBU – TECH 3341-3344, 2011; ITU BS.1770, 2011) was introduced and explained to large group of working professionals in the audio industry. R128 was invented to reduce, and eventually remove, loudness war from broadcasting industry, as it became mandatory later that year in television broadcasting in Finland.

The main problem still remains in consumer applications, as there is no standardized loudness normalization method applied to all consumer applications. Although many services such as Spotify, YouTube, Apple Music and countless others are already implementing various loudness normalization methods, as long as there is no de-facto loudness normalization method that always keeps the perceived volume at the same level, it can encourage loudness war.

This thesis aims to research if, when the loudness factor is removed, people actually prefer this louder, hyper-compressed music over a more dynamic version of the same piece of music. Research was conducted with 25 participants (15 female, 10 male) who listened to 100 pairs of sound clips. The participants were asked to rate both difference in audibility and preference between the two clips, where one clip was processed using dynamic range manipulation or tonal shaping, and the other was not manipulated. Research was done as a group study in a controlled listening environment, and was exploratory in nature, so hypotheses of the results were not set beforehand.

The main research questions are as follows:

- Is loudness war justified? Specifically, do people actually prefer “louder” music over less compressed, more dynamic music, or is loudness war only hurting the recordings?

- Does loudness standardization, by removing loudness from the listening factor, change listeners opinion on the sound recording?

The auxiliary questions added during the research:

- Are equalizers or limiters better tools for achieving pleasant loudness?

- Does the order (AB vs BA) of the samples matter?

2 LITERATURE REVIEW

In 2012 there was very little written about the subject of loudness war in recorded music in the scientific community. As the phenomenon of loudness war is relatively new, to find out what there was written about loudness war, I had to delve into it's timeline, used tools, formats, definitions, industry professional thoughts and reactions, and initially finding this material was not easy. At the same time, one of the strengths of the literature was that, by being so new, it is most likely valid and current.

To research loudness war more closely, I first tried to tackle it from the industry and technical point of view instead of psychological side, so I attended an Audio Engineering Society's (AES) loudness-seminar at the Finnish National Broadcasting Company Yleisradio (YLE) premises and Chartmakers mastering studios in Helsinki on March 12th 2013. There was a lot of current information on the subject, such as how the R128 broadcasting loudness standard (see page 11-12 for more detailed description) is becoming mandatory on television broadcasts in Finland by the end of 2013.

We were presented with visions of how R128 (or similar future variant, such as *ReplayGain,* the US standard ATSC A/85, or the Japanese standard ARIB STD-*B32*) might affect the music industry and the future recordings and mastering levels. We were presented insights into the industry professionals' view on this subject and heard differences between hyper-compressed and less compressed material by German metal band Rammstein and some other artists. To my surprise, the differences were not very big, only very subtle. Thomas Lund, the HD Development Manager at TC Electronic, provided some extra material to read on the subject, and most of that material was from AES.

2.1 LOUDNESS TRENDS IN MUSIC

Vickers (2012) mentions in his AES convention paper that commercially distributed music is getting louder and louder every year. This phenomenon is called “loudness war” or “loudness race”. It is usually often justified with the catch phrase “louder is better” (Frink, 2000). He and other audio professionals, such as Lund (2012) refer to this sort of material as "hyper-compressed" or over-compressed. Hyper-compression implies that the overall dynamic range changes of the recorded piece is minimized to maximize the perceived loudness. To quote Vickers:

> If louder songs have a commercial advantage, leading to a loudness war in which everyone hypercompresses (effectively negating the advantage), and if the resulting lack of musical dynamics ends up hurting the overall music industry, this would be a classic social dilemma. But if players use hypercompression because of a mistaken belief about the commercial advantage of louder songs, this is no longer a dilemma, merely a tragedy. (2012, p. 17)

In a research study done by Spanish researchers (Serrá et al. 2012), they compared 464 411 pieces of music in a computerized matter ranging from years 1965 to 2005 from variety of popular genres including rock, pop, hiphop, metal and electronic. They discovered that the loudness median seemed to have risen from -22 dBFS to -13 dBFS ("decibel, Full Scale", see page 12) in the measured time period, which is a 9 dB increase and an average increase of 0.13 dB per year.

Outside the Audio Engineering Society (AES) the loudness war has not had much research in the scientific community, as it has not really existed in recorded music until the rise of popularity of the CD in the early 1990s. Vickers also notes that "while vinyl LPs increased in level by perhaps 4 dB over 40 years, the average CD levels went up by almost 20 dB in 20 years". Bob Katz (2002, p. 187-188) has similar views in his book.

On the psychological side, Parker et al. (2012) have studied the effects of expectations on loudness and loudness difference using pure tone testing, where they subjected the listeners to believe the coming sounds were going to be loud or quiet, and they found out that if people were expecting loud sounds, they rated the sounds quieter, and vice versa, and Siegel and Stefanucci's (2011) pure tone study also implied that the expectations can affect the perception of sensory intensity, which also confirms this. The results were related to Parducci's (1965) range-frequency theory and Helson's (1964) adaptive-level theory, even though the some of the experiment results were contrary to adaptive level theory. Fenton and Wakefield (2012) wrote about the objective profiling of perceived punch and clarity in produced music, where they used two stimuli (songs by Nickelback and Sugababes), and compared how people perceived clarity and punch in song that has more dynamic (Sugababes) and less dynamic range (Nickelback) and noticed a degree of correlation observed between the subjective test scores and objective Inter-Band-Relationship (IBR) descriptor.

One of the early “culprits” to loudness war was the Israeli signal processing technology manufacturer Waves Audio Ltd's “L1 Ultramaximizer”, that came in to the market in 1994. Amir Vinci, the product manager of Waves Audio, said in an interview (KVR Audio, 2013):

> The first big hit was the L1 Ultramaximizer, which was an amazing time saver for mastering engineers who used to have to zoom in and use a virtual pencil to chop off outstanding peaks in order to reach optimal level without clipping. The L1 made the process so fast and painless and the results far louder then they could achieve before and also better sounding due to a small "lookahead" making the transient of the dynamics process smoother.

2.1.1 Loudness war in broadcasting

Loudness war was a much bigger issue in both radio and television broadcasting in the past. Somich (2009) noted in his paper, that loudness war in general is not a new phenomenon. In the 1930-1960s there was a similar loudness war called “Modulation Wars” with AM radio stations, where the radio stations tried to make the signal as clean and loud as possible by different means of dynamic range compression. Vickers (2012, 3) notes that the term “loudness war” may have been coined by Robert Orban, when he used the term in his 1979 article discussing excessive compression and limiting for FM radio broadcast.

