Loudness Normalisation: Paradigm Shift or Placebo for the Use of Hyper-Compression in Pop Music?

Malachy Ronan
University of Limerick
Castletroy Limerick
Ireland
malachy.ronan@ul.ie

Robert Sazdov
University of Limerick
Castletroy Limerick
Ireland
robert.sazdov@ul.ie

Nicholas Ward
University of Limerick
Castletroy Limerick
Ireland
Nicholas.Ward@ul.ie

ABSTRACT

Loudness normalisation has been heralded as a tonic for the loudness wars. In this paper we propose that a side effect of its implementation may be a greater awareness of sound quality. This side effect is explored through an analysis of the manner in which music is listened to under this new paradigm. It is concluded that the conditions necessary for sound quality judgments have been provided but that the existing preference for hyper-compression may affect the de-escalation of its use in the pop music industry. The aesthetic concerns of hyper-compression are examined in order to determine the sonic trade-offs or perceived benefits inherent in the application of hyper-compression. Factors considered include: (i) loss of excitement or emotion, (ii) audition bias in listening environments, (iii) hyper-compression as an aesthetic preference, (iv) the increased cognitive load of hyper-compression, and (v) the ability of loudness variation to define musical structures. The findings suggest that while loudness normalisation may help de-escalate the loudness wars, listener preference for hyper-compressed music may be more complex than simply a competitive advantage relating to loudness bias.

1. INTRODUCTION

Loudness wars have been pursued to varying degrees in many reproduction formats including vinyl, radio and digital storage mediums [1], [2] resulting in a negative impact on the perceived sound quality of these formats by industry experts [1], [3]. The paradigm of loudness maximization is primarily attributed to the psychoacoustic phenomenon where two identical tracks played back at differing amplitudes will result in the louder signal being perceived as better [4]. This quality judgment is primarily due to the non-linear response of the ear [5] in which louder reproductions result in more frequencies being audible at low and high frequencies [1]. This loudness bias has led to reports of enhanced perception of spaciousness and depth in signals differing by as little as 0.2dB by expert listeners (Katz in [2]). It is this association with sound quality that has led to the widely held belief that music with a reduced dynamic range, resulting in a louder average signal when presented using peak normalization, will sell better [4]. However, this belief was questioned in [4] where it was concluded that listeners are more sensitive to changes in the musical content focusing on melody, harmony, instrumentation, lyrical content, texture and emotion rather than loudness changes when choosing what music to buy. Though this is contrary to popular belief among record company executives in the ‘pop music industry’, the difficulty and expense associated with carrying out a full scale quantitative experiment ensures that this belief may be difficult to overcome [4]. However, while listener preference is based on the musical content, sound quality judgments are biased by loudness differences. For this reason loudness variance is strictly controlled in perceptual listening experiments [6], [7].

Given the sound quality bias resulting from loudness differences, some mastering engineers, whose primary function is to verify or enhance the sound quality of music prior to release [3], have been advocating a fixed monitoring level for a number of years [3], [8]. This fixed monitoring level led Katz to propose a new metering system based on a reference level of 83dB SPL. The argument for a fixed monitoring level suggests that auditioning music at a fixed level facilitates an awareness of sound quality issues between songs and allows the mastering engineer to determine the amount of dynamic range compression (DRC) to be applied prior to auditioning [3]. A fixed monitoring level solution has been made available to the general public in the form of loudness normalization software [9], [10]

The introduction of loudness normalization across the broadcasting industry in Europe [11] and the television broadcast industry in America [12] coupled with the inclusion of the ‘Sound Check’ loudness normalization software on iTunes [10] and open source solutions such as ‘ReplayGain’ [9] implemented on streaming services like Spotify [13], removes the competitive advantage previously afforded to aggressively compressed masters. Loudness normalisation in software audio players operates by accessing the metadata of digital storage files to extract the average level of a track across its entire length using loudness algorithms such as the ITU-R BS.1770 loudness algorithm [11]. This number is then compared against the reference target level and the file playback level is scaled up or down by the audio player in order to match this target output level. This results in all files being perceived at an equally loud level by the listener thereby removing the competitive advantage created by extreme dynamic range compression.

