THE EFFECTS OF DISTORTION ON THE PERCEPTION OF LOUDNESS IN LIVE SOUND

SE Durbridge Department of Engineering, University of Derby, UK
AJ Hill Department of Engineering, University of Derby, UK
J Taylor d&b audiotechnik, Gloucestershire, UK

1 INTRODUCTION

Psychoacoustics in live sound directly influences the perception of both mix engineers and audience members. The perception of loudness may have a direct relationship with how individuals perceive their environment, therefore optimizing loudness in live sound is a potentially useful tool for producing the best audio environment while maintaining safe listening levels. The distortion exhibited by loudspeaker systems operating at high levels may influence loudness perception, and the effect of distortion may be independent of sound levels.

This paper provides a cursory glance into the perception of distortion as it relates to apparent loudness. A brief description of the key physiological and psychoacoustic factors of loudness perception are given in Section 2. Following this, distortion is described (Section 3) followed by a description of methodology and results stemming from a listening test exploring these matters (Section 4).

2 LOUDNESS PERCEPTION

Accounting for the perception of loudness may be important when attempting to cultivate an optimal listening environment at safe listening levels. Absolute limits of noise exposure and safe sound levels at events are commonplace. Optimizing the loudness of mixes may be a next logical step in improving the live experience.

2.1 Equal Loudness Contours

Fletcher and Munson[1] developed equal loudness contours (with refinements provided by Robinson and Dadson[2]) and showed that humans have a frequency dependent sensitivity to the loudness of pure tones. Equal loudness contours relate sound pressure level (SPL) to subjective loudness, expressed in Phons[3].

Critically, the equal loudness contours highlight that human hearing sensitivity reaches its peak in the 3-5 kHz range, while sensitivity falls off at low and high frequencies (Figure 1). That said, perceived loudness can be significantly boosted by a small change in low-frequency content, as the contours in this range are very close together (although to achieve equal loudness to that in the more sensitive regions, much higher SPLs are required).

It is suggested that sound reproduction systems should ideally be designed to accommodate for this relationship, as at lower SPLs it may be necessary to increase the gain of LF and HF content differently, relative to the desired loudness level[4].

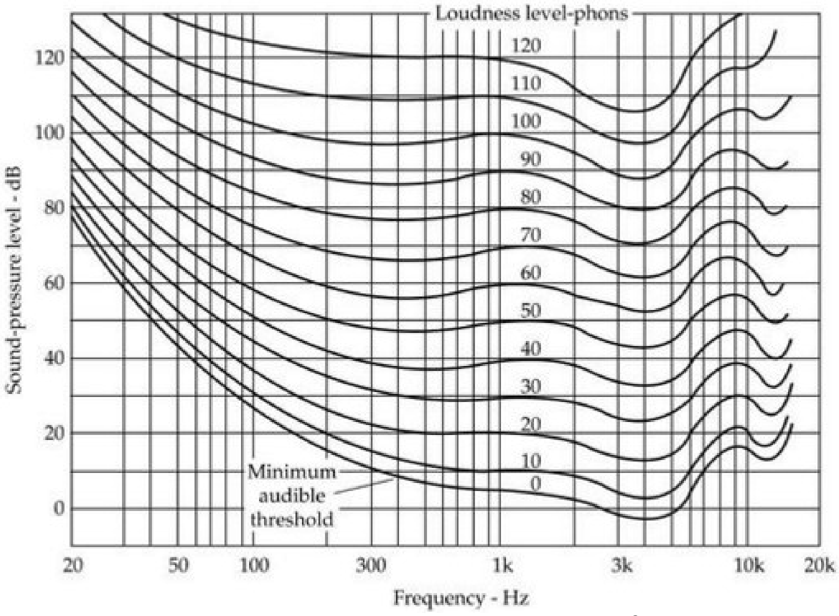


Figure 1 Equal loudness contours[2]

2.2 Auditory Masking

Loudness may be dependent on bandwidth for complex signals such as music. The perception of complex signals is influenced by masking. Auditory masking is the temporal incapacity to clearly perceive certain signal content, due to other signal content affecting the hearing mechanism.

The threshold of a tone being masked by a sinusoid-based stimulus at different frequencies is dependent on the bandwidth of the masking stimulus up to a critical bandwidth (CB)[5]. This led Fletcher and others to develop models of the basilar membrane based on overlapping bandpass filters with contiguous passbands. Figure 2 shows the increase in CB with frequency. Fletcher's experiment made a number of unrealistic assumptions about the use of pure tones and since then, the use of notched-noise methods for filter calculation has been implemented[6]. Figure 3 shows the hearing threshold of a masked signal as a function of the width of the notch in the noise-based masking signal.