Adriaensen explained at the 2011 Linux Audio Conference presentation, that “radio listeners and TV viewers hate it when they have to adjust the volume all the time”. A similar problem would occur with music listeners, if they listened to music that has drastic difference in loudness between recordings. Loudness normalization or loudness standardization in theory could void the loudness race. Television broadcasts had a similar problem that the commercials or “in-between programmes” were significantly louder than the actual programme (EBU R128, 2012).

The loudness race does not exist anymore as widely in television and radio broadcasting anymore, because of the current standardization, that recommends playback volume of -23 LUFS (Loudness Unit Full Scale, a weighted decibel unit set to a reference scale defined in ITU BS.1770 standard) programme loudness and true peak volume of -1dBTP (decibel True Peak, four times oversampled peak level, also defined in the ITU BS.1770). In Finland use of this standardization became enforced for TV broadcasting in 2013, and has already been applied in several other European countries.

Travaglini, Alemanno and Lantini (2012) were studying the listening comfort zone in broadcasting, and the study indicates that comfortable Loudness Range (LRA, the difference in loudness level between the soft and loud parts of a programme in decibels) is fairly big: “the average preferred LRA value is 17 for both stereo and 5.1 surround sound” (p. 8) and that people preferred softly/mediumly compressed movie trailers over hyper-compressed ones:

> If we look at the selections made for the TRAILER group, we see that people tend to prefer the medium dynamic soundtracks as well: NARROW was chosen 22.7% of the time, MEDIUM 46.4% and WIDE 30.9% of the time. This suggests that a medium dynamics compression is perceived as comfortable by the average listener whilst tracks with soft compression are slightly preferred by 1 person out of 3. It also indicates that usually over-compressed content are not perceived comfortably. (p. 8)

2.1.2 Consumer market loudness normalization

In consumer applications a loudness standard such as R128 or A/85 does not yet exist. There is a proposed loudness normalization standard by David Robinson (2001) called “ReplayGain”, that suggests use of same headroom of -14dB RMS sinusoidal wave, but it or other standards have not yet been widely accepted by the consumers, manufacturers or developers.

Products supporting ReplayGain have already been released by Dolby Laboratories (Wolters, Mundt & Riedmiller, 2010). Some companies have already incorporated similar propitiatory features in their own music player software, such as Spotify and Apple iTunes. Vickers (2012, 20) also notes in his paper that loudness normalization must become mandatory for it to become effective:

> Loudness normalization may not necessarily solve the problem unless its use becomes widespread. If loudness normalization is merely a selectable option (particular if the default is “off”), or if it is unavailable on a significant number of devices, there may still be some pressure to obtain a perceived competitive advantage by increasing the loudness during mastering.

In 2014, Apple launched iTunes Radio that had loudness normalization turned on by default and couldn't be turned off. Shepherd wrote about it in October 2014 under the title “Has the loudness war been won?”, and quoted on Katz's comment on the subject “The debilitating loudness war has finally been won… The last battle will be over by mid-2014”. Apple aquired Beats Music in 2014, and both Beats Music and iTunes Radio were discontinued and integrated to Apple Music service in January 2016 (TechCrunch, 2015). Shepherd also noted in March 2015 that Google's video streaming service YouTube started applying loudness normalization to their service.

2.1.3 Mastering

Bob Katz's book *Mastering Audio* (2002) defines in the introduction that “mastering is the last creative step in music production, the bridge between mixing and replication.” (p. 11) Mastering is usually the phase where the records are made loud, and Hamidovic (2012, p105) says in his book that this should already be taken into consideration while mixing:

> If you’re aiming for your mix to be LOUD, it helps to mix the bulk of it through a rough mastering chain. [...] as great as mixing at a low level, with all the headroom in the world is, you have to concede that your clients will likely want the master somewhat smashed in this day and age.

Finnvox mastering engineer Minerva Pappi on the other hand explains in her interview that “I try to make records loud in a pleasant way, and then equalizers are more important tools than compressors/limiters” (Visio 3/2008, own translation).

Apple Computers have guidelines titled “Mastered for iTunes” (2012). It explains how mastering engineers can maximize the audio files' fidelity when converted to file formats used in the Apple iTunes store.

2.1.4 Hypercompressed audio and radio playback

One drawback of hyper-compressed audio is that it is often played back on radio stations that apply their own processing to the signal. Both Katz (2002, 272) and Vickers (2012) quote Robert Orban's 2001 article “What happens to my recording when it's played on the radio?” It shows that hyper-compressed music actually sounds quieter on the radio, which negates the need for hyper-compression.

On the other hand, mastering engineer Ian Shepherd (2012) noted that some new remasters have been more dynamic instead of being more compressed. He showed an example of Green

Day 2012 remaster of the album “American Idiot” that is more dynamic than the original 2005 album, but acknowledges that there is not a massive difference in sound. Shepherd also noted in his April 2014 video that increases of a mere 0.5 dB in volume can increase the pleasure factor of a sound clip.

2.2 LOUDNESS DEFINITIONS

As a lot of the thesis has to do with loudness, it might be a good idea to explain it a bit. Loudness is used to measure the amplitude of sound. Loudness is a subjective measure, and thus a “very complicated matter” (Olson, 1972, p1) of which there are several different measurements and scales. Nave (2000) writes “Sound loudness is a subjective term describing the strength of the ear's perception of a sound” and that “loudness is not simply sound intensity.” Loudness is affected not only by sound pressure level, but also by frequency, bandwidth and duration of sound. Shepherd mentioned in his April 2014 video-article that increasing loudness by mere 0.5 dB can increase the perceived pleasantness.

Almost all loudness and sound intensity is measured with logarithmic decibel scale (dB), with different weighting. Weighting is done using filtering to reduce measurement anomalies. For example to measure room loudness a sound pressure level SPL A-weighted (IEC 61672-1, 2002, p31) scale is often used. It has a low frequency filter so that the high intensity bass frequencies do not skew the measurement results.

Loudness should not be confused with sound intensity or sound power. Sound power is the rate of sound energy measured in Watts (W), and sound intensity is the sound power per unit area and measured in W/m2.

2.2.1 Decibel (dB, dBFS etc), SPL and RMS

Decibel is a logarithmic unit that expresses the ratio of two values. In electrical circuits, change in voltage ratio by a factor of 2 (-50% or +200%) approximately corresponds to a 6.02 dB change in level.