The widespread implementation of this solution has led some commentators to declare that the loudness wars are over [14], [15]. However, this may be an over simplification of the issue in which additional aesthetic factors which contributed to the loudness wars are ignored. Indeed, some engineers have provided anecdotal evidence that the introduction of loudness normalization in Europe has begun to change the way music is being produced and engineered [16]. This would imply that the introduction of loudness normalization has affected the ability of producers and engineers to judge sound quality. However, during the loudness wars, there were frequent calls for an end to the war by numerous authors and music industry professionals, on the grounds that they could perceive a detrimental effect on the quality of music which they attributed to the over compression of dynamic range [3]. This sentiment was not echoed by all involved in music production and music deemed to be of compromised dynamic range is still being produced today [17]

The temptation to over compress the dynamic range led some engineers to compare the use of compression to drug addiction [18]. This addictive nature of compression and its subsequent abuse, has led to the creation of the term 'hyper-compression' [2] to describe the over-use of compression in order to achieve a perceptually louder sound recording [1], [2]. Whilst defining compression as an addiction may be inflammatory, it implies that continued exposure to hyper-compressed music may alter the internal quality meter of the music listener, biasing them towards hyper-compression. In this respect, the definition of a 'normal' amount of DRC is a perpetually moving target dictated by the production processes of the day. However, as loudness normalisation may allow the listener to isolate the subjective quality aspects of the music, in a similar manner to a mastering engineer, the listener may become more aware of sound quality.

This paper examines how loudness normalisation may change the way we listen to music by exploring the issues affecting sound quality judgments in domestic listening environments. An investigation of the aesthetic concerns of hyper-compression is conducted to understand the sensory and cognitive issues underlining the proliferation of hyper-compression during the loudness wars. By understanding the aesthetic motivations for the recent loudness war, we may gain an understanding of the sonic fingerprint of the internal reference of popular music listeners for sound quality in the 21st century.

2. MUSIC LISTENING IN THE AGE OF LOUDNESS NORMALISATION

One of the factors credited with facilitating the loudness wars was the perceived unwillingness of the listener to adjust the level dial when one track was marginally louder than another thereby conceding better sound quality to the louder track [1]. In order to determine the intensity change threshold at which listeners will adjust level, Riedmiller et al. conducted an experiment to examine the 'comfort zone' of listeners which was described as the range within which loudness changes from preferred listening level are considered acceptable [19]. Listeners reported that an increase of 2 or 3dB brought them out of their 'comfort zone' resulting in an increased likelihood of the audio being reduced in level [19]. However, a decrease in level of 6dB was required to prompt listener action. Another experiment conducted by Benjamin in a domestic listening environment sought to gauge the range of levels around the preferred listening level that are accepted as matching the preferred listening level [20]. These domestic listening environments had little background noise and Benjamin found that a +2.91/-3.78dB level change was enough to prompt listeners to describe the level as noticeably louder/quieter while +6.22/-9.22dB results in the level being perceived as too loud or too quiet [20]. However, Norcross et al. found that listeners were much more sensitive to level changes with subjects on average detecting JNDs of 1.24dB between different programs and JNDs of 0.5dB in the same program [21]. Given that loudness normalisation in broadcast is allowed to deviate +/- 2 LU in America [12] and +/- 1 LU in Europe [22], it can be surmised that under a loudness normalisation paradigm, listeners will be using their level control primarily to set their preferred listening level.

While it may be speculative to suggest that listeners utilising loudness normalisation software in domestic listening environments will be capable of judging sound quality to a comparable degree as that of a mastering engineer with specialized equipment, sound quality judgments are relative measures. Therefore, the playback system that the listener uses, regardless of its perceived quality or cost, becomes the reference by which sound quality is judged. Berger conducted an experiment to determine the preference for the different reproduction formats available [23]. The findings concluded an increasing preference with each additional year for the sound of MP3s over vinyl and other digital formats of higher quality [24]. Berger concluded that students that regularly download music from the internet and have grown up with the iPod as their main music player, prefer the sound of MP3s [23]. These results suggest that over an extended period of time the reproduction format will be absorbed into long-term memory and act as a sound quality reference.