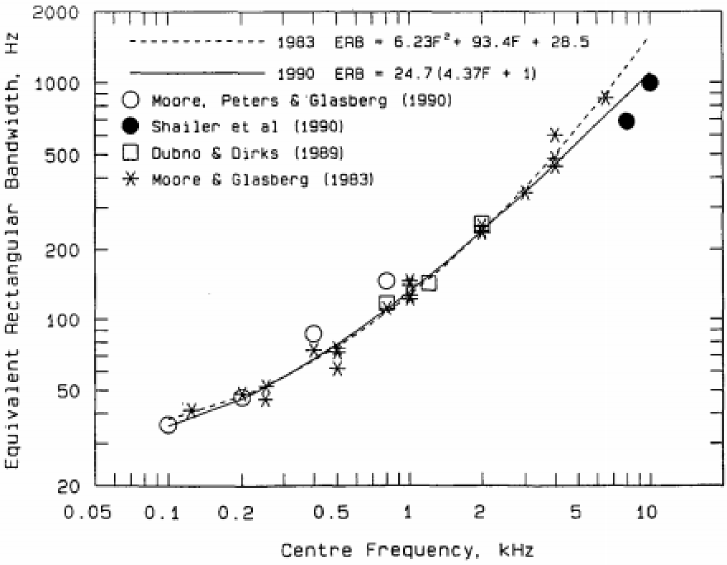


Figure 2 ERB filter bandwidth with respect to centre frequency for various filter models[34]

The notched noise filter calculation assumes that the auditory filter response is symmetrical about the centre frequency and that the increase in bandwidth is linear with the increase in level. Paterson e*t al* and Moore *et al* suggest a method for calculating signal thresholds for the notched-noise method using asymmetrical filter shapes, which is manifested as the Equivalent Rectangular Bandwidth model[6-8]. This model introduces a change in the shape of LF and HF filter overlap with respect to amplitude, increasing the width of a filter's LF overlap with respect to increasing amplitude while sharpening the HF overlap (not a uniform change with frequency).

Masking (due to narrowband noise) at low frequencies maintains a relatively consistent slope with increasing level, where the high frequency slope changes significantly. This suggests that masking at high levels is not only asymmetrical, but greatly effects perception of HF signal content (Figure 3).

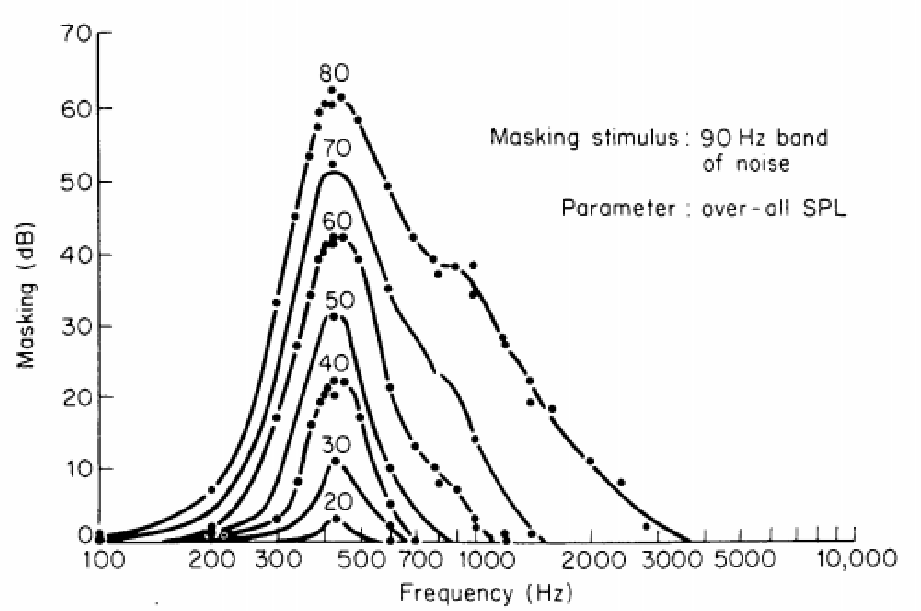


Figure 3 Masking patterns at different levels for a narrow band stimulus at 450Hz[34]

Masking is a temporal effect, and has three time-based descriptors: simultaneous masking, backwards masking, and forward masking. Simultaneous masking refers to masking occurring in time with the reference signal. Backwards masking refers to masking of a sound by a following stimulus that has occurred within milliseconds of the reference. Forward masking refers to the masking of a sound by a stimulus which ended before that sound[3]. Toole suggests that simultaneous masking is crucial in perceiving the original content in a signal that has been distorted[11]. This may be a critical factor in the role of distortion in perceived loudness in sound reinforcement systems, as will be explored in Section 4.

2.3 Loudness Units

Loudness Units (LU) are a perceptually-motivated metric defining the loudness of audio program material. The calculation of LU is an open standard published by the ITU[12]. Loudness Units Full Scale (LUFS) is calculated using the sequence shown in figure 8 and is expressed in decibels. A full description of LUFS calculation procedure can be found in the relevant ITU recommendation[12].