Decibels are essentially measured in three different realms: physical, electronic and digital. dBFS is short for “decibel, full scale” and it is used to measure digital signals. dBv is short for “decibel, voltage” and is used to measure voltage in electronics. SPL is short for “Sound pressure level” and is used to measure physical loudness intensity. (Matthews, 1999, p. 73).

dB RMS is short for “decibel root mean square”, a statistical way of calculating unweighted loudness. In electronics, RMS is defined as “the most common mathematical method of defining the effective voltage or current of an AC wave” (Rouse, 2005). For example the RMS is 0.707 times the peak value for a sine wave and 1.000 for a square wave. In the digital realm, dBFS is used to measure the peak values, and dB RMS (or other weighted scales) is used to measure loudness.

2.2.2 Perceived loudness: Fletcher-Munson curves, Equal Loudness Contours and Phon

Matthews (1999) has different explanations for loudness. He also does acknowledge that level increase of 6 dB equals to twice as loud (p. 76), but when perceived by the human ear, “somewhere between 5 and 10 dB difference in intensity is twice as loud” (p. 77). Ears have non-linear frequency responses in loudness perception. This non-linearity is called “Equal Loudness Contours”.

The Equal Loudness Contours were established and determined in 1933 by Bell Laboratories engineers Harvey Fletcher and Wilden A. Munson (1933, p. 394). This is why they are also often known and referred as “Fletcher-Munson curves”. The curves have been refined and redefined in ISO 226 standard (2003) visualized in Figure 1 in red, and are slightly different than the original Fletcher-Munson curves as visualized below in Figure 1 in blue. Equal Loudness Contours also are level and frequency dependent. A level of 20 phons curve compared to level of 60 phons is demonstrated in Figure 1 in green.

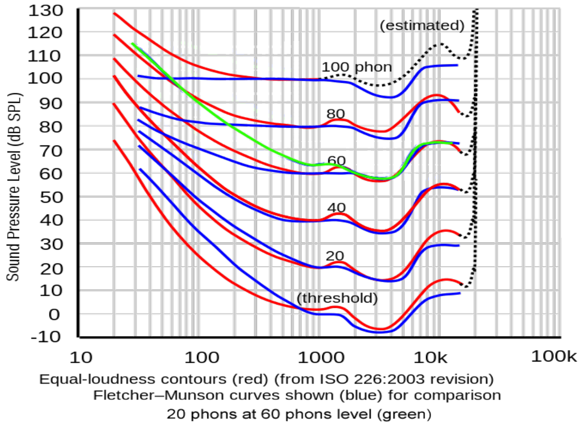

FIGURE 1: Fletcher-Munson curves (blue) vs Equal-loudness contours (red) and 60 phons vs 20 phons (green). Modified from public domain picture by Lindosland (2005), Equal-loudness contours, wikipedia

Sone and phon are units of loudness level for pure tones. A sone is 1 kHz tone at 40 phons. Phons and Sones are defined in ISO 532 (1975), and Wolfe (2010) defines phon as such:

> Phon is a unit that is related to dB by the psychophysically measured frequency response of the ear. At 1 kHz, readings in phons and dB are, by definition, the same. For all other frequencies, the phon scale is determined by the results of experiments […] judged its loudness to equal that of a 1 kHz signal

2.2.3 Loudness Normalization, R128 and LUFS/LKFS

Loudness normalization is a method where the volume level is normalized from the perceived loudness instead of the peak values of the signal. The main benefit with Loudness normalization over peak normalization is that perceived loudness will stay more consistent, whereas with peak normalization loudness will vary, when with peak normalization it happens vice versa, as demonstrated in Figure 2.

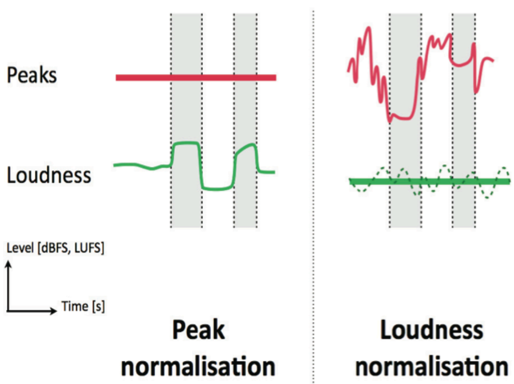

FIGURE 2: Peak normalization vs Loudness Normalization (EBU Tech 3343, 2012)

For broadcasting, loudness normalization exists. R128 is a new broadcasting loudness standard that is used in the European Union. It is based on the International Telecommunication Union's (ITU) "Algorithms to measure audio programme loudness and true-peak audio level" paper, also known as ITU BS.1770. There are similar standards around the world based on the ITU BS.1770 (TC Electronic, 2015), such as ATSC A/85 (United States), ARIB TR-B32 (Japan), OP-59 (Australia).

R128 defines LU (Loudness Unit) and defines recommendations for program loudness, with the main focus being that Programme Loudness Level shall be normalised to a Target Level of -23.0 LUFS (Loudness Unit Full Scale) with Target Level permitted deviation of ±1.0 LU (EBU R128, 2012, p. 4).

LU is a K-weighted measure. K-weighting has a 12dB/oct highpass filter at 100 Hz and 3dB high shelf boost at 2 kHz for measurement (EBU Tech 3343, 2012, p. 9). LUFS has an alternative name LKFS (Loudness K-Weighted Full Scale), but they are synonymous.

2.3 RESEARCH QUESTIONS

Taking the above literature into consideration, the current study aims to solve Pappi's claim that equalization is a better way to achieve more pleasant loudness than limiters, and Parker et al's hypothesis that expectation has a role in the order of the stimuli, and most importantly, whether loudness normalization can make loudness war obsolete when loudness factor has been removed and the stimuli has been adjusted to equal volume.

3 RESEARCH METHODS

The research method implemented measures of perceived difference and preference. Since Pappi mentioned in her interview that equalizers play a more important role in perceiving loudness than limiters, tonal shaping was included in the test. Participants in the current study listened to short clips of music instead of pure tones, to measure perceived loudness and preference, as music compromises of complex tonal structures (Hudson, 2011, p. 8). Many of the pure tone-based loudness perception tests (Parker et al. 2012; Siegel & Stefanucci, 2011) were not the most relevant resources for the current study compared to music perception papers (Travaglini, Alemanno & Lantini, 2012; Fenton & Wakefield, 2012).

However, both research fields provided good indications for research methodology that could be used in similar tests. Both used either headphones or speakers for the listening test. In addition, a survey similar to Siegel et al (2011) was used, but instead of 1-100 scale, a Likert scale of 1-5 was used for preference and 0-5 for difference, with 0 being “no audible difference”. Blank answering sheet included in Appendix 1.

