PROFILING THE DISTORTION CHARACTERISTICS OF COMMERCIAL MUSIC USING AMPLITUDE DISTRIBUTION STATISTICS

A Wilson Acoustics Research Centre, University of Salford, UK
B Fazenda Acoustics Research Centre, University of Salford, UK

1 INTRODUCTION

Sales figures suggest that the vast majority of music sales in recent decades are in digital media formats. As such the study of digital music signals is an area of much research, yet the understanding of quality-perception in matters of recorded audio lags behind many other areas of music signal analysis.

Within the manufacturing and services industries, quality can be broadly defined as the degree to which a set of inherent characteristics fulfils requirements. In audio, these requirements vary with the application. For example, a studio loudspeaker may aim to reproduce the input signal with minimal distortion – the inherent characteristics are well defined (linear complex frequency response, low harmonic distortion etc.) and the design and production is a straightforward engineering challenge that can be accomplished with sufficient resources. However, a hi-fi loudspeaker may aim to reproduce sound in a way that is most pleasing to most listeners – the inherent characteristics are less well defined and may include similar requirements to studio loudspeakers but with greater emphasis on subjective parameters and listening tests.

Based on previous literature, the quality of recorded music is expected to rely on a combination of subjective and objective parameters relating to the signal itself, the listener and the nature of the presentation[1]. In this manner, the quality of a musical recording is considered to be similar to the nature of quality in food products[2].

There have been a number of studies which can provide evidence to support the view that distortion in audio signals is related to the impression of quality. Often, it is hard-clipping that is studied as it is a well-defined phenomenon analytically[3,4]. Another category of distortion is often referred to as soft-clipping, which is less well-defined and can be described by a variety of different analytical models[5].

Statistical measures of the distribution of sample amplitudes can be used as measures of distortion, dynamic range compression, apparent loudness and other aspects of signal amplitude. The study of these measures can therefore be used in the study of audio quality. Signal amplitude distribution can be presented in the form of a probability mass function (PMF). A PMF shows the probabilities of a discrete random variable occurring at discrete values, in this case, the probability of a sample occurring at a given quantisation level. From the PMF, certain characteristics of the audio signal can be detected, such as clipping of the signal and errors in the analogue-to-digital conversion[6]. In some cases, the distribution is represented as an "amplitude histogram", where histogram bins are chosen based on decibel increments[7,8]. This approach can lack the required detail in high-amplitude values which is particularly relevant due to the fact that signal levels have increased in recent decades, in what is often described as a "loudness war"[9].

This paper presents an investigation into the perception of various categories of distortion present in commercial music recordings and the relationship to perceptions of audio quality. In addition, an investigation into the efficient representation of the PMF of high-definition audio formats is briefly discussed.

2 AMPLITUDE DISTRIBUTION IN DIGITAL AUDIO SIGNALS

A Probability Density Function (PDF) and Probability Mass Function (PMF) relate to the probability distributions of continuous and discrete variables respectively. Both terms are often used interchangeably in audio. There are some important differences which can be highlighted by considering the following example. For a sine function, written as $A\,sin(\omega t + kx)$, where the symbols have their usual meaning, the PDF can be expressed by the following expression, independent of frequency[10]. When A=1 and plotted for $-1 \le x \le 1$, the result is shown in Figure 1.

$$\frac{1}{\pi\sqrt[2]{A^2 - x^2}}$$

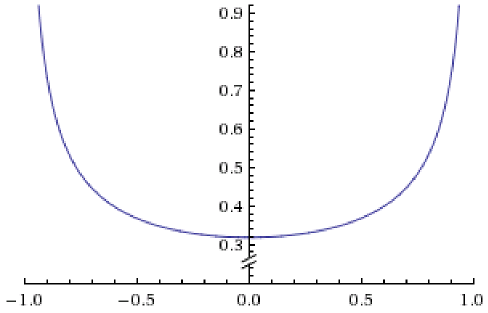


Figure 1 - Probability Density Function of Sine function

When sampled, the nature of PMF varies with the sampling rate, the frequency of the sine wave and the phase offset between the two systems. Consider a sine wave of frequency 1000Hz and a sampling frequency of 4000Hz; there are four samples per period, always at the same amplitudes. The examples below show the effect of a phase shift of π/4.

Example 1		Example 2	
amplitude	Probability	Amplitude	Probability
1	0.25	0.707	0.50
0	0.50	-0.707	0.50
-1	0.25		

These distributions do not resemble the PDF of a sine function due to the periodicity of the sampling with reference to the waveform. To obtain a PMF that approximates the PDF, the sine wave must be sampled evenly across the amplitude range. This can be seen in the following example, where f = 997Hz, a prime number, and the relative periodicity of sampling w.r.t. the wavelength is reduced.