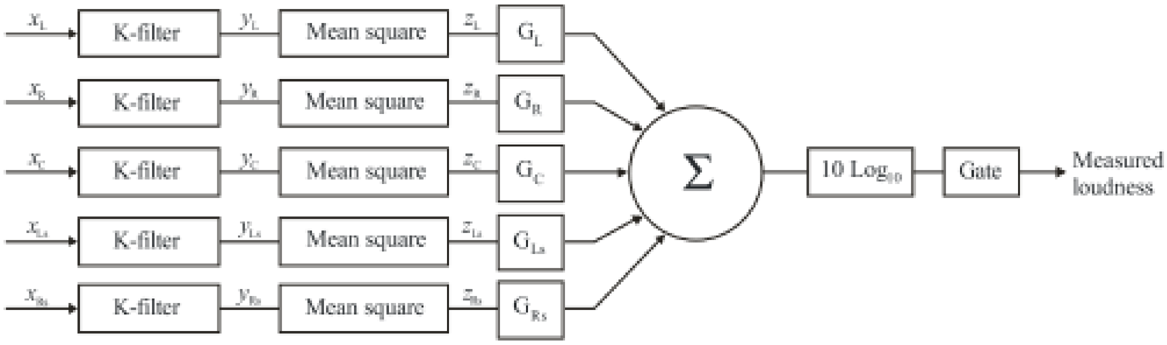


Figure 4 Simplified block diagram of LUFS calculation[12]

A series of filters are used to replicate human perception, and window-based gating and averaging of the signal is used to calculate a signal's loudness with respect to summed channels. Overlapping windows of differing lengths can be implemented to smooth the response of LUFS calculation, and are displayed as short term and long term values[13].

An absolute gating threshold of -70dB is applied to counteract periods of silence or low-level noise, which may reduce the effectiveness of loudness calculation. The method of averaging the loudness of a track allows for the equal comparison of loudness of tracks, independent of the crest factor or dynamic range of a track.

2.4 Loudness & Dynamics

To build a complete picture of loudness, it is important to consider the effect of signal dynamics on loudness perception. The dynamics of a signal can be described by crest factor (CF), dynamic range (DR) and loudness range (LRA). The CF is the relationship between the peak and the average level of the signal[4], of which the average is the RMS level of the signal (Equation 1).

$$CF = x_{\max}(dB) - x_{RMS}(dB) \tag{1}$$

The DR of a signal (or system) is defined as the difference between the peak signal and minimum signal - usually the noise floor (Equation 2).

$$DR = x_{\max}(dB) - x_{\min}(dB) \tag{2}$$

The LRA of a track is calculated by taking the average of a number of LUFS measurements, excluding the lowest 10% and highest 5% of readings[15].

Wendl & Lee found that changes in CF correlated with changes in signal loudness and quality[14]. Similar results have been found in the form of greater perceived loudness with increasing signal compression (i.e. lower CF)[27,33].

Two factors of loudness perception that directly relate to live sound are the stapedius reflex (SR) and the loudness overflow effect (LOE). The SR is the internal limiting function of the human auditory system, which operates as a muscular compression of the stapes in response to high signal transfer from the outer ear[,18]. SR is typically triggered around 70 dB SPL, and results in a reduced amplitude transfer to the inner ear. Some signal compression may also occur in the cochlea, which may also accommodate for a difference in dynamic range between the measureable system and a listener's perception[3]. This inevitable disagreement between measurements and perception is rarely mentioned in the discussion/debate surrounding live sound system loudness.

Additionally, the bandwidth of nonlinear distortion may have a direct impact on the perceived dynamic range of sound at different listening levels, decreasing with narrower bandwidth nonlinear behaviour[18]. This is analysed as the loudness overflow effect (LOE), which suggests that nonlinearity contributes to loudness sum and in turn increases perceived dynamic range[18]. LOE can, however, be active in relation to wideband nonlinearity or inhibited due to narrowband nonlinearity. The number of audio sources can also influence LOE. The nonlinearity of the human auditory system may provide an important comparison to this[23].

3 DISTORTION

Distortion is defined as "…any change in the waveform or harmonic content of an original signal as it passes through a device. The result of nonlinearity within a device."[4] Two of the analytical methods used to quantify the nonlinearity of a system are total harmonic distortion (THD) and intermodulation distortion (IMD).