Each sound clip had two questions: Preference between two clips (A and B) and audible difference between the clips from “no difference” to “clearly audible difference”. The pilot and experimental group participants had a paper survey with the same questions. In the beginning of the questionnaire, the participants were asked if the sound system was at a comfortable level, and participants were not able to control the volume during the experiment.

3.1 Participants

A small pilot group of 4 persons was used to test the audio changes against extreme subtlety obviousness. The experimental study had 25 participants (10 male, 15 female) from varying backgrounds. The listening was done at the Jyväskylä University's motion capture studio located in the basement of Musica building in October, 2015. The listeners were in groups of 2-5 persons, sitting in a row of chairs approximately 2 meters away from a pair of Genelec 8030A speakers.

3.2 Stimuli

The stimuli were from the debut album of Helsinki-based band Wylli Groove Federation (http://www.wylligroove.com). It was recorded and mixed it in 2012 at Finnvox studios in Helsinki, and permission was acquired from the band to use the material for this thesis study. Two bars of the instrumental intro from the song “Shout Out For Love” were used and the clip was from the pre-mastered mix, approved by the band.

The survey contained 100 different versions of the same song processed in different ways. The unprocessed clip (A) and the processed clip (B) were played in random order of AB or BA. The A/B comparisons were done with differing amounts of dynamic range compression and different tonal shapings.

For dynamic range attenuation of -1 to -20dB was used with Waves L1 and L2 Ultramaximizers, and five instances of Waves Renessaince Compressor in series. For tonal shaping, a Waves Q10 Equalizer with +-1dB to +-10dB at 100 Hz, 1000 Hz and 10000 Hz was used. See Appendix A for full processing list and Appendix B for example screenshots.

The clips were level matched to the unprocessed version using the AudioSuite Trim in Avid Pro Tools 10 software. The clips were faded in and out by 500ms and they had a 1 kHz sine wave beep in between to indicate the change of clip. Clips had 10 seconds of silence in between, so the participants had time to write down their answers.

4 RESULTS

The surveys gathered total of 4800 answers, that were transferred from paper to digital format, sorted and analyzed. Result analysis, sorting and graphs were done using Apache OpenOffice, Smultron, Microsoft Excel, MathWorks Matlab and Adobe Photoshop.

As the order of the data was mixed in the survey, it needs to be clarified that in context of this text when talking about the clips, A is always the processed clip and B is always the unprocessed clip. In the preference graphs 1 means preference to A, 5 means preference to B and 3 means no preference to either clip. In the difference graphs 0 is “no audible difference” and 5 is “major audible difference”. Captions have been added to the graphs to clarify this.

4.1 CONTROL TEST

To test participants against bias, the test had six unprocessed-unprocessed (BB) pairs. In mean difference testing (M = 1.25, SD = 0.22), the ratings data indicated participants were mostly listening and performing the test correctly, with majority of the answers being 0 or 1. The mean preference data (M = 2.97, SD = 0.21) also clearly indicated no preference to either clip.

4.2 DIFFERENCE

To test Parker et al's claim that expectation would affect the results, questions 45-53 and 54-62 were the same stimuli, but in different order (AB vs BA). When testing for correlation, the r-values were 0.69 and 0.79 with p <0.01, so at least with this stimuli, the differences in results were so small, that it was concluded the order of the stimuli did not matter, and there was no need to do further testing in different order.

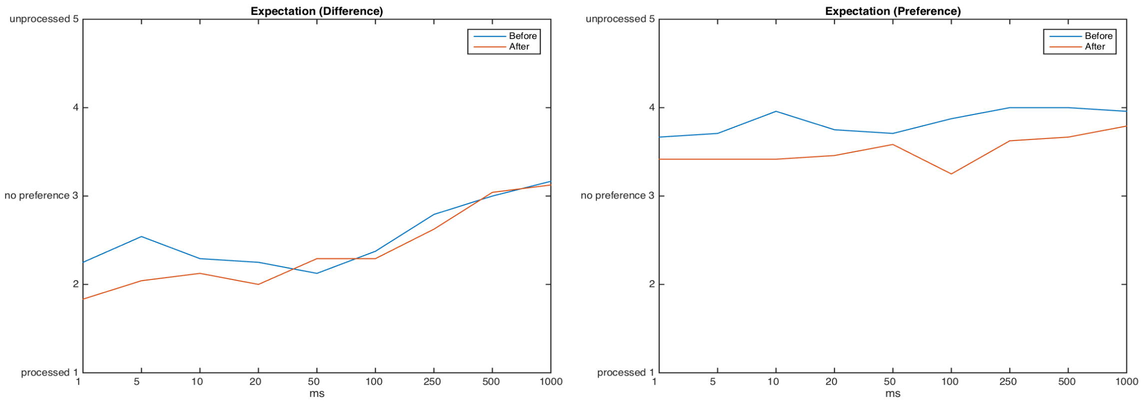

FIGURE 3: Expectation, difference (a) and preference (b)

4.2.1 Equalization

To test Pappi's claim that equalization causes a more pleasant result than limiters when trying to achieve loudness, the data between all three equalizer processing bands, 100 Hz (M = 2.09, SD = 1.16), 1 kHz (M = 3.01 , SD = 1.25), and 10 kHz (M = 3.16, SD = 1.08), was compared. All of the EQ processing had similar trends in results. When going towards the extremes, the results made a parabola shape as shown in Figure 4a and 4b on all three frequency bands. The difference increased almost linearly when deviating from the unprocessed samples.

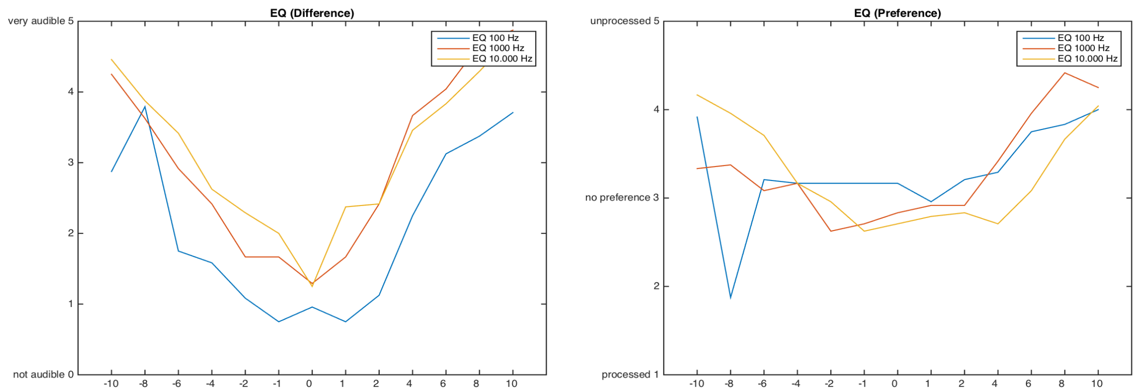

FIGURE 4: EQ processing, difference (a) and preference (b)

