Categorisation of distortion profiles in relation to audio quality

Wilson, AD and Fazenda, BM

CHARACTERISATION OF DISTORTION PROFILES IN RELATION TO AUDIO QUALITY

Alex Wilson, Bruno Fazenda

Acoustics Research Centre,
School of Computing, Science and Engineering
University of Salford
Salford, UK
a.wilson1@edu.salford.ac.uk

ABSTRACT

Since digital audio is encoded as discrete samples of the audio waveform, much can be said about a recording by the statistical properties of these samples. In this paper, a dataset of CD audio samples is analysed; the probability mass function of each audio clip informs a feature set which describes attributes of the musical recording related to loudness, dynamics and distortion. This allows musical recordings to be classified according to their "distortion character", a concept which describes the nature of amplitude distortion in mastered audio. A subjective test was designed in which such recordings were rated according to the perception of their audio quality. It is shown that participants can discern between three different distortion characters; ratings of audio quality were significantly different ($F(1,2) = 5.72, p < 0.001, \eta^2 = 0.008$) as were the words used to describe the attributes on which quality was assessed ($\chi^2(8, N = 547) = 33.28, p < 0.001$). This expands upon previous work showing links between the effects of dynamic range compression and audio quality in musical recordings, by highlighting perceptual differences.

1. INTRODUCTION

While a single, consistent definition for quality has not yet been offered, it has an accepted meaning when applied to certain restricted circumstances, such as audio reproduction systems. Measurement standards exist for the assessment of audio quality [1, 2] however such techniques typically apply to the measurement of quality with reference to a golden sample; what is in fact being ascertained is the perceived reduction in quality due to destructive processes. One such example is in the case of lossy compression codecs in which the audio being evaluated is a degraded copy of the reference and the deterioration in quality is measured [3].

This study is concerned with the audio quality of "produced" music where there is no fixed reference and quality is evaluated by comparison with all other samples heard. In this manner, "audio quality" is perhaps better related to "product quality", as considered in consumer research, food science and sensory profiling in general. In these cases quality is based on multi-modal perception - partly influenced by objective parameters, such as sugar level in drinks, and partly by issues such as branding and packaging [4].

The assessment of audio quality in musical recordings, especially that of popular music, is therefore thought to be based on both subjective and objective considerations. The weighting of these two factors can vary by individual, depending on experience and expertise [5].

1.1. Signal statistics

The distribution of sample amplitudes in digital audio signals has been shown in a number of previous studies - such works have displayed the probability mass function (PMF) for a number of digitised recordings of popular music. A PMF shows the probabilities of a discrete random variable occurring at discrete values. Particular characteristics can be observed in the PMF, such as clipping of the signal and errors in the analogue-to-digital conversion [6]. Often, this distribution is represented as an "amplitude histogram", where bins are chosen based on decibel increments [7, 8]. Some summary features have been suggested [9, 10], however such a logarithmic approach lacks the required detail in high-amplitude values. A detailed investigation of high-amplitude distributions is particularly relevant due to the fact that signal levels increased in recent decades, in what is often described as a "loudness war" [11].

1.2. PMF distortion

Throughout the literature there is rarely much attempt to analyse this distribution in the required detail and provide summary features. Previous work by the authors has provided one such summary feature of the PMF, which relates to the level of distortion and the perception of audio quality [5].

Hard-limiting and dynamic range compression have been studied in relation to listener preference [12, 13]. Since these parameters are encompassed by the PMF of an audio segment, the previous study by the authors attempted to gather them into a higher-level feature. Since the PMF describes many possible states (here, it is the 2^{16} quantisation levels in a 16-bit audio signal), a histogram was generated with 201 bins, providing a good balance of runtime, accuracy and clarity of visualisation. In order to evaluate the shape of the distribution, particularly the slope and the presence of localised peaks, the first derivative was determined. For the ideal distribution this had a near-Gaussian form so the goodness-of-fit to a Gaussian distribution was obtained for each sample (ratio of the sum of squares of the regression and the total sum of squares). This was used as a feature describing loudness, dynamic range and related distortions, referred to previously and herein as 'Gauss'.