There are two variations of the basic THD equation. THD_F is described as the comparison between the harmonic content of a waveform compared to the fundamental (Equation 3). THD_R is described as a comparison between the harmonic content of a waveform and its RMS value (Equation 4)[19]. When a system is probed with a sine wave, the output of the system is compared with one of these methods and the relationship is often described as a percentage[4].

$$THD_F = \frac{\sqrt{\sum_{n=2}^{\infty} |H(f_n)|^2}}{|H(f_1)|} \tag{3}$$

$$THD_R = \sqrt{\frac{\sum_{n=2}^{\infty} |H(f_n)|^2}{\sum_{n=1}^{\infty} |H(f_n)|^2}} \tag{4}$$

Where $|H(f_n)|$ is the magnitude response of the system at the n^{th} harmonic of the fundamental frequency.

Physical measurements of THD% are often taken with comparison to the noise floor of the measurement, which is described as THD+N[20].

IMD is described as the distortion that occurs when multiple spectral components interact with a system nonlinearity[3]. This is often measured by introducing two sine waves into a system and comparing the system's output to the input at different frequency intervals[20].

Both methods of measuring distortion display the addition of spectral content into a signal, as a result of the nonlinear behavior that is measured. This extra harmonic content may change the dynamic properties of the signal, and influence the psychoacoustic factors discussed in section 2.

Analytical methods of measuring distortion show how a nonlinear system interacts with pure tones, but cannot be used to evaluate complex signals. The measurement of nonlinear behavior does not directly correlate with the distortion content seen when comparing clean and distorted spectrograms of audio signals (complex signals, as opposed to the pure tones used for THD and IMD measurements). Toole suggests that conventional distortion measurements, such as THD and IMD, do not correlate with the perceptions when listening to music[11]. This is supported by Henin & De Santis[10]. This indicates that analyzing an audio signal for THD and IMD levels isn't acceptable for examining the effect distortion has on perceived loudness of live sound reinforcement systems.

3.1 Rnonlin

Another approach to quantifying signal distortion is to use a perceptual model[8-10]. A comparison of distortion perception models has shown that Rnonlin results correlate well with listener perception[10]. Simply put, Rnonlin quantifies the amount of perceptible distortion in a signal by comparing the signal with and without added distortion[9]. The method for calculating Rnonlin (figure 5) features filtering, including the use of a gammatone array of ERB-n wide filters described in section 2. A maximum cross-correlation of overlapping (three window wide) blocks are weighted and averaged, resulting in a value of normalised similarity (i.e. a value of 1 features no distortion, and 0 no similarity). A full explanation of Rnonlin and its calculation can be found in a series of papers by Moore *et al*[7].