There were two clear anomalies in the data, and those were the dips in difference at -10dB and preference in -8dB results in 100 Hz, as can be seen the blue line in Figures 4a and 4b. They were one of the first questions in the questionnaire, so the participants most likely were not familiar with the test yet, which most likely caused this dip that also showed up in the correlations ($r(13) = 0.19$, $p = 0.53$). When ejected from the calculations, the entire 100 Hz data correlated highly ($r(11) = 0.95$, $p <0.01$). Overall, the other two bands were also highly correlated (1 kHz: $r(13) = 0.95$, $p <0.01$, 10 kHz: $r(13) = 0.83$, $p <0.01$).

4.2.2 Limiter

With limiters, there were two tests; first a comparison of Waves L1 with 1ms release time ($M = 2.80$, $SD = 1.42$), Waves L1 with 100ms release time ($M = 2.70$, $SD = 1.23$) and Waves L2 with automatic release time ($M = 2.48$, $SD = 1.17$). The second was comparing -10 dB of gain reduction with variable release time.

When comparing the limiter data with -10 dB of gain reduction with different release times to each other, the curve was almost increasingly linear in difference (Fig 3a), but preference stayed about the same (Fig 3b). An unforeseen effect in the gain reduction stimuli was seen around -11 dB of gain reduction with this stimuli. At this level, results shift dramatically in difference and preference, indicating a -10 dB gain reduction tolerance. As this measure was used to test ordering of the stimulus presentation, the effects were negligible in the analysis (see page 15-16).

All three had very similar trends, including a very high number of “no preference” answers and low audibility difference until the data went to -12 dB of gain reduction, after this point the data moved very linearly to favor the unprocessed sample, as seen in Figure 5. Although there were a lot of “audible difference” answers starting as early as -6 dB gain reduction on all three processor settings, there clearly was a tolerance to some gain reduction until it became unbearable around the -12 dB mark with this stimuli. All three had highly correlated answers between difference and preference (L1 1ms: $r(11) = 0.97$, $p <0.01$, L1 100ms: $r(11) = 0.96$, $p <0.01$, L2 ARC: $r(11) = 0.95$, $p <0.01$).

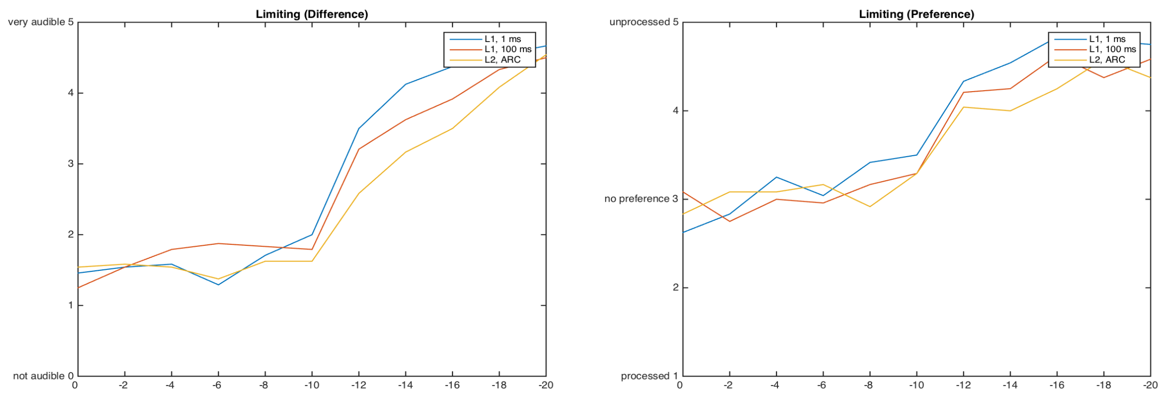

FIGURE 5: Limiters, difference (a) and preference (b)

4.3 Serial compression

The serial compression data (M = 1.36, SD = 0.51) unfortunately had a design flaw from the beginning, was evaluated as not reliable, and thus discarded.

4.4 CORRELATION

At first one of the goals was to find out of any trends or correlations between the limiter and equalizer answers. It was quickly realized, that since the values and scales were different between both limiters and equalizers, and difference and preference graphs were measuring different things, crosschecking was not very reliable. There were also surprisingly few correlations between the individual answers on difference and preference (most r-values hovering between 0.1 and 0.5), but when doing trend calculations for all answers, there were very significant correlations between difference and preference (r >0.8 and p <0.05) on both equalizers and limiters, as presented above.

When comparing the correlation between equalizers and limiters, a cut-off point had to be established because the equalizer had a parabola results and limiters had linear results after the threshold. When correlating the +0 to +10 dB equalizer results to the -8dB to -20 dB gain reduction limiter results, the correlations were quite high between all bands ($r(7)$ = 0.90, p <0.01), thus resulting in similar end results.

5 DISCUSSION

To go back to the beginning of the thesis, to answer the question “Is loudness war justified” is not a simple “yes/no” matter. It can be answered with both “yes” and “no”, but mostly “maybe”, as it depends on the situation. But this study still fortifies current two main views on loudness war: If the listener has control over the volume knob and/or if the loudness has been normalized, they will set it to their preferred level, thus the loudness war in this scenario is not justified. If the listener has no control over the volume knob and the loudness has not been normalized, the loudness war is justified.

Since loudness normalization has been adapted already in television broadcasting and normalization is slowly being adapted to commercial applications, where you do not need to touch the remote to turn down the volume during loud parts, it has been successful in defeating the loudness war in those fields. In those cases, loudness war is no longer justified, but only destroying the music. But what about in other consumer applications, where this some sort of normalization is not in effect? There loudness war, unfortunately, has a small justification, as differences as minor as 0.5 dB can achieve more pleasurable listening experience. And if the mastering engineer can increase the loudness of a clip by over 10 decibels without negative repercussions, as with the test stimuli of this experiment, the person responsible for the sales of the music piece in question will most definitely try.

To answer the question “does loudness normalization affect opinions on hyper compressed music”, the answer clearly is “yes”. When the loudness difference of the limited material was removed, the test subjects' opinions were clearly negative when going to the extremes, that would be gained from hyper compressed material. If it would not have had any effect, the preference would have hovered around 3 in the preference tests.