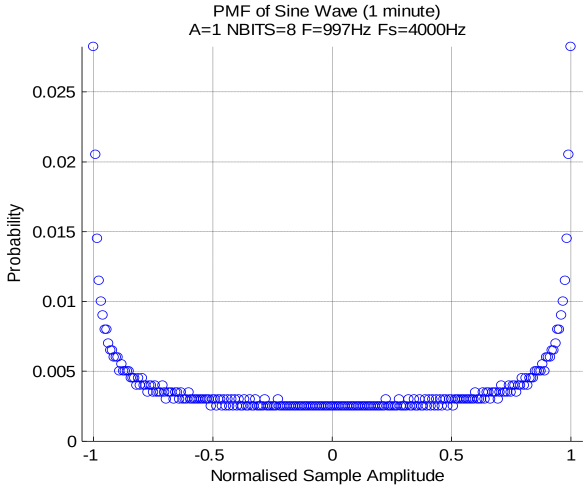
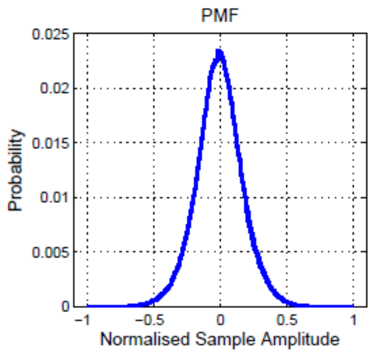


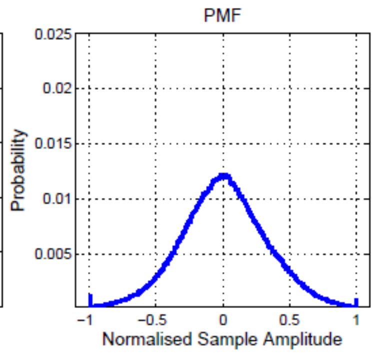
Figure 2 - PMF of sine wave, when similar to PDF

2.1 Analysis of CD audio

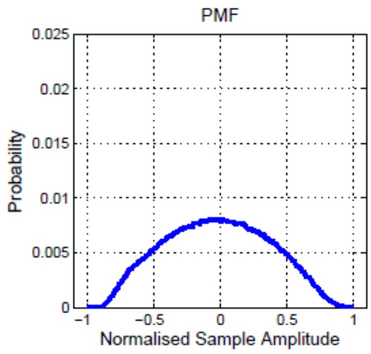
When considering audio stored on compact disc, at a sampling rate of 44.1kHz and a resolution of 16-bits, the distribution of these audio signals has been approximated by a number of models, such as a Gaussian or Laplacian distribution[11]. When the measured audio signal contains periods of silence at the start and end of the song, such as a gradual fade out, the distribution is dominated by a large peak centered about the zero-amplitude region. By contrast, Figure 3 shows the PMF of four audio samples (when evaluated as a histogram of 201 bins, for clarity), where the audio is a 20-second segment centered about the second chorus of the song. Importantly, no fade in or fade out has been applied. While the area under the curve is identical by definition, these four plots represent a variety of amplitude distributions over a ~30 year period.



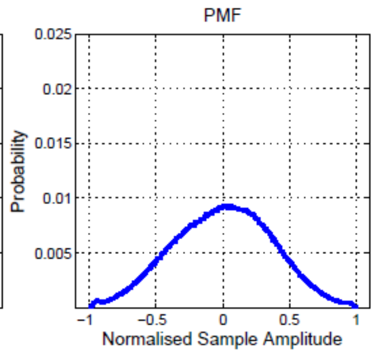
(a) *Europe, 'The Final Countdown', 1986*



(b) *Prodigy, 'Firestarter', 1996*



(c) *Slipknot, 'Duality', 2004*



(d) *Katy Perry, 'Roar', 2013*

Figure 3 - Histograms of CD audio amplitude distributions

2.2 Distortion Character

From the analysis of CD audio we consider three typical configurations of the PMF; while Figure 2a represents a typical predicted distribution and Figure 2b shows the well-defined 'hard-clipping', (although it appears subtle due to the histogram operation), Figures 2c and 2d show distributions which manage to produce even louder signal levels, without hard-clipping, by means of 'softer' distortions and time-dependant dynamic range compression.

We see that there are a number of possible outcomes when attempting to maximise the perceived loudness of digital music signals, as is often desired in modern music productions. By considering both distortion type as well as the intensity (distortion level) we obtain a two-dimensional paradigm. From this we refer to three discrete categories of 'distortion character', as illustrated in Figure 4.

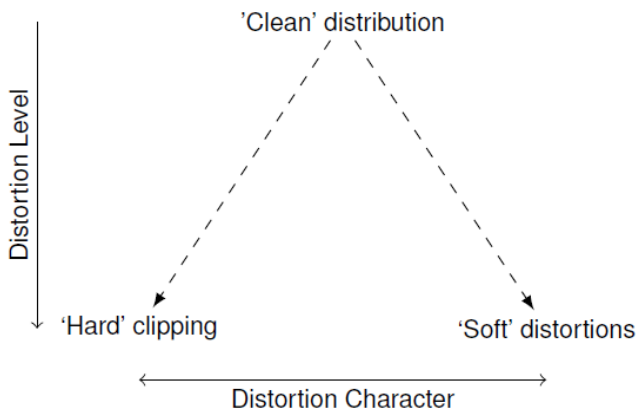


Figure 4 - 2D distortion paradigm, showing discrete categories

This is referred to as distortion character since the problem is comparable to character recognition. For example, while the letter **W** can be defined as such, **W** and *W* are still recognised as equivalent symbols, despite changes in styling. Even the naming of the symbol varies with language in ways which imply a particular style, such as "double-U" or "double-V". In this study, the shape of the PMF curve is the 'character' in the problem and comparable PMFs are considered to contain similar amplitude characteristics, despite subtle variations in both shape and description.

2.3 Audio dataset

The dataset of audio which was analysed contained audio clips from 321 songs by 229 artists. There is an average of ten samples per calendar year from 1982 until 2013. As described in Section 2.1, these were 20-second clips centered about the second chorus of the song, where possible. In other studies which consider the analysis of a large volume of musical samples, the datasets used contain digital representations of audio recordings originally released in analogue formats, for example, music from the 1970's and earlier[12,13]. What is not always made clear is how these samples were gathered – digital re-releases of previously released material can be considered to be re-masters and the amplitude distribution will not necessarily be comparable to the original release. For this reasons, only original digital content was used in this study, dating back to the commercial introduction of the Compact Disc in 1982.