In contrast to recording studio environments, the reference signal in domestic listening environments must compete with other signals for the listener's attention. Blauert defines noise as sounds which give rise to unwanted auditory events [25] which, in this context, may constitute any signal that interferes with the listeners perception of music from their playback system. The findings of Pearsons et al. indicate that preferred listening levels increase as background noise increases [26]. Furthermore, Bradley found that for every 1dB of background noise, listeners raised the level by 0.7dB to remain at the preferred listening level [27]. These preferred listening levels, were found to be an average of 61dB Leq [27]. Coincidentally, Ventry et al. in [20] found that in a hearing test booth designed to have low background noise, the most comfortable loudness level for speech was found to be 49.3dB(A) but the task may also involve a different level of attention to a domestic listening environment. This illustrates that the preference for listening level is highly dependent on background noise and the listener's task.

It is proposed that the ability to listen to music at consistent preferred loudness levels in a domestic environment may act as a method of auditory training facilitating a greater awareness of sound quality between different artists auditioned contiguously. While sound quality judgments have traditionally been conducted under experimental conditions or in recording studios, it is proposed that this new music listening paradigm may alter listeners' perception of 'good' sound quality. It is further proposed that this may lead to a more active interest in sound quality in much the same manner as hi-fi enthusiasts in the 1950s pursued the notion of 'presence' [1].

Given that loudness normalisation has created the conditions necessary to facilitate a greater awareness of sound quality, it is important to understand how music listeners will be assessing the sound quality of their music collection. Figure 1 illustrates the process involved in product sound quality judgments where subsequent to the perception of the event, the process is affected by non-auditory modalities, cognition, action and emotional factors until we arrive at an auditory event sound [28]. This mental representation of the acoustic event is referred to as a stream and this stream must compete with other streams for the attention of the listener [29]. These additional streams are outside of the control of the music producer and provide the reason for the low correlation between what we actually hear and the acoustic event [25]. However, while the listener is absorbing input from external factors in a domestic listening environment, the resulting judgment is compared against a reference held in the long-term memory (Figure 1). In this manner, if presented with two versions of the same track, the listener's internal reference may bias the resulting sound quality assessment in favour of the sonic characteristics of their music collection [22]. If their music collection contains primarily hyper-compressed material then it has been suggested that the listener may be biased towards the sonic traits associated with the resulting sound quality [30]. The factors influencing listener preference for dynamic range compression have been proposed to include: prolonged exposure to hyper-compressed music, musical genre preference, perceptually salient sound quality attributes resulting from hyper-compression and the education and training of the listener [30].

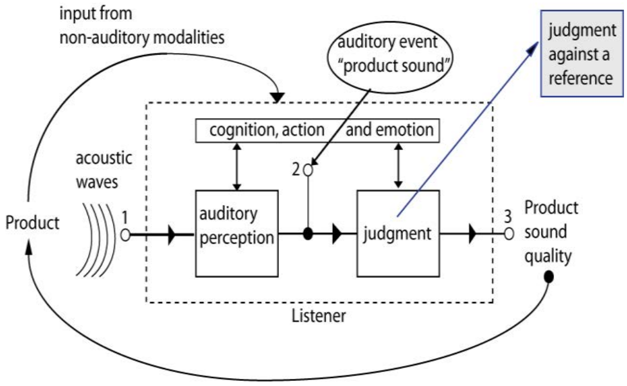

Figure 1 product sound quality judgments adapted from [28]

3. INVESTIGATING THE AESTHETIC CONCERNS OF HYPER-COMPRESSION

Taking into account the improved listening conditions of domestic listeners as a result of loudness normalisation, what follows is a discussion of the aesthetic concerns of hyper-compression highlighting the possible sonic trade-offs or perceived benefits inherent in the application of hyper-compression.

3.1. Loss of excitement and emotion

One of the concerns about hyper-compressed music is that it reduces the emotional and dramatic impact of music, resulting in less active listening which may prevent listeners from bonding with music [2]. This raises the issue of whether variation in loudness is a primary mechanism used to perceive emotion in music. There are numerous theories outlining how emotion relates to cognitive processes and a good summary is available in [31].