This feature was shown to relate significantly to subjective impressions of audio quality, when participants were asked to rate "the audio quality of the sample" [5]. While this feature can distinguish between distorted and non-distorted audio signals and give an approximation of the amount of distortion, the difference between different types and causes of distortion is not clear from this feature alone. This paper will describe a method which can be used to determine other aspects of distorted audio in addition to level.

2. TYPICAL DISTORTIONS IN MASTERED AUDIO

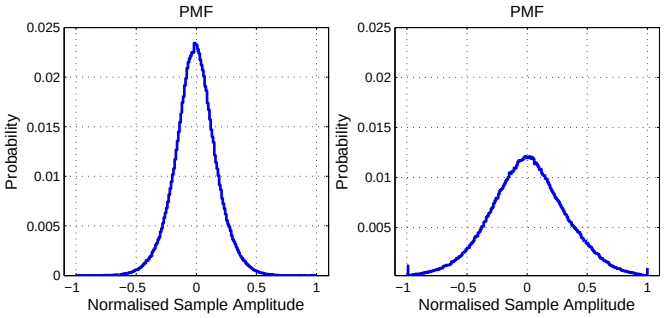
An examination of commercial music samples was undertaken in order to identify typical outputs of the mastering process and its visible imprints on the PMF. The nature of the sample amplitude distribution is influenced by the aforementioned loudness war, in which the perceived loudness of digital music signals has increased since the launch of the CD in 1982, at the expense of reduced micro-dynamics, achieved using dynamic range compression [10]. Despite the popular term, this may be thought of more as a "loudness race", as this increase takes place primarily in the 1990's and has remained at an escalated level since, in a state of *détente*.

Shown in Figure 1 is the PMF for a selection of audio samples. While the area under the curve is identical by definition, the shape varies. Figure 1a shows a distribution which is typical of its time. Due to the nature of its dynamic range, a distribution of this shape is often considered to be an ideal, neutral distribution, in relation to issues of loudness and dynamic range compression.

While hard-clipping of the waveform becomes increasingly popular during this "loudness race", as in Figure 1b, it becomes less common in recent years. Other PMF distributions have become popular, featuring a similar loudness level while avoiding hard-clipping. This can be achieved in a number of ways, one of which is to apply limiting to individual instrument groups during the mix process, or the use of multi-band limiters in the mastering chain. The awareness of inter-sample peaks has also lead some engineers to avoid the implementation of hard-clipping [14].

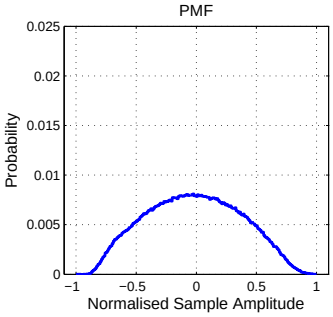
If a mix has been clipped the subsequent processing in the mastering stage, including equalisation, further dynamics processing, stereo-enhancement and downsampling from high sample rates, can cause this clipping to be spread out over a wider amplitude range, in regions around the maximal values, as in Figure 1d.

Figure 1c shows an example of a distribution highly warped in comparison to typical distributions and therefore the Gauss value is very low. Distortion, across the full mix, is evident on audition. It was worth noting that this album involved the same producer, mix-engineer and mastering-engineer as 'Death Magnetic', the 2008 album by Metallica which was responsible for a popular, if at times ill-informed, backlash against the loudness war [15, 16]. This demonstrates how a team of engineers can impart a distortion characteristic on productions and that this characteristic can be identified by listeners.

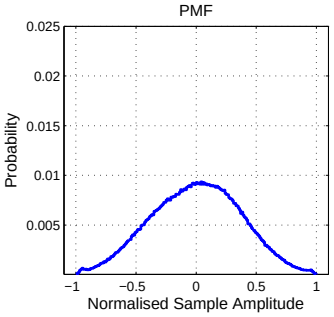


(a) *Europe, 'The Final Countdown', 1986*

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



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



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

Figure 1: *Examples of probability mass functions for digital music, showing a variety of production outcomes*