3 CLASSIFICATION

For this study, only three categories of distortion character are considered – the three categories shown in Figure 4. All examples were labelled by visual inspection of the PMF along with audition.

The designing of a classifier, in this case, has two objectives.

- To label unseen samples with the appropriate distortion character, using a consistent metric
- To provide information on which objective features were used to perform this labelling

3.1 Feature extraction

In order to determine which features can be used to classify audio by distortion character, a series of features were extracted, mainly amplitude features due to the nature of the problem. The choice of features is also influenced by earlier studies into the nature of audio quality in which a number of these features were found to be correlated to subjective quality ratings[1]. These include predictions of the emotional response of the listener[14]. Feature extraction was aided by the use of the MIRtoolbox[15].

Table 1 - Features extracted from CD audio clips

Category	Feature	Description
Common measures of amplitude	Crest Factor	Ratio of peak amplitude to rms amplitude
	Loudness[16]	According to ITU BS. 1770-3
	Top1dB	Proportion of samples with amplitudes between 0dBFS and -1dBFS
Spectral features	Rolloff85[17]	Frequency at which 85% of the spectral energy is contained below
	Harsh Energy	Proportion of total spectral energy that is contained in the 2kHz to 5kHz band
	LF Energy	Proportion of total spectral energy that is contained in the 20Hz to 80Hz band
PMF features	PMF	Probability Mass Function, evaluated as histogram with 201 bins
	Centroid	First moment of PMF
	Spread	Square root of second moment of PMF
	Skewness	Third moment of PMF
	Kurtosis	Fourth standardised moment of PMF
	Flatness	Ratio of geometric and arithmetic means of PMF
	PMF_d	First derivative of PMF
	Gauss[1]	A measure of how well a Gaussian distribution can be fitted to PMF_d
Other	MIRemotion[14]	Eight objective predictions of listeners emotional response – Happy, Sad, Tender, Anger, Fear, Activity, Valence, Tension

3.2 Classifier design and performance

Classification was designed using Orange, a data-mining toolkit for python[18]. As the initial set of features extracted contains over 400 features, this number was reduced by means of recursive feature elimination[19], described as follows.

i. A list of features is provided and a linear support vector machine (SVM) is obtained.
ii. The features are ranked according to their weights in the SVM solution.
iii. The lowest-ranked feature is removed from the list.
iv. Repeat these steps until desired number of features remain.

This algorithm returned the ten features most relevant to distortion character classification from the initial set shown in Table 1; these were as follows. gauss, kurtosis, flatness, the 1st, 197th, 199th and 201st elements of PMF and the 1st, 79th and 200th elements of PMF_d. Since the PMF was sampled as a histogram of 201 bins, we see that the relevant features of the PMF are found to be, mostly, the extreme regions.

A new SVM implementation was created, with a multi-class configuration and using only the ten features output by the RFE process. The parameters of the SVM are automatically optimised using LIBSVM's procedures[20]. The dataset was randomly split so that 50% was used for training and 50% for testing. The trained classifier was tested using 10-fold cross-validation. The classification accuracy was 0.795, with area under ROC curve of 0.888. The confusion matrix for this test is shown in Table 2, with correct classifications in shaded cells.

Table 2 - Confusion matrix of classifier

		Predicted			
		clean	hard	soft	*Recall*
Real	Clean	73	5	4	0.89
	Hard	2	28	5	0.80
	Soft	5	12	27	0.61
	Precision	0.91	0.62	0.75	

We see that both recall and precision is highest for the ‘clean’ category, which indicates that these samples show a high degree of conformity which allows them to be easily recognised. The ‘hard’ category shows high recall also, as the hard clipping is easily determined by the features of PMF_d. The precision is lower, since samples with hard clipping can also have a more general PMF distribution indicative of other types, leading to misclassification into this category. In a similar manner, recall is low for the ‘soft’ category as this group is less well-defined and can be mis-classified more easily. Precision, however, is quite high, since other groups are unlikely to be classified into this group.

4 LISTENING TEST

A subjective listening test was performed in order to test the hypothesis that there is a perceptual difference between these three categories of audio signals. The null hypotheses are listed here.

1. There is no difference in quality ratings of the different audio clips.
2. There is no difference in quality ratings of the different distortion character groups.
3. There is no difference in how words are used to describe the quality ratings of different distortion character groups.

4.1 Test design

Of the 321 audio samples which were analysed, 63 were used in a listening test in which participants were asked to report their impression of the quality of the recording. The following questions were asked for each sample.

1. How do you rate the audio quality of this sample?
2. Please choose two words which describe the attributes on which you assessed the audio quality.

Participants rated the audio quality of each sample on a 5-point scale, with 5 as highest. For question 2, participants were provided with a list of commonly used terms as a reference but were encouraged to provide their own terms. The total number of participants was 22 and the mean age was 24.2 years (std.dev = 4.5 years). Mean test duration was 38 minutes (std.dev = 11 minutes).