One theory proposes that emotions and cognitive processes are integrated through the use of somatic markers where visual images are marked with an emotional association, which helps to increase the speed with which decisions can be made [32]. Some of the mechanisms connecting music and emotion may be automatic and some may require cognitive processing. The varying loudness levels lead to an automatic emotional response and unexpected events provide a heightened sense of arousal followed by a cognitive appraisal [32]. This heightened sense of arousal has been attributed to a physiological response correlated with the approach of a sound source towards the listener [33]. At high intensities, loudness sensitivity increases providing more acute arousal cues for sources closer to the listener [33]. It is these defensive systems that Bradley and Lang believe determine our expression of emotion through sound [34]. Schubert found that 60% of variation in arousal response can be attributed to the musical features of loudness and tempo [35] both of which relate to the awareness of the speed of approaching objects. Juslin and Västfjäll also attribute psychophysical cues such as loud and sudden events to increased arousal which aids the perception of emotion [36]. Hyper-compression will reduce 'loud sudden' events through the restriction of dynamic range implying that listeners will not receive the priming cues required to perceive emotion in music.

Dean et al. sought to determine if changes in listeners perception of arousal correlated with changes in intensity/loudness [34]. The intensity profiles of a number of recordings of varying genre were reversed and both versions presented to the listeners. The results showed that intensity had a direct impact on perceived arousal. This suggests that intensity profiles are a major source of continued perceived arousal in music and while arousal is but one component of emotional response to music, it is considered an important one. Furthermore, the reduction of dynamic range was shown to have an adverse effect on the emotional impact of music [37]. In this manner, we can deduce that varying intensity does provide emotional cues to the listener and restricting the dynamic range may affect the emotional cues available to the listener. It seems plausible then to suggest that when we restrict dynamic range through hyper-compression, we are removing one of the composer's tools for conveying emotion.

Further evidence of the connection between emotion and dynamic range is presented in [38] in which it is proposed that melody and rhythm contain few emotional cues for the listener. Instead, it is proposed that music reflects the dynamic patterns of emotion including tension and release, motion and rest, preparation and fulfillment, and sudden change among others. All of these parameters are lost in hyper-compressed music, which includes little variation of loud and quiet with the presentation remaining at a similar level for the duration. A further theory states that emotion is created through the subtle deviation from our expectations [39]. Mandler concurs, attributing emotion to the creation of expectations in the listener and then the violation of these expectations in subtle ways [40]. Cabrera states that building tension and expectation involves increasing the overall 'size' of the sound by increasing the intensity of the sound and Huron states that an augmentation of sound sources further helps to achieve this [41]. In this manner, the ability to create tension and expectation without audible artifacts in hyper-compressed music is diminished due to the proximity of the RMS level to the 0dBFs peak level, leaving little headroom for intensity changes with the introduction of additional instrumentation. However, the findings of Bishop et al. suggest that listeners of classical music have the ability to imagine loudness variation in the absence of corresponding environmental input [42]. This indicates that in the absence of intensity differences through hyper-compression, the listener may imagine the required cues in order to derive the emotional response intended by the composer. However, further research is required to determine the effect of hyper-compression on the perception of emotional cues.

3.2. Audition bias in listening environments

The listening environment used to audition music may have a profound effect on a listener's preference for DRC. In an environment, the signal carries the desired information (the music) and all other signals present are considered undesired information or 'noise' [25]. Environmental noise in locations such as shops, factories, cars and noisy city apartments consists of broadband noise composed of a number of signals from varying sources fused into a single perceptual construct. Broadband environmental noise is a much more efficient method of masking sounds than tonal masking at each of the critical bands [43]. Therefore, in order for music to be accurately perceived, it may need to provide a consistent average level in order to remain above the noise floor at all times. This is due to frequency or simultaneous masking, in which loud signals mask other signals at nearby frequencies creating a masking threshold within which the quieter signals will become inaudible [40]. Given the correlation between preferred listening level and background noise outlined in section 2 and the ability of loudness normalisation to match program loudness, a hyper-compressed master with a narrow dynamic range may be preferred in environments with greater levels of noise.

Wagenaars presented subjects with an uncompressed stimuli followed by the compressed stimuli and asked subjects to rate the sound quality [44]. It was found that the compressed signal was moderately preferred to the uncompressed, only under noisy conditions. Another experiment in [44] examined the effect of compression in a normal living environment where only the degree of compression (ratio) and the sound level were adjusted. This experiment confirmed that only under noisy conditions was compression preferred and there is a perceived decrease in sound quality when compression increases. It is worth noting that this experiment was conducted in 1986 when the loudness wars had not yet begun [45]. A repeat of this experiment with listeners that have lived through the loudness wars may provide an indication of the effect of the loudness wars on the preference for DRC. This preference for DRC in noisy environments may impact the growing number of listeners using portable media players as their primary means of listening to music [46].