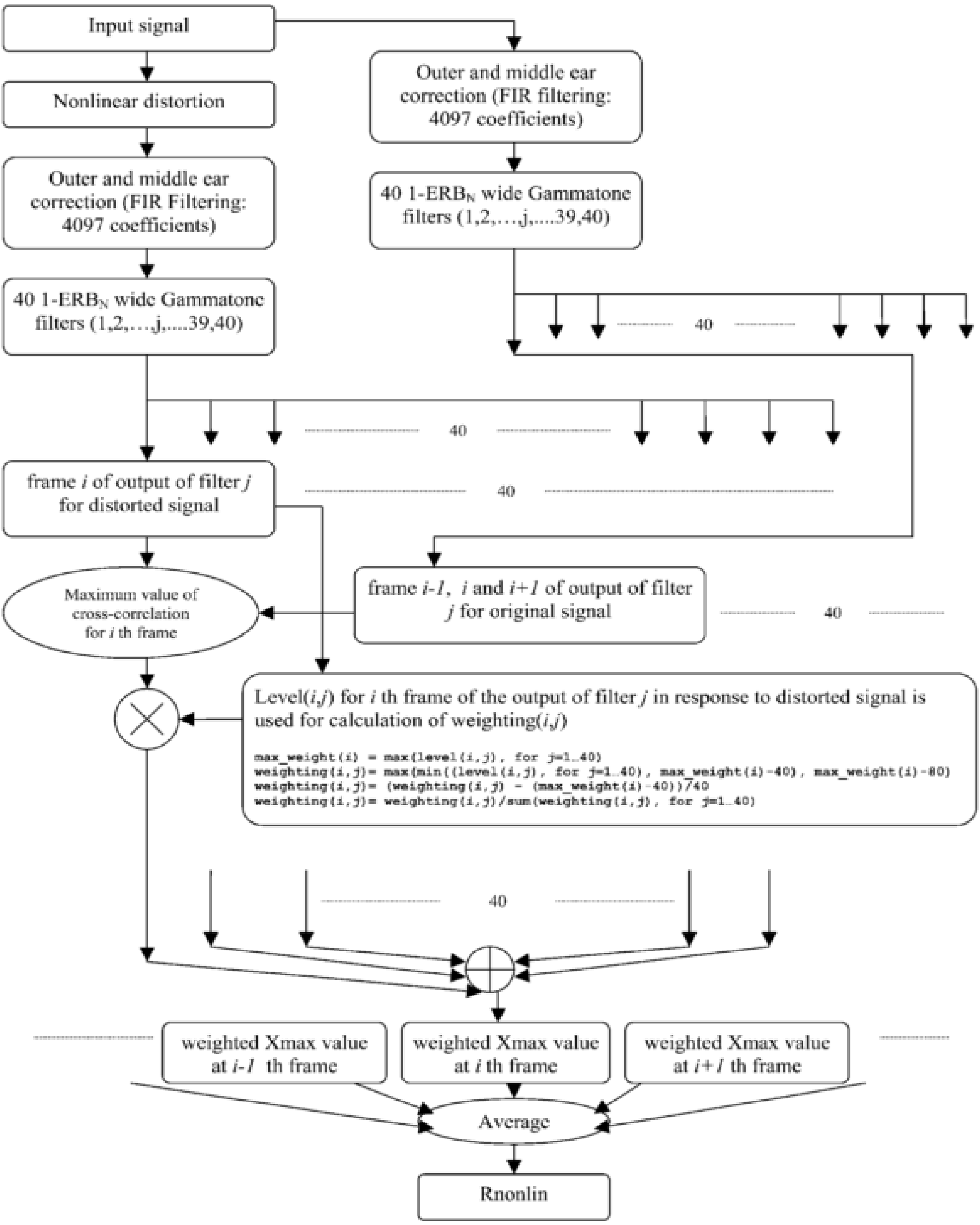


Figure 5 Block diagram of Rnonlin calculation algorithm[9]

4 PRACTICAL EXPERIMENTS

In order to properly examine the effects distortion has on loudness for real live sound reinforcement systems, a series of subjective evaluations was devised, with appropriate signal analysis applied using the methods described in Sections 2 and 3.

4.1 Method

A blind listening test was designed using bespoke software developed in MATLAB which aimed to show if distorting music would have a consistent effect on a track's perceived loudness. Twenty-five participants took part in a headphone-based test. Additionally, ten participants took part in a loudspeaker-based test in an identical format to the headphone-based test. Listeners were asked to compare two versions of 30-second song samples, repeated for ten tracks (four of which were used twice). One version of each song was the reference version (no processing applied) while the other versions were distorted versions of the reference (distorted using a chosen nonlinear function). All tracks were peak normalised to -10 dBFS. The quantifiable attributes of the ten tracks used is shown in Table 1.

Loudness and Distortion Characteristics Change for Listening Test							
Track	LRA	CF	Rnonlin	LUFS	Genre	Device	Threshold
1	-0.0025	0.02	0.995	0.02	Triphop	Half Clip	0.7
2	-0.0311	-2.2	0.967	2.25	Triphop	Fuzz Exp	Sign
3	0.4948	2.7	0.942	-2.61	Electronica	Norm Sig	J = 0.7
4	0.0189	-1.7	0.996	1.77	Acoustic	Soft Clip	0.6 / 1:2
5	0.045	-6	0.958	6.14	Acoustic	Hard Clip	0.45
6	-0.0951	-3.5	0.987	3.49	Hiphop	Soft Clip	0.5 / 1:4
7	-0.432	-1.2	0.981	1.11	Drum & Bass	Soft Clip	0.6 / 1:2
8	-0.4547	-1.1	0.987	0.99	Drum & Bass	Hyp Tang	Tanh
9	-0.1399	-3.9	0.996	4.12	Acapella	Hard Clip	0.6
10	-0.0748	-1.7	0.999	1.8	Acapella	Soft Clip	0.6 / 1:2

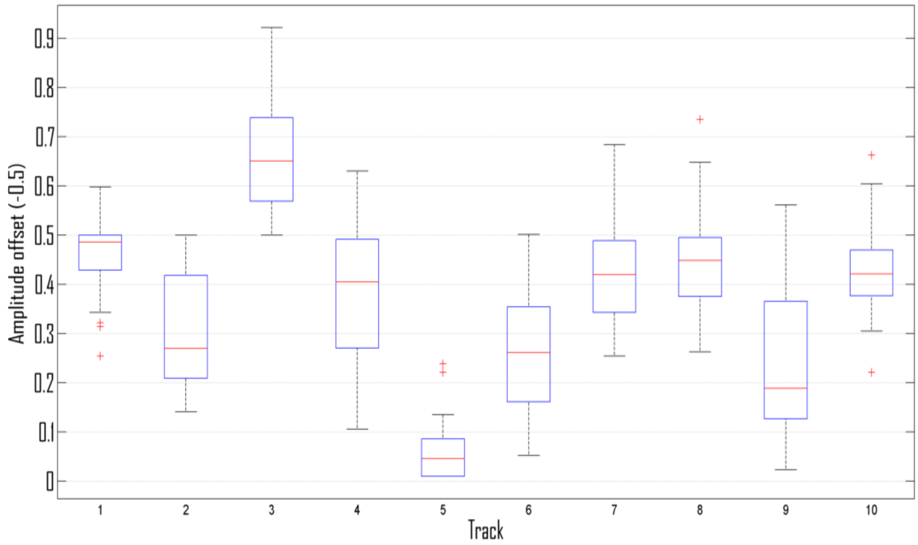
Table 1 Loudness and distortion characteristics for the listening tests

Listeners were instructed to compare the two versions of each track (one reference, one distorted) and to try to balance the level of the distorted track to match the reference track. The testing system was rigorously maintained to be the same for each batch of tests, with the headphones calibrated so that consistent playback level was maintained for all participants and matched to the loudspeaker tests.