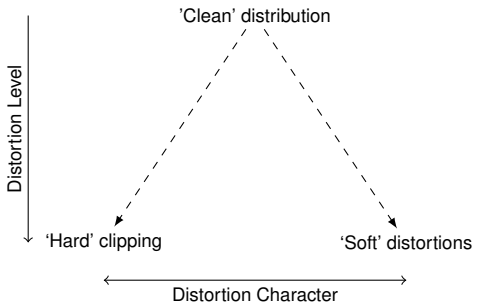


Figure 2: *2D distortion paradigm demonstrating three discrete characters*

2.1. Distortion character

It becomes apparent that hard-clipping is one of a number of possible outcomes when attempting to maximise the perceived loudness of digital music signals. Incorporating this distortion type into a two-dimensional paradigm with distortion amount introduces the notion of distortion character, as illustrated in Figure 2.

This is referred to as distortion character by the authors since the problem is essentially one of character recognition. For example, while the letter W can be defined as such, $\mathbf{W}$ and $\mathcal{W}$ are still recognised as equivalent symbols. In this case, the shape of the PMF curve is the 'character' in the problem and comparable PMFs are considered to contain similar amplitude characteristics.

2.2. Test hypotheses

This paper describes an investigation into the perception of different distortion profiles in relation to audio quality. In constructing a subjective test and its subsequent analysis, the following null hypotheses are used as a basis for the design.

- #1 There is no difference in quantitative quality ratings of the different audio clips.
- #2 There is no difference in quantitative 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.

Test hypothesis #1 was rejected in previous work by the authors featuring a similar experiment [5] but stands as the basic null hypothesis in this work.

2.3. Audio dataset

The dataset of audio examples is comprised of 321 songs by 229 different bands or artists. There is a mean of ten samples from each year between 1982 and 2013. The clips used for analysis are 20-second excerpts centred about the second chorus.

The audio is being collected as part of a larger study into the nature of quality-perception. While other studies have included digital audio files representing music from earlier periods [8, 9, 10] there is usually not much explanation as to how they have been sourced. In particular, what is often not acknowledged is that samples which represent music made prior to the advent of digital media would have been remastered for release on digital media. When these remasters were made is important; remastered audio does not typically retain the amplitude characteristics of the original release. To address this issue, only audio from original CD releases is considered here, i.e. from 1982 onwards.

There are only two samples from 1982, due in part to the fact that many of the earliest CD releases were re-issues of material recorded in previous years. Both of these releases feature an emphasis system, designed to compensate for deficiencies in the A/D conversion process, which at this time was based on earlier, 14-bit technologies. The signal had been subject to pre-emphasis, and de-emphasis was to be performed on-board the player. For this study, a de-emphasis filter was designed in order to retrieve representative amplitude statistics for any samples featuring pre-emphasis. Based on an available circuit analysis [17], the filter was an IIR design, constructed by use of the Yule-Walker method.

2.4. Labelling of distortion character

For simplification, only three categories of distortion character are considered - clean, hard and soft, as in Figure 2. The clean and hard categories are quite well-defined analytically, however the soft character is a set of PMF shapes having high dynamic range compression but without hard-clipping, such as Figure 1d, where small deformations in the PMF can be seen just below the extreme levels. Two options were available for labelling the dataset;

1. The samples could be labelled analytically, since hard clipping, or lack thereof, can be determined by the values of the PMF in its extreme values, after normalisation.
2. The samples could be labelled by an expert panel, by simultaneous audition of the signal and visual inspection of the PMF.

Method 2 was used due to the subjective nature of the problem and the fact that method 1 makes assumptions about the nature of the soft distortion character, which is harder to categorise. As a result of this labelling approach a classifier was designed to blindly label samples, as learned by the initial classification of the expert panel.

3. CLASSIFICATION

3.1. Feature extraction

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

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

Table 1: *Features used in objective analysis*

Feature	*Description*
Crest factor	Ratio of peak amplitude to RMS amplitude
Loudness	According to ITU BS. 1770-3 [19]
Top1dB	Proportion of samples between 0dBFS and -1dBFS
Rolloff	Frequency at which 85% of spectral energy lies below [20]
Harsh energy	Fraction of total spectral energy contained within 2k-5kHz band
LF energy	Fraction of total spectral energy contained within 20-80Hz band
MIRemotion	Objective predictions of emotional response - Happy, Sad, Tender, Anger, Fear, Activity, Valence, Tension [21]
PMF	Evaluated as a histogram with 201 bins - see Section 1.2
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 mean and arithmetic mean of PMF
PMF_d	First derivative of PMF
Gauss	Measure of distortion in PMF_d feature [5]

Table 1 shows features which were extracted from each sample in order to train the classifier. These are mainly amplitude-based features due to the nature of the problem. The evaluation of certain features was aided by the MIRtoolbox [18].