Lund found that listeners have a well-defined dynamic range tolerance (DRT) that varies with environmental noise level. DRT is a continuum containing the preferred average window plus some peak headroom [46]. Within this environment specific DRT continuum, the listener can understand speakers, instruments are heard clearly and there are no sudden loud sounds. However, if the level regularly fluctuates outside this range the listener becomes annoyed. Therefore material with a wide dynamic range may prove unsuitable to communicate the desired information to the listener in noisy environments due to the risk of causing damage to the listener's ears. Given the increasingly noisy environments in which we live, and the preference for a 5-10dB louder level when listening on headphones rather than loudspeakers [46], listeners may prefer the consistent level associated with hyper-compression in order to prevent hearing damage from loud sudden events in wide dynamic range material. In this manner, the varying environmental noise levels that accompany portability coupled with in-ear headphones that provide little or no isolation from noise may have facilitated a trend for hyper-compressed music during the loudness wars.

Wolters et al. suggested that consumers want their content on portable media players to be presented in a form that matches their listening environment [47]. In an effort to address this issue, Vickers suggested that manufacturers provide playback compression options particularly for devices that operate in noisy environments such as car radios. A DRC option on listening devices would likely remove any competitive pressure to make 'louder' masters and provide a more user-oriented listening experience.

3.3. Hyper-compression as aesthetic preference

As previously discussed in section 2, hyper-compression may be an aesthetic preference due to extended exposure to the attributes associated with this type of processing. It has also been linked to genres such as grunge, heavy metal, glitch and shred in which hyper-compression, clipping, aliasing and distortion are the tools used to achieve this aesthetic [2]. Mastering engineers in these genres may intentionally clip the A/D converters in order to increase loudness and add distortion to the signal [48]. Furthermore, while the output of analogue distortion is driven by the frequency of the input and therefore harmonically related, digital distortion such as aliasing is related to the sampling frequency resulting in sum and difference signals un-correlated to the original signal being introduced [49]. The introduction of these inharmonic components may lead to the enhanced perception of dissonance, a desirable component of these genres. This genre specific preference for digital distortion contrasts with the industry wide call for a reduction of the peak level read from a peak program meter (PPM) to -1dBFs in order to avoid clipping when creating a lossy codec, performing sample rate conversion and when playing through a DAC [16]. Given the bias for dissonance created by digital distortion, the promotion of loudness normalisation may not remove hyper-compression from the aesthetic of these genres.

The introduction of the Solid State Logic console heralded a revolutionary design approach in which a compressor was available on every channel [1]. This console has been credited with changing the sound of popular music [1] but it was the console's bus compressor that earned the moniker the 'good' button [18] highlighting an aesthetic preference for DRC among musicians and engineers. Mix bus compression was alleged to make the recording sound like a 'record' or akin to the songs heard on the radio and to please the artists, the engineers added the compression at the mixing stage [18]. This positive association with DRC has been suggested to have been learned in the long-term memory [30] and associated with a preference for hyper-compression. Producers such as Rick Rubin, who produced Metallica's hyper-compressed album 'Death Magnetic', are associated with a sound quality that is lively, loud, exciting and perceived as 'jumping out of the speakers' [50]. However, Rick Rubin is not the first producer to use compression in order to obtain a sound that jumps out of the speakers. Joe Meek famously applied what was considered outrageous amounts of compression (for the 1960s) in order to create this effect [1].

Given that Electronic Dance Music (EDM) has no direct natural reference in the real world, the reference for judging the amount of compression to use is provided by other professionals working in the genre. As EDM is a relatively new genre associated with digital audio, which is linked to hyper-compression [1], practitioners most likely learned their craft and how to use compression during the loudness wars. This may have affected their preference for DRC and as there is no real world equivalent to many EDM instruments, the music released becomes the reference. If this music is hyper-compressed, then it is likely that other artists emulating this sound will employ similar levels of DRC.

The effect of recording on the globalisation of performance in classical music has been well documented [51] and given the ubiquitous access to the same technology, a globalisation of sound among genres is likely. Furthermore, the trends in classical music performance over the twentieth century are clearly preserved on recordings [51] which provokes the question of how the aesthetic preference for hyper-compression in this era will be perceived by future generations?