The test took place in the listening room at University of Salford, which conforms to appropriate standards[21]. Playback of audio used Sennheiser HD 800 headphones; the frequency response was measured using a Brüel & Kjær Head and Torso Simulator (HATS). Low-frequency rolloff in the response below 110Hz was compensated using an IIR filter, designed using the Yule-Walker method. This then facilitated the addition of a notch filter at 0Hz. The perceived loudness of all audio samples was normalised according to ITU BS. 1770-3[16]. The presentation volume to participants was set to deliver a sound pressure level of 82dB(A), as measured using the HATS and sound level meter.

4.2 Results

With 63 audio samples and 22 subjects, these 1386 auditions were gathered and analysis was performed on this dataset. A one-way ANOVA was performed with post-hoc multiple comparison and Bonferroni adjustment applied. As shown in Figure 5, the mean quality rating is higher for the ‘clean’ category compared to the other two, while ‘hard’ and ‘soft’ distortion categories are rated similarly ($F(1, 2) = 5.72$, $p = 0.00$, $\eta^2 = 0.008$). This provides evidence in support of rejecting test hypotheses #1 and #2, however the effect size is considered to be small, as $\eta^2 < 0.01$. This is influenced by the narrow use of the scale and contributions from other variables, as seen in earlier tests[1].

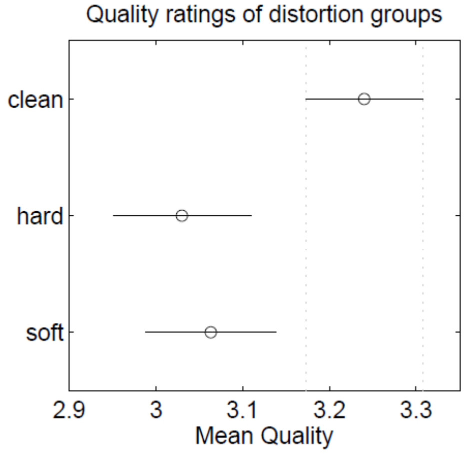


Figure 5 - Mean quality ratings, grouped by distortion character

In total, 255 words were used over the course of the 1386 unique auditions, after spelling had been corrected and equivalent terms collated. In this lexicon, many words are not used often, some being unique to a single participant. The five most frequently occurring terms are shown in Table 3 and account for 19.7% of all descriptions requested. A Chi-Square analysis was performed and shows that the words chosen to describe the quality of each distortion character differed significantly from chance ($\chi^2(8, N = 547) = 33.28, p = <.001$). This data provides evidence in support of rejecting test hypothesis #3. In Table 3, frequencies highlighted in bold (with '>' or '<') are either significantly greater than (>) or less than (<) the expected counts.

Table 3 – Frequency count of quality descriptors (chi-square analysis)

		Groups			
		clean	hard	soft	TOTAL
Words	*Distorted*	**21<**	**47>**	**59>**	127
	Punchy	53	37	**34<**	124
	Clear	49	30	45	124
	Full	29	28	30	87
	Harsh	**42>**	20	23	85

5 APPLICATION TO HIGHER BIT-DEPTHS

It is now common for an artist to release music online in a variety of formats - 'high-definition' audio is often provided as an alternative to formats based on lossy compression. The merits of this alternative are debateable. Shown in Figure 6 is the waveform of a 32-bit 96k WAV file of a music release, freely distributed from the artists website in early 2014. Samples at 0dbFS are highlighted in red. Hard-clipping is evidently a desirable stylistic choice, unrelated to audio format.

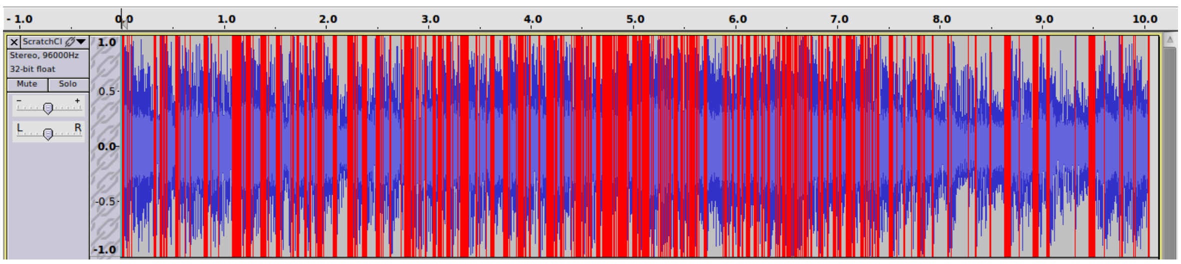


Figure 6 - Waveform of commercially released 32-bit audio (Left channel)

In order to compare the signal statistics of PCM audio at various bit-depths, a collection of audio was generated for study. Using multitrack audio sessions freely available online[22,23], the sessions were mixed using 32-bit float arithmetic and then exported as 32, 24 and 16-bit PCM audio files. In order for the audio signal to occupy the full dynamic range of the format, a small amount of clipping was permitted. The peak output level was typically +0.5dBFS, resulting in only a small number of clipped samples on export.

Figure 7 shows the PMF of the 24-bit version of one such audio sample. As the plot contains 2^24 data points, it is not possible to make out many details, however it can be seen that the number of counts for unique sample amplitudes is quite low; a given amplitude level contains an integer number of counts and the plot appears to have steps caused by these integers.

What is not immediately clear is the vast number of sample amplitudes with a probability equal to zero. This can be observed in Figure 8, a histogram where the number of unique sample amplitudes which display a given number of occurrences is displayed. The cumulative probability distribution of this histogram data, shown in Figure 9, shows that 99% of amplitude levels have three or fewer occurrences.