3.2. Classifier design

Statistical analysis was aided by the use of Orange, a data-mining toolbox for Python [22]. Orange can also be used as a visual programming environment and this was used for prototyping. Based on this prototyping stage, the decision was made to use support vector machines (SVM) for classification. This decision was made as it is a well-known method which can address both aims of the classifier, as mentioned in section 3.1, and as described below.

3.2.1. Optimisation

For optimisation and reproducibility, the final classifier was also implemented using Orange but with the Python-scripting interface. The SVM implementation in this package is that of LIBSVM [23]. As the initial set of features extracted contains over 400 features, this number was reduced by means of recursive feature elimination (RFE) [24]. The algorithm is described below.

1. A list of features is provided and a linear SVM is obtained.
2. The features are ranked according to their weights in the SVM solution.
3. The lowest-ranked feature is removed from the list.
4. Repeat these steps until desired number of features remain.

This algorithm was used to return the ten features most relevant to distortion character classification from the initial set shown in Table 1. All 321 audio clips were used in this analysis. The features found to be most important in classifying distortion character

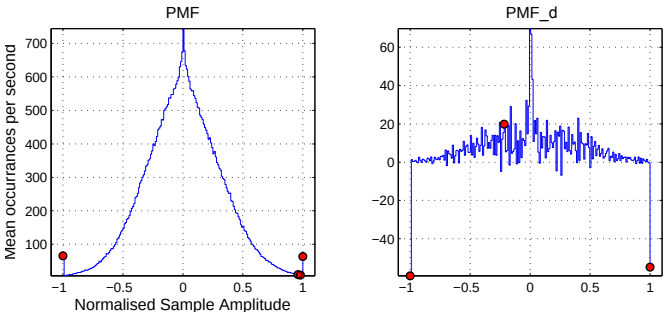


Figure 3: *PMF and (PMF)′ feature sets, in which the features important to distortion character classification are highlighted by circles*

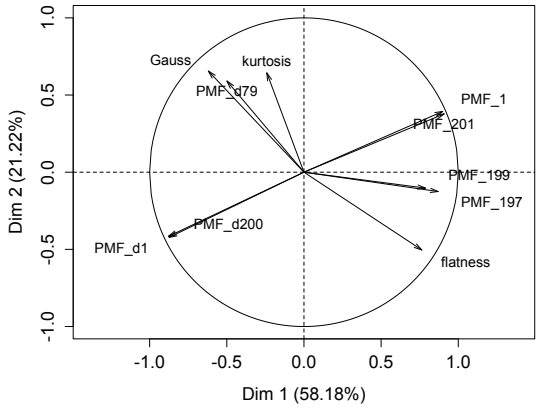


Figure 4: *Correlation of features with principal components*

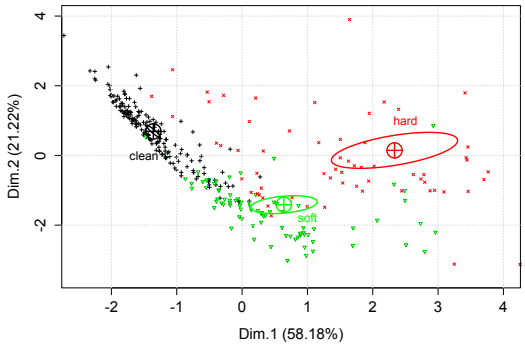


Figure 5: *Individuals factor map, showing the distribution of the three different groups*

were the following: gauss, kurtosis, flatness, the 1st, 197th, 199th and 201st elements of PMF and the 1st, 79th and 200th elements of PMF_d. Features that are a subset of PMF and PMF_d are shown in Figure 3, highlighted in PMF and PMF_d of audio which displays clipping, as evident from the values of these features.