3.4. Increased cognitive load of hyper-compression

A side effect of using DRC to balance relative levels is the creation of what has been termed a more coherent result due to the ability of compression to fuse individual sound sources together [52]. Moore reported that a number of studies found that coherent amplitude changes tend to perceptually fuse sounds together whereas un-correlated changes segregate the sounds [53]. When the components are amplitude modulated in the same way, they become fused together into a single percept and they are listened to as a whole [53]. This can also lead to difficulty in perceiving individual sound sources within a complex signal, presented to the listener as a reproduced scene. This relates to the brain's ability to use the common modulation of sources as a means of grouping them together or, conversely, the lack of common modulation as a means of identifying them as individual sources. When this process is interrupted by the compression of a group of sources, a common modulation envelope is applied to all of the instruments in the reproduced scene. The type of common modulation applied to the signal is determined by the time constants selected on the compressor and may be largely responsible for the widely acknowledged 'gluing effect' when applied to a group of sources.

When using analogue compressors or digital processors with non-linear distortion modeling such as dynamic convolution [54], the processor will apply harmonics to the signal determined by its topology or modeling. This application of harmonics of fixed relative power to a group of instruments may provide a coherent coloring to the sound scene. However, there is currently no research to support the notion that harmonic modulation will support perceptual grouping when both sources are modulated coherently [53]. Differences in timbre support stream segregation [55]. When two melodies are played together using different timbres, listeners find it easier to identify changes in the melody than when both timbres are similar. It seems possible then that the application of fixed harmonics to a group of instruments may make it more difficult to perceptually segregate sources. Further research on this point is needed. This proposed trade-off between DRC and stream segregation resembles the trade-off between pitch and timbre when determining the perceptual organisation of a sequence of notes [56]. This trade-off states that when listeners are presented with a sequence played by two instruments, with a large difference in timbre and a small difference in pitch, the percept heard will be that of grouping by timbre. The opposite is also true.

This fusing together of individual sound sources through the use of compression elicits a varying preference response among listeners and engineers [18]. These subjective responses may be related to the manner in which the listener engages with the material. Critical listeners, such as musicians and engineers, often complain about hyper-compression [3], [16] as they may need to be able to hear all of the instruments separately and the heavy application of compression may obscure the individual stream segregation of these sources. On the other hand, listeners relegating music to background sound may prefer the music to fuse together perceptually into a single source as the role of the individual instruments may be of less concern and wide dynamic range may draw too much attention to the music. These application specific listening modes may dictate the degree of compression preferred by different listeners.

3.5. Loudness variation defining musical structure

One of the arguments against aggressive DRC is that some genres, such as the alternative music of the early nineties and late eighties, relied on loudness differences between verse and chorus as a structural element [2]. In acoustic ensembles, when loud notes fall on the down beat, they provide a tonal stability to the music [57] and hyper-compression may remove the strength of the downbeat perceived by the listener thus removing some of the composer's artistic intent. Furthermore, melodies are often accompanied by a level boost and a 20 ms lead time in acoustic ensembles [57] to enable streaming of the lead instrument(s) as the foreground sound. Due to the excessive use of DRC during the loudness wars, these performance attributes are simulated with macro-dynamic volume automation post-compressor [3] and via the nudge function in Pro Tools [58] respectively. While the loudness wars all but ended the ability to use loudness as a defining musical structure, new structures were born out of necessity. Engineers began to use the frequency range as a method of defining musical structure where choruses are signified by an expanding of the frequency bandwidth of the music at the low and high frequencies and a widening of the sound stage in contrast to the verses [59]. In this regard, the legacy of the loudness wars on music production processes has yet to be discussed and future research is needed.

4. CONCLUSION

This paper examined the factors affecting music listening under the loudness normalization paradigm, highlighting the manner in which loudness normalisation may facilitate a renewed interest in sound quality. Given that loudness normalisation has been associated with an end to the loudness wars, the aesthetic concerns of hyper-compression were discussed in order to further understand the reasoning for the use of hyper-compression beyond simply increasing loudness. In this manner, this paper explored whether the application of hyper-compression can truly be eradicated through the proliferation of loudness normalisation. The findings suggest that while loudness normalisation may help in this regard, listener preference for hyper-compressed music may be a more complex issue than simply relating to loudness bias. Further research is needed concerning the aesthetic factors contributing to listener preference for hyper-compression and the music production techniques created as a direct result of the loudness wars.