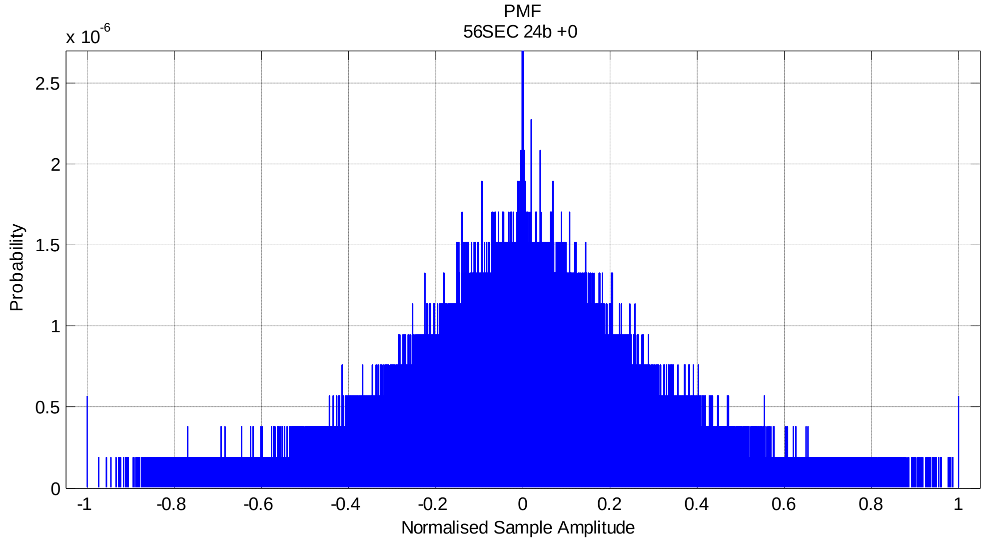
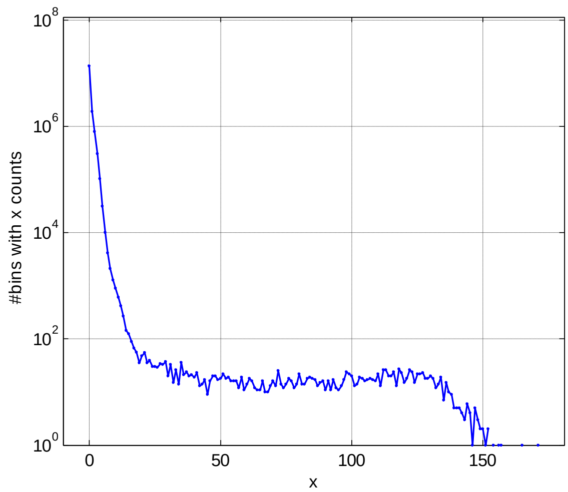


Figure 7 - PMF of 24 bit audio



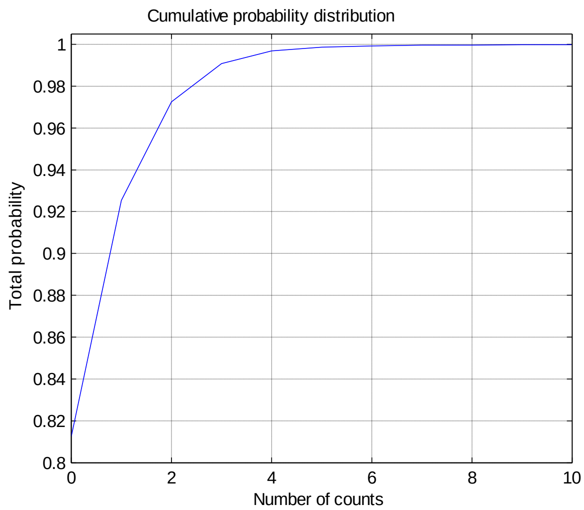


Figure 8 - Histogram of PMF in Figure 7

Figure 9 – Cumulative probability distribution of histogram in Figure 8

5.1 PMF of 24bit audio compared to 16bit

When comparing the 24-bit and 16-bit files, some shown in Figure 10, it can be seen that hard-clipping is more evident in the graphs of 24-bit audio. This is simply because the number of counts at the extreme amplitude values can be considerably large when compared to the very low number of counts in other regions. Note the change in y-axis scale between 24-bit and 16-bit plots.