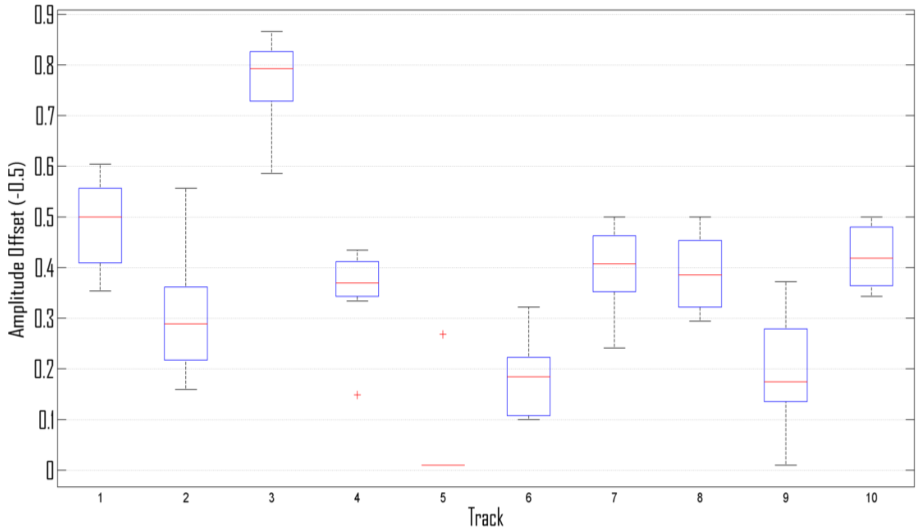
4.2 Results analysis

Figure 6 shows the listening test results for the headphone and loudspeaker-based tests. The data presented indicates the amplitude offset chosen by the listeners for distorted signal so that it matches the reference signal in perceived loudness.

When comparing the same tracks with different distortions in Figure 6(a) (1 & 2, 4 & 5, 7 & 8, 9 & 10), the pairs exhibit different IQRs and mean offsets. The more severe nonlinear devices have larger IQRs than less severe nonlinear devices, even for those with very large Rnonlin values. The difference in IQR is not necessarily genre-dependent for each nonlinearity, and additionally it can be seen by the pattern consistency between the data sets that the effect of distortion is independent of playback medium (although the IQRs of loudspeaker data are interestingly more compact, potentially indicating great performance consistency for real-world sound systems, not those simulated over headphones).



(a)



(b)

Figure 6 Listening test results (indicating distorted signal amplitude offset to match reference signal loudness) for tests over (a) headphones and (b) loudspeakers

Figure 7 shows the statistical loudness level ranges (derived by calculating LUFS values with data from figure 6(a)) compared to the static loudness levels of each track. It can be seen that between two of the same track with different distortions (i.e. tracks 1 & 2, 4 & 5, 7 & 8, 9 & 10) versions with more severe distortions have a greater range of loudness values that are perceived as equal to the loudness of undistorted versions. In most instances where a compressive distortion (1, 2, 4, 5, 6, and 10) is used, an increase in distortion will cause listeners to perceive louder tracks to be equal in loudness to the undistorted tracks. In the case of track 3 where an expansive distortion was used, all listeners perceived the quieter distorted track to be of equal loudness to the undistorted counterpart which was louder.