5. REFERENCES

[1] G. Milner, *Perfecting sound forever: an aural history of recorded music*. New York: Faber and Faber, 2010.

[2] E. Vickers, “The loudness war: Background, speculation, and recommendations,” in *Audio Engineering Society Convention 129*, 2010.

[3] B. Katz, *Mastering Audio: the art and the science*, 2nd ed. London: Focal Press, 2007.

[4] E. Vickers, “The loudness war: Do louder, hypercompressed recordings sell better?,” *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 346–351, 2011.

[5] H. Fletcher and W. A. Munson, “Loudness, Its Definition, Measurement and Calculation,” *J. Acoust. Soc. Am.*, vol. 5, no. 2, pp. 82–108, 1933.

[6] ITU-R BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” International Telecommunications Union, Geneva, 1997.

[7] S. Bech and N. Zacharov, *Perceptual Audio Evaluation-Theory, Method and Application*. Chichester, UK: John Wiley & Sons Ltd., 2006.

[8] B. Katz, “Integrated Approach to Metering, Monitoring, and Leveling Practices, Part 1: Two-Channel Metering,” *J Audio Eng Soc*, vol. 48, no. 9, pp. 800–809, 2000.

[9] Hydrogen audio, "ReplayGain 1.0 specification," 28-Mar-2014. .

[10] Apple, "iTunes: About Sound Check," 23-Sep-2010. .

[11] ITU-R BS.1770-3, "Algorithms to measure audio programme loudness and true-peak audio level," Geneva, 2012.

[12] "ATSC A/85," 2013.

[13] "Spotify," 2014. .

[14] H. Robjohns, "The end of the loudness wars?," *Sound on Sound*, no. February 2014.

[15] G. Reierson, "The loudness war is over," *Mix*, 02-Aug-2011. .

[16] R. Katz, *ITunes music mastering high resolution audio delivery : produce great sounding music with Mastered for iTunes*. Waltham, MA: Focal Press, 2013.

[17] "Dynamic Range database," *Album list*. .

[18] B. Owsinski, *The Mixing Engineer's Handbook*, 2nd ed. Boston, USA: Course Technology, 2006.

[19] C. Robinson, S. Lyman, and J. C. Riedmiller, "Intelligent Program Loudness Measurement and Control: What Satisfies Listeners?," in *Audio Engineering Society Convention 115*, 2003.

[20] E. Benjamin, "Preferred Listening Levels and Acceptance Windows for Dialog Reproduction in the Domestic Environment," in *Audio Engineering Society Convention 117*, 2004.

[21] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie, "The Subjective Loudness of Typical Program Material," in *Audio Engineering Society Convention 115*, 2003.

[22] EBU-R128, "Loudness normalisation and permitted maximum level of audio signals," European Broadcasting Union, Geneva, Aug. 2011.

[23] N. Kelly, "Manic compression: it's killing rock'n'roll -- but the kids like it," *Irish Independent*, 2012.

[24] D. Dougherty, "The Sizzling Sound of Music," 03-Jan-2009. .

[25] J. Blauert, "Cognitive and aesthetic aspects of noise engineering," presented at the Internoise, 1986, vol. 1, pp. 5–14.

[26] K. Pearsons, R. Bennett, and S. Fidell, "Speech Levels in Various Noise Environments," California, EPA-600-1-77-025, May 1977.

[27] J. S. Bradley, "Acoustical measurements in some Canadian homes," *Can. Acoust.*, vol. 14, no. 4, 1986.

[28] J. Blauert and U. Jekosch, "Sound-quality evaluation - A multi-layered problem," *ACUSTICA*, vol. 83, no. 5, pp. 747–753, 1997.

[29] A. S. Bregman, *Auditory scene analysis : the perceptual organization of sound.* Cambridge, Mass.: MIT Press, 2001.

[30] M. Ronan, R. Sazdov, and N. Ward, "Factors influencing listener preference for dynamic range compression," in *AES 137 Convention*, Los Angeles, in press.

[31] W. F. Thompson, *Music, thought, and feeling: understanding the psychology of music*. Oxford: Oxford University Press, 2009.

[32] A. R. Damasio, *Descartes' error: emotion, reason and the human brain*. London: Vintage, 2006.