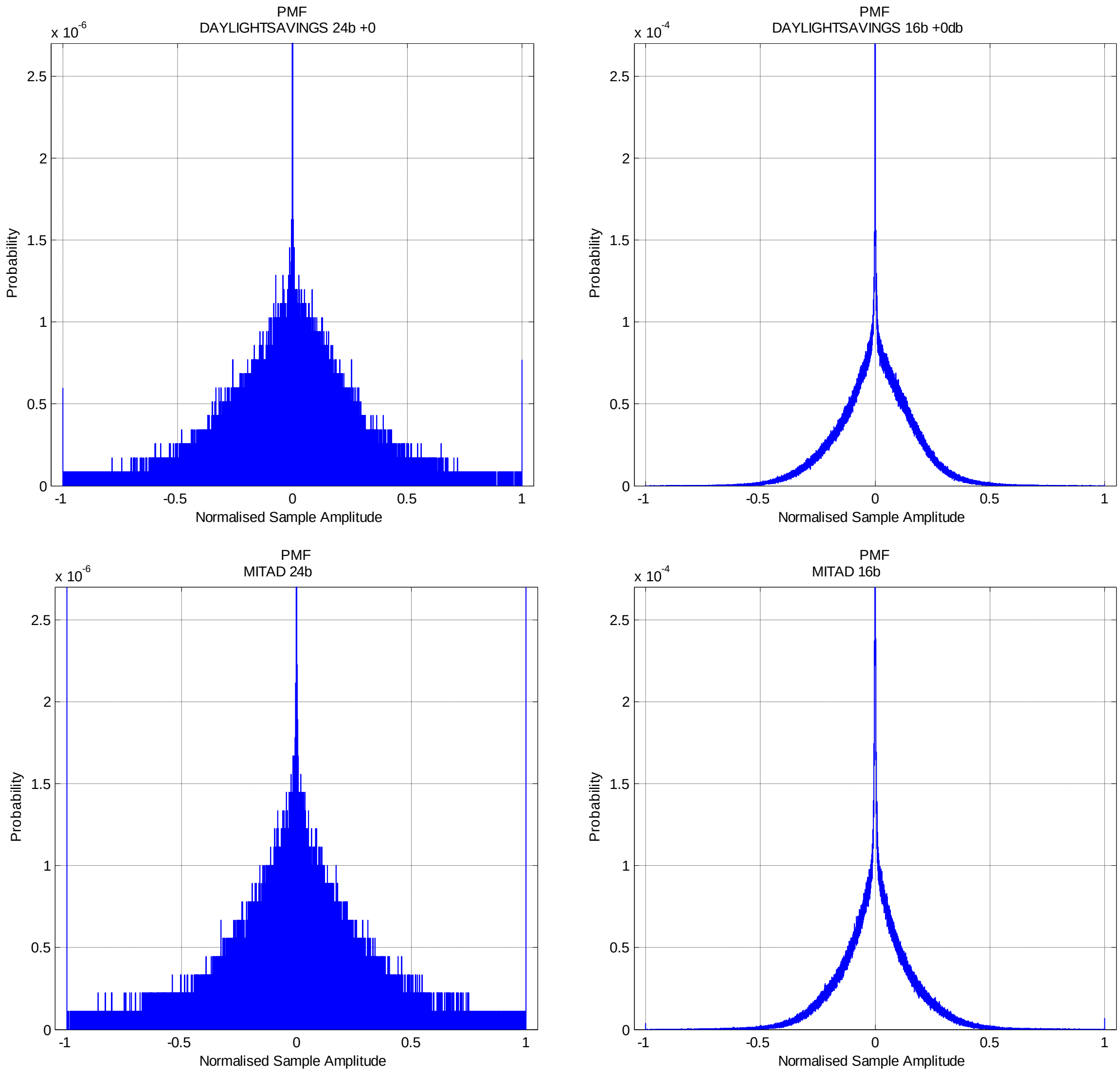


Figure 10 - Comparison of PMF of 24bit (LEFT) and 16-bit (RIGHT) audio

5.2 Sparse representation

In this 24 bit WAV file the number of amplitude levels with zero occurrences represents a high proportion of the total. This number is referred to as the 'sparsity' of the function. This sparsity can be exploited to apply data compression, which is useful when comparing a large number of PMFs from different audio files, as was done in Section 3. Sparsity, S, can also be determined as $S = \frac{n_{total} - n_{unique}}{n_{total}}$, where n represents the number of quantisation levels in the amplitude range. The measured sparsity of each audio signal at each bit-depth is shown in Table 4.

Table 4 - Sparsity of example digital audio files

Song	Length	S(16)	S(24)	S(32)
56 seconds	2:00	0.34605	0.81355	0.99880
Daylight Savings	4:25	0.27740	0.72705	0.99750
Heartbeats	4:23	0.28235	0.71280	0.99750
Lolita	4:23	0.32735	0.74250	0.99750
Mitad Del Mundo	3:24	0.24205	0.76630	0.99800

Consider a vector of length 2^24 where many elements are equal to zero. A sparse representation would contain only the non-zero elements as well as their indices in the vector. For this 1-D array, the indices will be from 1 to 16,777,216 and so 94% of these indices will be 8-digit numbers.

$$\mathrm{PMF}_{1\mathrm{D}} = [0\ 0\ 2\ 0\ \ldots 1\ 0\ 0\ 3 \ldots\ 2\ 2\ 1\ 0 \ldots\ 0\ 1\ 0\ 0 \ldots] \ \rightarrow\ \mathrm{PMF}_{2\mathrm{D}} = \begin{bmatrix} 0\ 0\ 2\ 0 \ldots \\ 1\ 0\ 0\ 3 \ldots \\ 2\ 2\ 1\ 0 \ldots \\ 0\ 1\ 0\ 0 \ldots \\ \vdots\ \vdots\ \vdots\ \vdots\ \ddots \end{bmatrix}$$

This 1-D array can be converted to a 2-D array, such as a square matrix of 4096*4096, where the n^{th} 4096 elements fill the n^{th} row. The number of digits required to index a particular value will be typically lower. For example, only when both co-ordinates are 1000 or greater will 8-digits be required to index. This allows savings to be made in terms of memory storage requirements.

This methodology can be extended to higher dimensions, such as arrangements of 256*256*256 or 64*64*64*64 in this case, however, the additional savings are relatively small. Table 5 shows the number of bytes required to store the PMF of 24-bit audio files in each of these described arrangements.