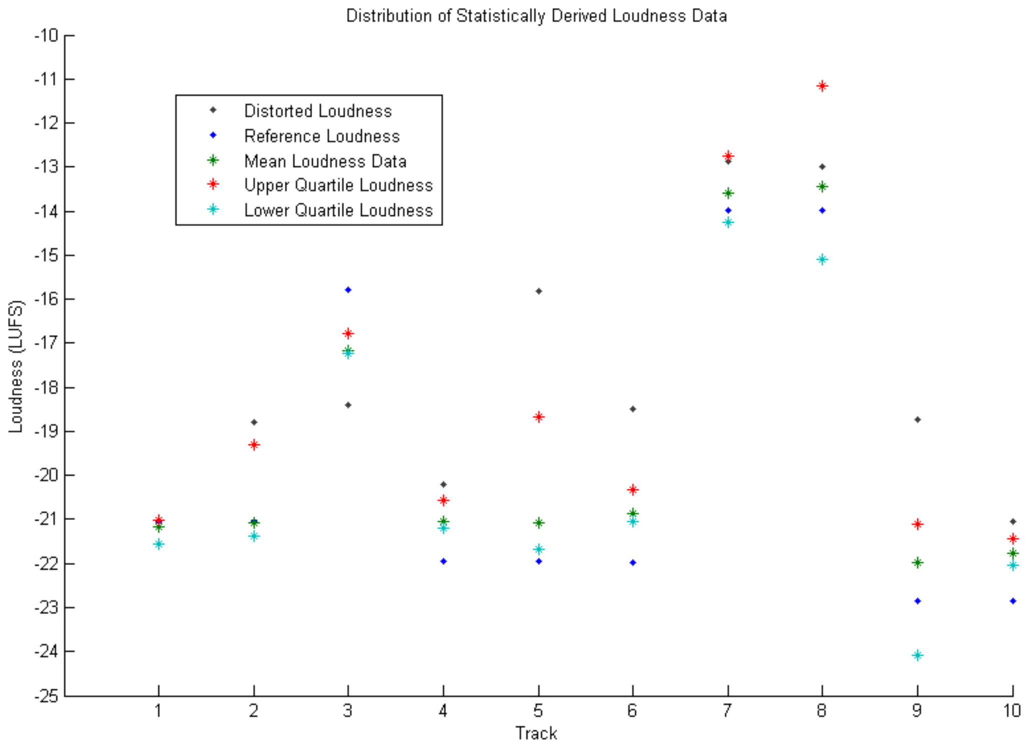


Figure 7 Distribution of loudness data (LUFS) for headphone-based listening tests

5 CONCLUSIONS & FUTURE WORK

The work presented in this paper has led to the following conclusions:

1. Distortion has an effect on the perception of loudness of music by way of increasing inter-listener difference in comparison based tests.
2. The base function (compression or expansion) and perceptual severity of the nonlinearity may play a part in inter-listener difference in perception.
3. Clipping causes music to become louder, but not equal in perceived loudness.
4. Clipping causes music to be perceptually equal in loudness to an undistorted reference, when the clipped music is louder that the undistorted music.
5. Polynomial nonlinear devices (8, 2 & 3) can be used to boost perceived loudness in a similar way to clipping, without degrading sound quality.
6. The effects of distortion on loudness do not appear to be genre-dependent as suggested in previous works.
7. The effects of distortion on loudness are very similar between loudspeaker systems and headphones.
8. Clipping loudspeaker systems will cause an increase in signal loudness, but will not be perceived as significantly louder without a relatively significant increase in loudness (i.e. soft clipping will increase loudness with less significance than hard clipping, but hard clipping causes a significant degradation of sound quality).

Further research into this topic must attempt to determine how different nonlinear devices influence engineers when processing music, to see if distortion has a direct influence on how live events are mixed. Research into the effective use of polynomial nonlinear devices may highlight the effectiveness of distortion as a tool, without significant degradation of sound quality. Ultimately, it must be determined whether any controlled distortion could be implemented in the digital domain.

6 AKNOWLEDGEMENTS

The authors would like to thank Simon Henin, Floyd Toole, Sam Wise and Brian CJ Moore for their correspondence over the course of this research.

7 REFERENCES

1. Fletcher, H., & Munson, W. A. (1933). Loudness, its Definition, Measurement and Calculation. *Journal of the Acoustical Society of America*, *5*, 82–108. Retrieved from http://psycnet.apa.org/?fa=main.doiLanding&uid=1934-01463-001

2. Robinson, D. W., & Dadson, R. S. (1957). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*. doi:10.1088/0508-3443/7/5/302

3. Rossing, T. D. (2007). *Springer Handbook of Acoustics*. doi:10.1007/978-0-387-30425-0

4. Everest, F. A., & Shaw, N. A. (2001). Master Handbook of Acoustics, 5th Edition. *The Journal of the Acoustical Society of America*. doi:10.1121/1.1398048

5. Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, *12*(1), 47–65. doi:10.1103/RevModPhys.12.47

6. Moore, B. C. J. (1996). Masking in the human auditory system. *Collected Papers on Digital Audio Bit-Rate Reduction*, 9–19.

7. Patterson, R. D. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788. doi:10.1121/1.388652

8. Leembruggen, G., Packer, N., Goldburg, B., & Backstrom, D. (1998). A Nonlinear model of the peripheral auditory system. *105th Convention of the Audio Engineering Society*, *0*.