[33] J. G. Neuhoff, "Perceptual bias for rising tones," *Nature*, vol. 395, no. 6698, pp. 123–124, 1998.

[34] R. T. Dean, F. Bailes, and E. Schubert, "Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation," *PLoS ONE*, vol. 6, no. 4, p. e18591, Apr. 2011.

[35] E. Schubert, "Modeling Perceived Emotion with Continuous Musical Features," *Music Percept.*, vol. 21, no. 4, pp. 561–585, 2004.

[36] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behav. Brain Sci.*, vol. 31, no. 05, pp. 559–575, 2008.

[37] A. Bhatara, A. K. Tirovolas, L. M. Duan, B. Levy, and D. J. Levitin, "Perception of emotional expression in musical performance.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 37, no. 3, pp. 921–934, 2011.

[38] S. K. Langer, *Philosophy in a new key: a study in the symbolism of reason, rite, and art, 3rd Ed.* Cambridge, Mass: Harvard University press, 1957.

[39] L. B. Meyer, *Emotion and meaning in music*. Chicago: University of Chicago Press, 1961.

[40] G. Mandler, *Mind and body : psychology of emotion and stress*. New York: W.W. Norton, 1984.

[41] K. N. Olsen, C. J. Stevens, and J. Tardieu, "Loudness change in response to dynamic acoustic intensity.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 36, no. 6, pp. 1631–1644, 2010.

[42] L. Bishop, F. Bailes, and R. T. Dean, "Musical Expertise and the Ability to Imagine Loudness," *PLoS ONE*, vol. 8, no. 2, p. e56052, Feb. 2013.

[43] E. Zwicker and H. Fastl, *Psychoacoustics: facts, and models*, vol. 22. New York: Springer-Verlag, 1990.

[44] W. M. Wagenaars, A. J. Houtsma, and R. A. van Lieshout, "Subjective evaluation of dynamic compression in music," *J. Audio Eng. Soc.*, vol. 34, no. 1/2, pp. 10–18, 1986.

[45] E. Deruty and D. Tardieu, "About Dynamic Processing in Mainstream Music," *J. Audio Eng. Soc.*, vol. 62, no. 1/2, pp. 42–55, 2014.

[46] T. Lund, "Control of Loudness in Digital TV," in *Proc. of the NAB-2006 Convention*, 2006.

[47] M. Wolters, H. Mundt, and J. Riedmiller, "Loudness Normalization in the Age of Portable Media Players," in *Audio Engineering Society Convention 128*, 2010.

[48] P. White and M. Houghton, "Crafting loud mixes that sound great," *Sound on Sound*, Mar-2012.

[49] F. Rumsey and T. McCormick, *Sound and recording*. Amsterdam: Elsevier/Focal, 2009.

[50] "Lars Ulrich takes on the Death Magnetic complainers." [Online]. Available: http://www.musicradar.com/news/guitars/lars-

ulrich-takes-on-the-death-magnetic-complainers-175581/.

[51] R. Philip ,, *Performing music in the age of recording*. New Haven: Yale University Press, 2004.

[52] M. A. Stone, B. C. Moore, C. Füllgrabe, and A. C. Hinton, "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, pp. 532–546, 2009.

[53] B. C. J. Moore, *An introduction to the psychology of hearing*. Bingley: Emerald Group, 2008.

[54] M. J. Kemp, "Analysis and simulation of non-linear audio processes using finite impulse responses derived at multiple impulse amplitudes," in *Audio Engineering Society Convention 106*, 1999.

[55] R. Cusack and B. Roberts, "Effects of differences in timbre on sequential grouping," *Percept. Psychophys.*, vol. 62, no. 5, pp. 1112–1120, Jul. 2000.

[56] P. G. Singh, "Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?," *J. Acoust. Soc. Am.*, vol. 82, no. 3, pp. 886–899, 1987.

[57] R. A. Rasch, "Synchronization in Performed Ensemble Music," *Acta Acust. United Acust.*, vol. 43, no. 2, pp. 121–131, Sep. 1979.

[58] F. D. Cook, *Pro tools 101: an Introduction to Pro tools 10*. Boston, MA: Course Technology PTR, 2012.

[59] *Pensado's Place #70 - Mixing Engineer and Producer, Michael Brauer*, # 70 vols. 2012.