Table 5 - Filesize (in bytes) of various PMF representations

Song	Length	1D	Sparse 2D	Sparse 3D	Sparse 4D
56 seconds	2:00	134,217,736	50,087,384	41,077,204	42,738,008
Daylight Savings	4:25	134,217,736	73,307,280	59,864,906	62,510,539
Heartbeats	4:23	134,217,736	77,127,184	62,793,461	65,578,985
Lolita	4:23	134,217,736	69,147,368	55,715,117	57,967,609
Mitad Del Mundo	3:24	134,217,736	63,163,992	88,421,999	90,547,144
Average			66,5666,42	61,574,549	63,868,457
Ratio of 1D			0.50	0.46	0.48

5.3 32-bit audio

As shown in Table 4, when compared to 16 and 24-bit audio, the sparsity of 32-bit audio is greatly increased. Knowing that the function is highly sparse, it would be naïve to attempt to determine the complete PMF, with all 4,294,967,296 quantisation levels. Furthermore, the memory requirements can be prohibitive.

A single 32-bit PMF can be considered as 256 smaller 24-bit PMFs (or 65,536 16-bit PMFs), each of which demonstrates a high degree of sparsity (some containing no non-zero elements). The full PMF can then be determined by evaluating these smaller amplitude ranges one-by-one, saving only the non-zero elements and their indices before moving to the next segment. When these sections are combined, this yields a 1-D sparse array which could then be transformed to higher-dimensional structures for added compression. This method could be used in related future studies to store PMF data prior to feature extraction.

6 DISCUSSION

Since the dataset used contained roughly ten songs per year from 1982 to 2013, plotting the proportion of samples in each category against the release year shows that changing attitudes towards CD mastering can be measured, such as the increase in loudness and related distortions during the 1990s and a shift from hard to soft clipping in recent years.

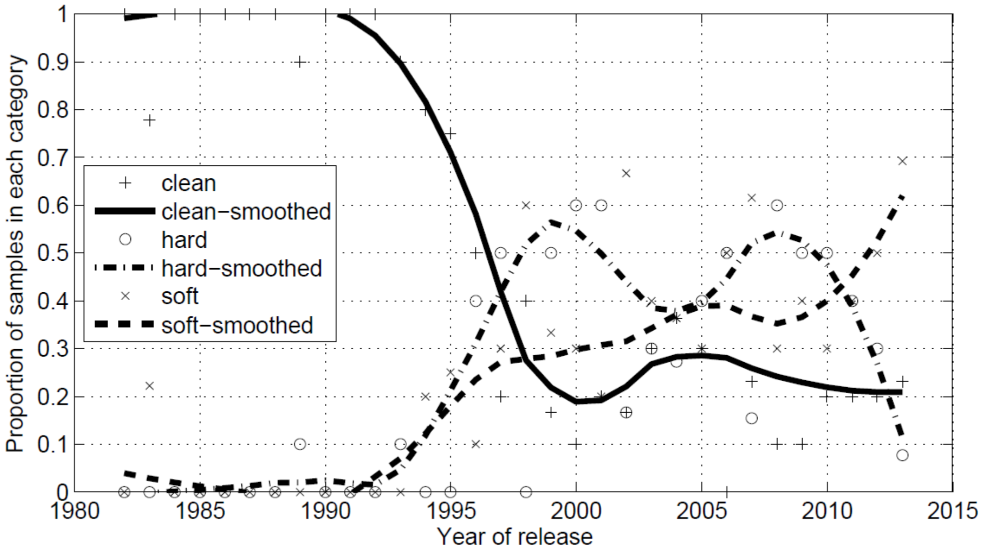


Figure 11 - Changes in production trends, 1982-2013

The trained classifier obtained an accuracy of ~80%. While this performance is good, the main aim of creating a classifier was to determine which objective aspects of the audio formed the basis for the initial labelling. The features found to be most relevant in the classification of these categories are all features relating to the PMF of the audio; the extreme values of the PMF and its first derivate are used to detect the presence of hard-clipping, the kurtosis of the PMF along with the value of the Gauss feature are used to detect the 'clean' samples, which are typically less loud and more dynamic that the other samples, leaving the 'soft' distorted samples.

It can be seen from the results of the listening test that there are measureable differences in how each of these three categories are perceived. The quality rating of the 'clean' samples is typically higher than the two distorted categories. This suggests that the perception of quality is related to increased dynamic range and related amplitude features. However, there was no significant difference in quality ratings for the two distorted groups. This suggests that any perceptual differences between them might not be quantitative. The analysis of descriptors of quality suggests this difference may be qualitative.

We see that 'distorted', 'punchy' and 'clear' were the three most-used terms, with similar numbers of occurrences. This indicates that distortion is a widely-understood concept upon which quality assessments are made and that the perception and audibility of distortion differs across the three groups. This result justifies the investigation into distortion character and other related concepts. Punch and clarity are two concepts that have received attention in recent studies[24]. It can be seen that the 'soft' category is less-often perceived as 'punchy', although there is no significant variation in the use of the term 'clear'.

The term 'full' was often used to describe quality ratings, although we see that this usage does not vary with distortion character. It is interesting that the term 'harsh' is used to describe the clean category significantly more often that would be expected by chance alone. This may be due to the following:

1. These samples are typically the earliest-released samples, from before the 'loudness war' (see Figure 11) and the changes to the typical spectrum of popular music since that time may contribute to variations in the perception of harshness[13].
2. As the perceived loudness of all samples was normalised, these more dynamic samples produced greater peak volumes, which may have been perceived as harsh by participants.