9. Tan, C. T., Moore, B. C. J., Zacharov, N., & Mattila, V. V. (2004). Predicting the perceived quality of nonlinearly distorted music and speech signals. *AES: Journal of the Audio Engineering Society*, *52*(7-8), 699–711. doi:10.1121/1.3384932

10. Santis, E. M. de, & Henin, S. (2007). Perception & Thresholds of Nonlinear Distortion Using Complex Signals. *A master's thesis, Aalborg University*. Retrieved from http://projekter.aau.dk/projekter/fbspretrieve/9852082/07gr1061_Thesis.pdf

11. Toole, F. E. (2009). Sound Reproduction Loudspeakers and Rooms. *Noise Control Engineering Journal*. doi:10.3397/1.3124162

12. International Telecommunication Union. (2012). *Algorithms to measure audio programme loudness and true-peak audio level*. *ITU-R BS.1770-3* (Vol. 3). Geneva. Retrieved from http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-2-201103-I!!PDF-E.pdf

13. EBU. (2011). Practical guidelines for Production and Implementation in accordance with EBU R 128. *Ebu- Tech 3343*.

14. Wendl, M., & Lee, H. (2014). The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor. *Audio Engineering Society Convention 136*. Retrieved from http://www.aes.org/e-lib/browse.cfm?conv=136&papernum=9021

15. EBU. (2011). Loudness Metering: "EBU Mode" metering to supplement loudness normalisation in accordance with EBU R 128. *Ebu – Tech 3341*, (August), 1–10.

16. Zakrisson, J., & Borg, E. (1974). Stapedius Reflex and Auditory Fatigue. Stockholm.

17. Gido, S. P., Schulein, R. B., & Ambrose, S. D. (2011). Sound Reproduction within a Closed Ear Canal: Acoustical and Physiological Effects. *130th Convention*, 1–19.

18. Simpson, A. J. R. (2010). Auditory Perception of Dynamic Range in the Nonlinear System. In *128th Convention of the Aud. Eng. Soc.* (p. 22).

19. Shmilovitz, D. (2005). On the Definition of Total Harmonic Distortion and Its Effect on Measurement Interpretation. *IEEE Transactions on Power Delivery*, *20*(1), 526–528. doi:10.1109/TPWRD.2004.839744

20. Hofer, B. E. (2005). Measuring Distortion in Switching Amplifiers. *Audio Engineering Society*, 1–6.

21. Evans, E. F. (2001). Basic Physiology of the Hearing Mechanism. *Audio Engineering Society*, 50–57.

22. Tan, C.-T., Moore, B. C. J., & Zacharov, N. (2003). The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals. *J. Audio Eng. Soc*, *51*(11), 1012–1031. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=12197

23. Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models*. *Psychoacoustics: Facts and Models*. doi:10.1007/978-3-540-68888-4

24. Simpson, A. J. R. (2010). Auditory Perception of Dynamic Range in the Nonlinear System. *128th Convention of the Aud. Eng. Soc.*, 22.

25. Cabot, R. C. (1984). Perception of Nonlinear Distortion. *Audio Engineering Society*. Washington, Vancouver: Tektronix. doi:10.1121/1.389559

26. Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal-Audio Engineering Society*, *50*(5), 331–342. doi:10.1016/S0378-5955(03)00347-2

27. Moore, B. C. J., Glasberg, B. R., & Stone, M. A. (2003). Why Are Commercials so Loud? - Perception and Modeling of the Loudness of Amplitude-Compressed Speech. *AES: Journal of the Audio Engineering Society*, *51*, 1123–1132.

28. Camerer, F. (2014). The Path To Loudness Nirvana. In *BBC Academy*. London: BBC Academy. Retrieved from https://www.youtube.com/watch?v=EAYnI8ZQeHc

29. Vickers, E. (2010). The loudness war: Background, speculation, and recommendations. *Audio Engineering Society Convention 129*, 1–27. Retrieved from http://www.aes.org/e-lib/online/browse.cfm?elib=15598

30. Ronan, M., Sazdov, R., & Ward, N. (2014). Factors influencing listener preference for dynamic range compression. *Convention Paper 9176 : 137th Conference of the Audio Engineering Society*, 1 to 9.

31. Hjortkjær, J., & Walther-Hansen, M. (2014). Perceptual Effects of Dynamic Range Compression in Popular Music Recordings. *Journal of the Audio Engineering Society*, *62*, 37–41.

32. Technology, O. (2014). Reflective Object Sensor OPB608 Data Sheet. Carlton, TX: TT Electronics. Retrieved from http://optekinc.com/datasheets/opb608.pdf

33. Hackley, R., Olson, H., & McCoy, D. (1969). Loudness Meter. *The Audio Engineering Society*, *45*(1), 346. doi:10.1121/1.1972376

34. Moore, B. C. J. (1993). Characterization of Simultaneous, and Backward Masking. *Audio Engineering Society*.

35. Durbridge, S. E. (2015). *Does distortion have an effect on perceived loudness?* Derby.

36. Cherry, E. M. (1983). Amplitude and Phase of Intermodulation Distortion. *J. Audio Eng. Soc*, *31*(5), 298–304. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=4579