To answer the question “does the order of the clips matter”, when tested, the difference between the clips in reverse order was so small, that it did not affect the results in any major way.

To answer Pappi's claim “equalizers give more pleasant result than limiters”, when comparing the results between the limiters vs equalizers, it was pretty clear that equalizers caused a linear result in terms of both audibility and preference, when limiters had a clear critical point threshold, after which the preference was clearly shifted towards the unprocessed sample. The limiter release time also had an effect on both preference and audible difference. So in that term, when the results are noticeable almost immediately with equalizers, using a limiter does give the mastering process and in that sense, the loudness war, a significant “free pass” when trying to achieve loudness until you reach the audible point of distortion.

With this stimuli, even when samples were equal in volume, the majority of the participant answers reflect that the limited samples were not audibly preferred until -12dB of gain reduction, when with any of the bass, middle or treble equalizer frequencies had a clear preference and audibility shift after just mere 4 dB of change.

There was several limitations and shortcomings with this test, that could be improved in future studies. This study was originally planned to include an online survey, which would have resulted in larger quantity of data. However due to technical difficulties and the project timeline, a survey was not possible. Since the technological fields are advancing at very rapid pace, I sincerely hope that similar studies are conducted using a bigger cohort of participants, as it might either fortify or give a totally different result than what this study yielded.

Although a group study might not have been the best format for the study, but it was almost the only good way to schedule it, as there was a few cancellations, schedule changes, etc. If this had been a study where participants were the sole participant in the room, it might have resulted in slightly different results, as participants were now asked to be silent in a small space for 45 minutes with a group of strangers, potentially skewing their answers due to stress or discomfort. Although stereo manipulation was not done to the clips, due to the group research settings, all of the listeners might not have had an optimal listening position.

One major disappointment of the research was that the serial compression data was surprisingly highly preferred by the listeners, but since it had such unreliable and small amount of test data, it was unable to provide definite results. Also an oversight was to use -10 dB gain reduction data for the order comparison test, since most listeners started listing a clear preference with -12 dB samples.

One limitation of the research was that the data was not purely randomized, but the data was always showed in the same order to the participants. Truly random data might have resulted in different results. A much better way would have been to make a ”game” or other interactive computer experience, where the listener would have gotten truly random data, that might have resulted in different results, plus the listener could have listened to different amounts of sample clips depending on how much time they wanted to use. Using a game-style interface would have made the test more social media friendly experiment by making it “shareable”, which could have potentially resulted in much bigger data pool.

6 CONCLUSION

Overall the results reflected that when the listening volume was set to a comfortable level and the loudness was normalized, listeners almost always preferred the unprocessed sample after they started hearing audible distortion or extreme tonal shaping. Since this research showed that loudness normalization is very effective in removing the loudness war effect, the most efficient way to get rid of loudness war would be to make loudness normalization mandatory, without an ability to turn it off. This can give indications for mastering engineers, record labels and software developers on how to approach levels of future recordings.

References

Apple Computer, Inc (2012) *Mastered for iTunes: Music as the Artist and Sound Engineer Intended*. Retrieved from http://www.apple.com/itunes/mastered-for-itunes/

Advanced Television Systems Inc (2011) *A/85: ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television.* Retrieved from http://www.atsc.org/cms/standards/a_85-2011a.pdf

Dye, C., Ralston, J., Wagner, A. (2007-2013) *Turn Me Up!* Retrieved from: http://ww.turnmeup.org

European Broadcast Union (2011) *R128: Loudness Normalisation and Permitted Maximum Level of Audio Signals*. EBU. Geneva, Switzerland.

European Broadcast Union (2011) *EBU Tech 3341: Loudness Metering: 'EBU Mode' metering to supplement loudness normalisation in accordance with EBU R128. EBU. Geneva,* Switzerland. Retrieved from: https://tech.ebu.ch/docs/tech/tech3341.pdf

European Broadcast Union (2011) *EBU Tech 3342: Loudness Range: A measure to supplement loudness normalisation in accordance with EBU R128. EBU. Geneva,* Switzerland. Retrieved from: https://tech.ebu.ch/docs/tech/tech3342.pdf

European Broadcast Union (2011) EBU Tech 3343: Practical guidelines for Production and Implementation in accordance with EBU R128. EBU. Geneva, Switzerland. Retrieved from: https://tech.ebu.ch/docs/tech/tech3343.pdf

European Broadcast Union (2011) EBU Tech 3344: Practical guidelines for distribution systems in accordance with EBU R128. EBU. Geneva, Switzerland. Retrieved from: https://tech.ebu.ch/docs/tech/tech3344.pdf

Fenton, S., Wakefield, J. (2012). Objective Profiling of Perceived Punch And Clarity In Produced Music. *Audio Engineering Society, 132nd Convention, 8626.* Budapest, Hungary.

Fletcher, H. and Munson, W.A. (1933) Loudness, its definition, measurement and calculation. *Journal of the Acoustic Society of America, 5*, 82-108.

Frink, M. (2000) Why Louder Sounds Better. Mix Online. Retrieved from http://mixonline.com/mag/audio_why_louder_sounds/

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* Harper, NY, USA.

Hudson, N. J. (2011) Musical beauty and information compression: Complex to the ear but simple to the mind? *BMC Research Notes, 4*, 9.

International Electrotechnical Commission (2002) *Electroacoustics - Sound level meters - Part 1: Specifications.* IEC. Geneva, Switzerland.

International Telecommunication Union (2012) Recommendation ITU-R BS.1770-3: Algorithms to measure audio peak programme loudness and true-peak audio level. ITU. Geneva, Switzerland.

ISO 226 (2003) *International Standards for Equal-Loudness Level Contours – Criteria for Noise Rating and Evaluation of Audio Equipment Brought Closer to Human Auditory Perception.* Retrieved from http://www.aist.go.jp/aist_e/latest_research/2003/20031114/20031114.html

ISO 532 (1975) *Method for calculating loudness level.*

Katz, B. (2002) *Mastering Audio: The Art and the Science*. Focal press Oxford, UK.

KVR Audio (2013, March 8) *Testing the softsynth waters – Interview with Amir Vinci from Waves.* Retrieved from http://www.kvraudio.com/interviews/testing-the-softsynth-waters---interview-with-amir-vinci-from-waves-21606

Mathews, M.V. (1999) *What Is Loudness / Music, Cognition and computerized sound: An introduction to psychoacoustics.* MIT Press Cambridge, MA, USA.