A new SVM implementation was created, using a multi-class configuration. The parameters of the SVM are automatically optimised using LIBSVM's procedures [23].

3.2.2. *Performance*

This data was randomly divided into two portions; 50% for training and 50% for testing. The trained classifier was tested using 10-fold cross-validation and achieved a classification accuracy of 0.795, with area under ROC curve of 0.888. The confusion matrix for this test is shown in Table 2.

Table 2: *Confusion matrix showing performance of trained classifier on test dataset of 161 samples*

		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	

Both recall and precision is greatest for the 'clean' category. This indicates that there is a conformity between these examples and, as such, they can easily be recognised.

Recall is high for the 'hard' category, as this clipping is recognised easily by the PMF_d features (see Figure 3). However, precision is lower, as samples with hard clipping may have any other general PMF shape as identified by the gauss, kurtosis and flatness features. This leads to misclassification into this category.

Similarly, recall is low for the 'soft' class as this category is composed of a collection of PMFs that could not more easily be labelled as either of the two other groups. Precision is still rather high, as other groups are unlikely to be misclassified as members of this category. The most common misclassification is that of 'soft' as 'hard', likely due to the lack of conformity in the 'soft' group and reasons described above.

Table 3: *Variance explained by first five dimensions*

Dim.	1st	2nd	3rd	4th	5th
% var.	58.18	21.22	7.76	5.72	3.50
Cumulative % var.	58.18	79.39	87.15	92.87	96.37

3.3. Principal component analysis

In order to investigate how the ten identified features vary across the proposed three characters, a Principal Component Analysis (PCA) is performed. This yields dimensionality-reduction; the ten dimensions can be reduced to a combination of orthogonal functions which explain as much of the original variance as possible. PCA is performed with the 'FactoMineR' package, using **R**.

58.18% of the variance in the data is explained by Dim.1, which relates mainly to features associated with hard-clipping. The first two dimensions account for 80%, with Dim.2 describing the kurtosis and general 'peakiness' of the distribution. The variance of each dimension and cumulative variance is shown in Table 3.

Figure 5 shows the individual audio samples, grouped by distortion character. The centroid of each group is shown, with ellipses representing 90% confidence. This shows that each group is distinctly defined in this space. The axis limits were chosen for clarity; some outliers are not visible.

4. RELATION TO SUBJECTIVE IMPRESSION OF AUDIO QUALITY

4.1. Test design and execution

Of these 321 audio samples which were analysed, 62 were used in a listening test in which participants were asked to report their impression of the quality of the recording. This was assessed using the following instructions 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 list of words provided is shown in Appendix A. In this paper, the frequencies of these words are used in a *post-hoc* investigation of the perception of the three different distortion characters, i.e. to avoid bias, participants were not made aware of the distortion character concept prior to, or during, the listening test.

The test took place in the listening room at University of Salford, a room which conforms to BS.1116-1 [1]. Audio was delivered via Sennheiser HD 800 headphones, the frequency response of which was measured using a Brüel & Kjær Head and Torso Simulator (HATS). Low-frequency rolloff in the response below 110Hz was compensated for using an IIR filter designed using the Yule-Walker method. This then facilitated the addition of a notch filter at 0Hz.

The loudness of all audio samples was normalised, according to current broadcast standards, after headphone compensation had taken place [19]. The presentation level to participants was 82dB(A), as measured using the HATS and sound level meter.

One additional clip was added to the beginning of each test to serve as a trial. A short break was automatically suggested when 40% of trials had been completed. Post-experiment discussion was typically led by the participant and offered valuable insight.

4.2. Participant demographics

The total number of participants was 22 (4 female, 18 male), tested over a period of five days. The participants were 13 experts and 9 non-experts, which was self-reported, based on their level of academic or professional experience in fields relating to acoustics and audio. The mean age of participants was 24.2 years (std.dev = 4.5 years), varying from 19 to 39. Participants were asked to indicate their preferred musical genres and it was observed that the participants had diverse musical tastes.

Test duration varied by participant, with a mean value of 40 minutes (std.dev = 11 minutes). As this contained the option of a short break, the effect of fatigue on the reliability of subjective quality ratings was considered to be negligible, according to suggested guidelines [25].