With artists releasing music in a variety of formats, many of the typical advantages of higher-definition audio formats are often not taken advantage of, due to aesthetic choices in the mixing and mastering processes, such as dynamic range compression, unless a separate master is produced. The study of amplitude distribution in higher-definition formats can be made easier by taking advantage of sparsity in the PMF to inform data compression. This eases the memory requirements of the test equipment and, for 24-bit formats, the computation time is relatively short. For 32-bit, computing the PMF is still very time-consuming when using a typical desktop computer.

7 CONCLUSIONS AND FURTHER WORK

A dataset of CD audio signals has been analysed and, based on observations, labelled by 'distortion character'. In order to be confident in the results pertaining to amplitude distributions of digital audio, and to report changes in audio production trends revealed by these distributions, it has been important to ensure that only original digital audio releases were used in the audio dataset.

The results of a subjective listening test show that participants can discern between these three groups. In this study, only the five most used quality descriptors are shown and only grouped by distortion character. Further analysis of these quality descriptors supplied by participants can be used to gain a more detailed insight into the differences between these groups and what features can be used to classify them. Of particular interest is the analysis of quality descriptors in relation to the actual quality ratings, which falls outside the scope of this paper. This ongoing investigation informs feature extraction processes which can be used to build a model of quality-perception in music recordings.

There is still further work to be done in the analysis amplitude distributions in higher-definition audio formats, which could inform testing into the perception of quality in these formats. A more efficient compression of PMF data for high-definition audio could be achieved by using prior knowledge of the amplitude distribution model to inform the process of allocating indices in the 2-D representation. For example, regions where both row and column indices are low can be reserved for quantisation levels expected to contain more than zero occurrences (such as near-zero amplitudes), while more 'expensive' regions of the matrix would be reserved for amplitude levels more likely to be unused, which, ideally, are at the extreme values. This is similar to techniques used in lossless audio encoding.

Overall, this study furthers the understanding of quality-perception in digital audio, by highlighting the relative importance of distortion and the perceptual differences between different types of distortion often used in the audio mastering process.

8 REFERENCES

1. A. Wilson and B. Fazenda, "Perception & evaluation of audio quality in music production", in Proc. Of the 16th Int. Conf. On Digital Audio Effects (DAFx-13), Maynooth, Ireland, Sept 2013
2. M. Ng, C. Chaya and J. Hort, "The influence of sensory and packaging cues on both liking and emotional, abstract and functional conceptualisations", Food Quality and Preference, Vol. 29, No. 2, pp.146-156, Sept 2013
3. N. Croghan et al, "Quality and loudness judgments for music subjected to compression limiting", Journal of the Acoustical Society of America, Vol. 132, no. 2, pp. 1178-1188, 2012

4. TJ. Cox et al, “Quality, timbre and distortion: perceived quality of clipped music”, in Proc. Institute of Acoustics – Conference on Reproduced Sound, Manchester, 2013
5. S. Enderby and Z. Baracskai, “Harmonic instability of digital soft-clipping algorithms”, in Proc. Of the 15th Int. Conf. On Digital Audio Effects (DAFx-12), York, UK, Sept 2012
6. E. Benjamin, “Characteristics of musical signals”, Audio Engineering Society Convention 97, 1994
7. M. Mijić et al, “Statistical properties of music signals”, Audio Engineering Society Convention 126, May 2009
8. J. Serrà et al, “Measuring the evolution of contemporary western popular music”, Scientific reports, Vol. 2, January 2012
9. E. Vickers, “The loudness war: background, speculation and recommendations”, Audio Engineering Society Convention 129, 2010
10. N. Cheremisinoff, L. Ferrante, “Practical Statistics for Engineers and Scientists”, CRC Press, 1987
11. Bell, Anthony J., and Terrence J. Sejnowski. "An information-maximization approach to blind separation and blind deconvolution." Neural computation, 1995
12. Deruty and D. Tardieu, “About dynamic processing in mainstream music”, Journal of the Audio Engineering Society, Vol. 62, No.1, 2014
13. PD Pestana, Z Ma and JD Reiss, “Spectral Characteristics of Popular Commercial Recordings 1950-2010”, Audio Engineering Society Convention 135, 2013
14. T. Eerola et al, “Prediction of multidimensional emotional ratings in music from audio using multivariate regression models”, International Conference on Music Information Retrieval, Kobe, Japan, Oct. 26-30, 2009, pp. 621-626, 20091
15. O. Lartillot and P. Toiviainen, “A matlab toolbox for musical feature extraction from audio”, Proc. Digital Audio Effects (DAFx-07), Bordeaux, France, Sept. 10-15 2007, pp. 237-244, 2007
16. ITU-R BS.1770-3, “Algorithms to measure audio programme loudness and true-peak audio level,” Tech. Rep., International Telecommunications Union, 2012.
17. G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, 2002.
18. J. Demšar et al, “Orange: Data mining toolbox in python”, Journal of Machine Learning Research, Vol.14, pp.2349-2353, 2013
19. I. Guyon et al, “Gene selection for cancer classification using support vector machines”, Machine Learning, Vol. 46, No. 1-3, pp. 389-422, March 2002
20. C. Chang and C. Lin, “LIBSVM: A library for support vector machines”, ACM transactions on intelligent systems and technology, Vol. 2, 2011
21. ITU-R BS 1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems”, Tech. Rep. International Telecommunications Union, Vol. 1, pp. 1-11, 1997
22. http://weathervanemusic.org/shakingthrough
23. http://www.cambridge-mt.com/ms-mtk.htm
24. S. Fenton and J. Wakefield, “Objective profiling of perceived punch and clarity in produced music”, Audio Engineering Society Convention 132, 2012