Nave, R. (2000) *Loudness*. Retrieved from http://hyperphysics.phy-astr.gsu.edu/hbase/sound/loud.html

Nyman, T. (2008) *Minerva Pappi masteroi levyjä Finnvoxilla – Viimeinen voitelu ja onnea matkaan.* Visio 3/08. Helsinki, Finland

Norcross, S. G., Lavoie M. C (2012). Loudness Normalization of Wide-Dynamic Range Broadcast Material. *Audio Engineering Society, 132nd Convention,* 8606. Budabest, Hungary.

Parducci, A (1965). Category Judgement: A Range Frequency Model. *Psychological Review*, 72, 6, 407-418.

Parker. S., Moore, J., Bahraini, S., Gunthert, K., Zellner, D. (2012). Effects of expectations on loudness and loudness difference. *Attention, Perception & Psychophysics*, *74*, 6, 1334-1342.

Robinson, D (2001) *ReplayGain 1.0 specification.* Retrieved from http://www.replaygain.org

Rouse, M (2005) *What is root-mean-square (RMS)?* Retrieved from http://whatis.techtarget.com/definition/root-mean-square-RMS

Serrà, J., Corral, Á., Boguñá, M., Haro, M, Arcos, J.L. (2012). Measuring the evolution of contemporary western popular music. *Scientific Reports, 2*, 521.

Shepherd, I. (2014, October 24) *Has the loudness war been won?* Retrieved from http://productionadvice.co.uk/has-the-loudness-war-been-won/

Shepherd, I. (2012, September 21) *New Green Day Masters – A Loudness War Victory* [Video file]. Retrieved from http://www.youtube.com/watch?v=0EIRF7QW0eA

Shepherd, I. (2013) *Dynamic Range Day – Loudness War Research*. Retrieved from http://dynamicrangeday.co.uk/research/

Shepherd, I. (2014, April 28) Think you know what your music sounds like? Think again... [Video file]. Retrieved from https://www.youtube.com/watch?v=mPhOH05EDMg

Shepherd, I. (2015, March 17) *YouTube just put the final nail in the Loudness War's coffin*. Retrieved from http://productionadvice.co.uk/youtube-loudness/

Siegel, E. H., Stefanucci, J. K (2011) A Little Bit Louder Now: Negative Affect Increases Perceived Loudness. *Emotion, 11*, 1006-1011.

Somich, J., Mishkind, B (2009) *Sound Processing: A History of Audio Processing.* The Broadcasters' Desktop Resource. http://www.thebdr.net/articles/audio/proc/

TC Electronic (2009-2013) *Loudness – Literature & Glossary.* Retrieved from http://www.tcelectronic.com/loudness/literature-glossary/

TC Electronic (2015) *Broadcast Standards – Loudness*. Retrieved from http://www.tcelectronic.com/loudness/broadcast-standards/

TechCrunch (2015, June 8) *Beats Music tells users to switch to Apple Music*. Retrieved from http://techcrunch.com/2015/06/08/beat-it/

Travaglini, A., Alemanno, A., Lantini, F. (2012). Defining the Listening Comfort Zone in Broadcasting Through The Analysis Of The Maximum Loudness Levels. *Audio Engineering Society, 132nd Convention,* 8613. Budabest, Hungary.

Vickers, E. (2010). The Loudness War: Background, Speculation and Recommendations. *Audio Engineering Society, 129th Convention*, 8175. San Francisco, CA, USA.

Wolters, M., Mundt, H., Riedmiller, J. (2010). Loudness Normalization in the Age of Portable Media Players. *Audio Engineering Society, 128th Convention*, 8044. London, UK.

Wolfe, J. (2010) *dB: What is a decibel?* Retrieved from http://www.animations.physics.unsw.edu.au/jw/dB.htm

7 APPENDIX A: SURVEY ANSWERS AND ANSWER SHEET

TABLE 1: List of survey answers

	Type	Processed		Type	Processed		Type	Processed
1	EQ 100 +10	B	36	L2 18	B	71	EQ 1000 +2	B
2	EQ 100 -10	A	37	L2 4	A	72	EQ 1000 -2	A
3	EQ 100 +8	B	38	L2 16	B	73	EQ 1000 +1	B
4	EQ 100 -8	A	39	L2 6	A	74	EQ 1000 -1	A
5	EQ 100 +6	B	40	L2 14	B	75	EQ 1000 0	
6	EQ 100 -6	A	41	L2 8	A	76	L1 100ms 20	B
7	EQ 100 +4	B	42	L2 12	B	77	L1 100ms 2	A
8	EQ 100 -4	A	43	L2 10	A	78	L1 100ms 18	B
9	EQ 100 +2	B	44	L2 0		79	L1 100ms 4	A
10	EQ 100 -2	A	45	L1 1ms 10	B	80	L1 100ms 16	B
11	EQ 100 +1	B	46	L1 5ms 10	B	81	L1 100ms 6	A
12	EQ 100 -1	A	47	L1 10ms 10	B	82	L1 100ms 14	B
13	EQ 100 0		48	L1 20ms 10	B	83	L1 100ms 8	A
14	L1 1ms 20	B	49	L1 50ms 10	B	84	L1 100ms 12	B
15	L1 1ms 2	A	50	L1 100ms 10	B	85	L1 100ms 10	A
16	L1 1ms 18	B	51	L1 250 10	B	86	L1 100ms 0	
17	L1 1ms 4	A	52	L1 500 10	B	87	L1 1000ms 20	A
18	L1 1ms 16	B	53	L1 1000 10	B	88	EQ 10000 +10	B
19	L1 1ms 6	A	54	L1 1ms 10	A	89	EQ 10000 -10	A
20	L1 1ms 14	B	55	L1 5ms 10	A	90	EQ 10000 +8	B
21	L1 1ms 8	A	56	L1 10ms 10	A	91	EQ 10000 -8	A
22	L1 1ms 12	B	57	L1 20ms 10	A	92	EQ 10000 +6	B
23	L1 1ms 10	A	58	L1 50ms 10	A	93	EQ 10000 -6	A
24	L1 1ms 0		59	L1 100ms 10	A	94	EQ 10000 +4	B
25	Staged 7	B	60	L1 250 10	A	95	EQ 10000 -4	A
26	Staged 0	A	61	L1 500 10	A	96	EQ 10000 +2	B
27	Staged 6	B	62	L1 1000 10	A	97	EQ 10000 -2	A
28	Staged 1	A	63	EQ 1000 +10	B	98	EQ 10000 +1	B
29	Staged 5	B	64	EQ 1000 -10	A	99	EQ 10000 -1	A
30	Staged 2	A	65	EQ 1000 +8	B	100	EQ 10000 0	B
31	Staged 4	B	66	EQ 1000 -8	A			
32	Staged 3	A	67	EQ 1000 +6	B			
33	Staged 0		68	EQ 1000 -6	A			
34	L2 20	B	69	EQ 1000 +4	B			
35	L2 2	A	70	EQ 1000 -4	A			