4.3. Results

With 63 audio samples and 22 subjects, these 1386 auditions were gathered and analysis was performed on this dataset. As this test was also concerned with variables outside the scope of this paper, an n-way ANOVA was performed. This revealed a significant effect of the variables relevant to this paper, in terms of the influence on quality ratings, which were investigated further.

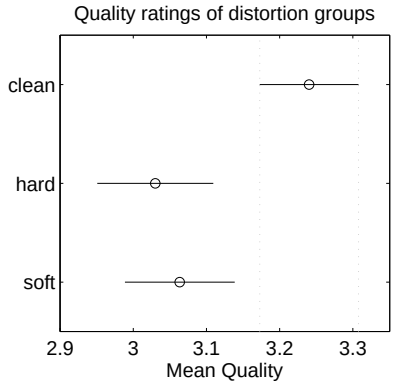


Figure 6: *1-way ANOVA: Quality, grouped by distortion character*

A one-way ANOVA was performed with *post-hoc* multiple comparison and Bonferroni adjustment applied. As shown in Figure 6, 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 [5].

4.3.1. Analysis of words used to explain quality ratings

While the provided list contained 41 words, in total, 255 words were used over the course of the 1386 unique auditions, after spelling had been corrected and equivalent terms collated (for example, 'compressed' and 'over-compressed' were equivalent in this context). In this lexicon, many words are not used often, some being unique to a single participant. While this study is comparatively small, connections between a words frequency and rank in a frequency table are found in other studies of linguistic corpora [26].

Figure 7 shows word clouds of the terms used to describe the participants' quality ratings, generated using **R**, along with the packages 'tm' [27] and 'wordcloud'. The five most frequently occurring terms are shown in Table 4 and account for 19.7% of all descriptions requested. In order to determine if there was significant variation in the frequency of each term across the three categories, a Chi-Square analysis was performed. The words chosen to describe the quality of each distortion character differed significantly (χ^2(8, N = 547) = 33.28, p = <.001). This data provides evidence in support of rejecting test hypothesis #3. In Table 4, frequencies highlighted in bold (with '>' or '<') are either significantly greater than (>) or less than (<) the expected counts.

		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

Table 4: *Frequency count (Chi square test analysis) of five most commonly used words*

(a) *'clean' category* (b) *'hard' category* (c) *'soft' category*

Figure 7: *Word clouds, showing the most used terms for each category. Larger/darker text indicates greater frequency.*

5. DISCUSSION

5.1. Feature reduction

The RFE process returned the most important features for classification. Perhaps unsurprisingly, all of these features are directly associated with the PMF. Other features such as crest factor or loudness (see Table 1) are indirectly encoded in the PMF, while spectral features were ranked lower. The emotion features 'Happy' and 'Anger' [21] were found to relate to audio quality in a previous study [5] and were included as they encode amplitude information. However these features were not highly ranked in this case.

The features found to be most useful are those bins of the PMF histogram close to extreme values which can detect the presence or absence of clipping, the gauss feature which can discriminate the 'clean' samples from the other groups, and kurtosis and flatness which can help to isolate the 'soft' category. PCA results in Figure 5 show that the three categories are well-defined by these features.

5.2. Production trends

Figure 8 shows the proportion of samples in each year of analysis which belong to each of the three distortion character categories. The scattered data is smoothed by local regression using weighted linear least squares and a second-degree polynomial model method, with rejection of outliers. As there is an average of ten audio clips per year, this data can be seen as illustrative rather than conclusive. From this plot, a number of discussion points are evident.

1. The transition from more dynamic signals to more compressed signals occurs throughout the 1990's.
2. The percentage of 'clean' samples has remained stable since.
3. Recent years have seen a move away from hard clipping, and towards the use of softer distortions in the final master.

Historical analysis of the gauss feature has shown that values range from an "early digital" phase to a "modern digital" phase, via a transition phase [5], referred to earlier as a loudness race. When distortion character is taken into consideration, a similar three-stage effect is observed in the clean category.