EQ: Waves Q1 using Q value of 2.0 at frequency (100, 1000 or 10000 Hz) and amount of boost (dB)

L1: Waves L1 Ultramaximizer with release time (ms) and amount of gain reduction (dB)

L2: Waves L2 Ultramaximizer with automatic release time (ARC)

Staged: 5 Waves Renessaince Compressors in series with amount of gain reduction

Processed column shows if the processed clip was before (A) or after (B) the beep

In clips 13 24, 33, 44, 75, 86 and 100 audio was identical in both A and B

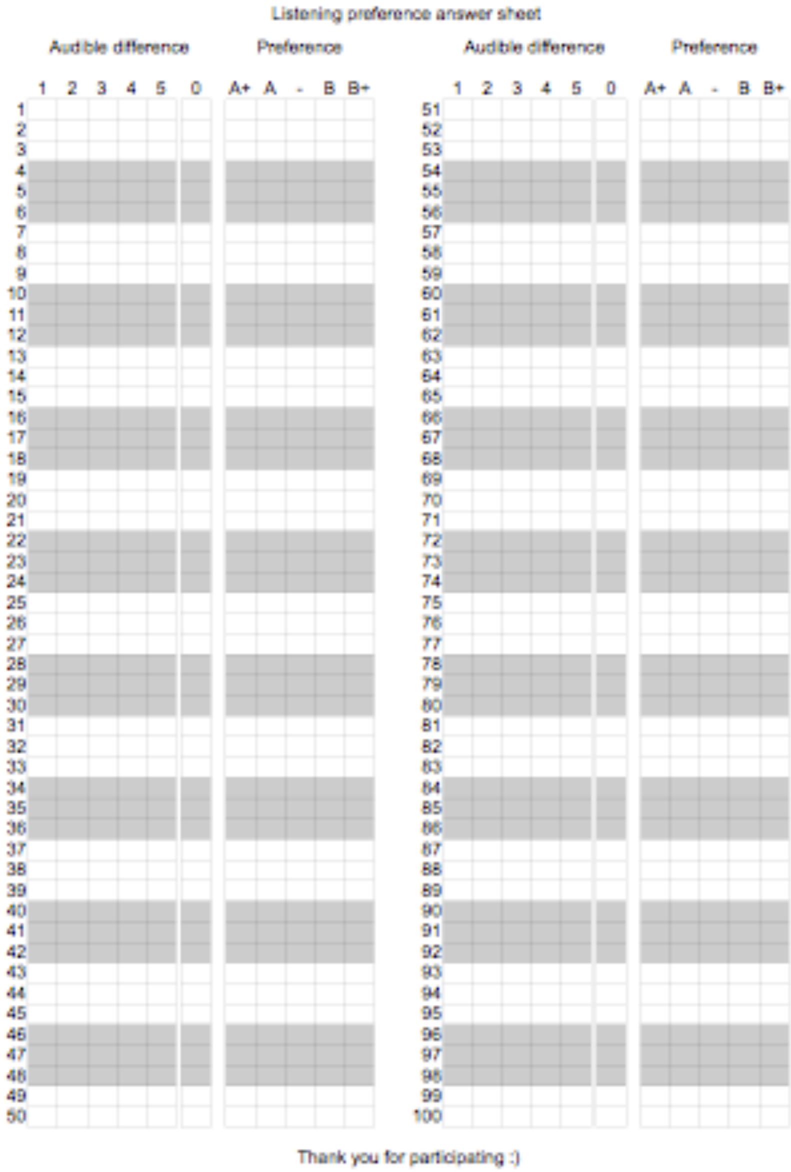

FIGURE 1: Blank answer sheet

8 APPENDIX B: EXAMPLE SCREENSHOTS OF PROCESSING

Example screenshots taken in Steinberg Cubase 6.5 and Avid Pro Tools 10.

FIGURE 1: Staged compression with give Waves Renessaince Compressors with gain reduction of -3 dB each.

Settings used were ARC, Opto, Warm, 1 ms attack, 100 ms release. Ratio (primary) and threshold (secondary) changed to get desired amount of gain reduction (GR). List of all the variants used:

-1dB	GR	each:	1.05:1	ratio.	Thresholds:	-21.5,	-22.5,	-23.5,	-25.0,	-26.0
-2dB	GR	each:	1.10:1	ratio.	Thresholds:	-23.0,	-25.0,	-26.5,	-28.5,	-31.0
-3dB	GR	each:	1.15:1	ratio.	Thresholds:	-23.5,	-26.5,	-29.5,	-32.0,	-35.0
-4dB	GR	each:	1.20:1	ratio.	Thresholds:	-24.5,	-28.0,	-32.0,	-36.0,	-40.0
-5dB	GR	each:	1.25:1	ratio.	Thresholds:	-25.5,	-30.5,	-35.5,	-40.0,	-45.0
-6dB	GR	each:	1.30:1	ratio.	Thresholds:	-28.0,	-34.0,	-40.0,	-46.0,	-52.0
-7dB	GR	each:	1.35:1	ratio.	Thresholds:	-31.0,	-38.0,	-44.5,	-51.5,	-59.5

FIGURE 2: Waves Q10 parametric equalizer with +8 dBb boost at 100 Hz and Q value of 2.0

FIGURE 3: Waves L1 Limiter with 1 ms release time and -11 dB of gain reduction

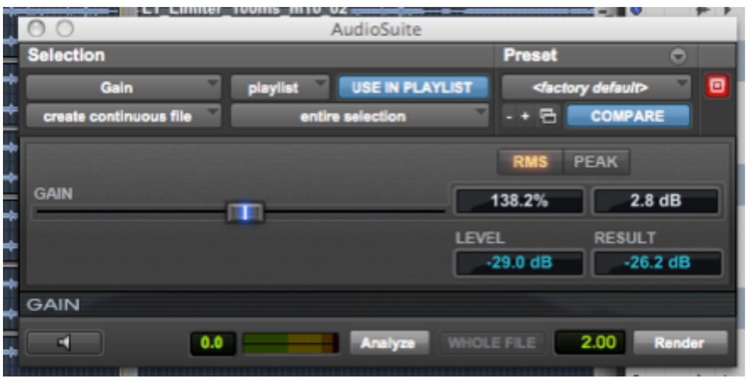

FIGURE 4: Example screenshot of level matching in Pro Tools using the Trim plugin