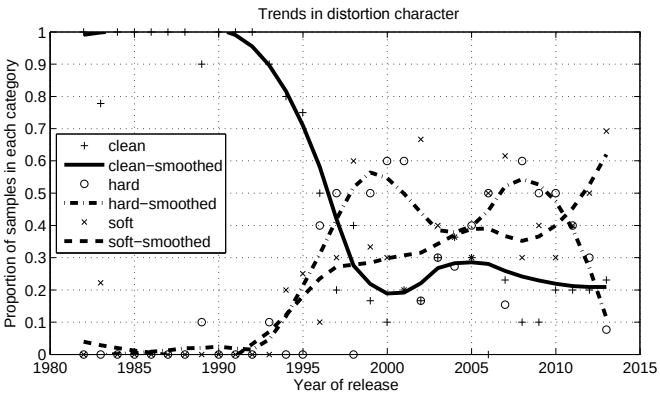


Figure 8: *Timeline showing changes in production trends and relative usage of each distortion character*

5.3. Differences in perception of each distortion character

While participants rated the quality of the clean samples higher than both distorted groups, there was no significant difference found between the two individual distorted groups. There was however, a difference in the way quality was perceived, as the distribution of descriptive terms varies between categories. From Figure 7 it can be seen that the number of words used for the 'clean' category is higher, whereas the word clouds of the other two categories are dominated by a small number of terms. A discussion of the influences of these words is provided below.

5.3.1. Distortion

'*Distorted*' was the most frequently occurring word overall. This indicates it is easily recognised by listeners and is a primary descriptor of quality, or lack therof. Table 4 shows it is used less than statistically expected by chance alone to describe the 'clean' category and more so for both other categories. This indicates that the three distortion characters do represent different levels of distortion, as illustrated in Figure 2 and Figure 5.

5.3.2. Punch and clarity

The adjectives '*punchy*' and '*clear*' are two of the most frequently used terms throughout, however they are used in varying amounts depending on the distortion character. This suggests the relative importance of such terms and also how they may be measured objectively, a task that has been investigated in recent literature [28].

'*Clear*' is used relatively less often for the examples with hard-clipping, although it is not significant. That the frequency of '*punch*' is lower for the soft category may simply be a result of other words being used more frequently. However, objectively, hard clipping would result in inter-sample peaks in subsequent stages of amplification and reproduction which could be interpreted as additional dynamic range, whereas this effect would not be so great for the 'soft' category.

5.3.3. Harshness and Fullness

The description '*full*' was used often but there is little variation in use across these three groups. This indicates that when participants used the word to explain why a particular numerical quality rating was awarded, the decision was not concerned with the distortion character but other factors. '*Harsh*' was often used by participants and there are a number of possible explanations for this.

- Participants' sensitivity to the headphones used, the response of which may have sounded unfamiliar to some participants
- This word was used more often for the 'clean' category, which is the dominant distortion character for the older material used. Changes in the typical spectrum of music recordings since this period may have had an influence [29].
- Additionally, under loudness-normalisation, the more dynamic nature of the 'clean' samples results in higher peak volumes and a transient response that some listeners may be less accustomed to. This is effectively the opposite of the common complaint among audiophiles that compressed music sounds flat and lifeless under loudness-normalised conditions.

5.4. Side-effects of loudness normalisation

This last point came up in post-experiment discussion with some participants and also in certain words used to describe particular songs. A smooth jazz sample was described as '*compressed*' and '*distorted*' by a number of participants (one using the term '*over-compressed*') as, when played at the same perceived loudness as other samples, it sounded unnaturally loud. Also, as the sample featured very subtle percussion the crest factor was lower than its distortion character would suggest. These issues indicate that, with loudness-normalisation, choosing a playback volume that does not bias against any one particular distortion character is difficult.

6. CONCLUSIONS

This work has set out to investigate whether commercial music samples can be categorised according to distortion level and type and does this categorisation further the understanding of audio quality in the context of modern commercial music.

It has been seen that the concept of a distortion character, informed by subjective perception, relates to certain objective measures of the PMF, namely particular regions as dictated by certain bins of the histogram, as well as summary features such as the statistical moments. The quantitative and qualitative aspects of quality ratings varied significantly for the three groups. This relationship between distortion character and quality ratings can contribute towards the understanding of quality-perception in the context of recorded music as well as inform attempts at evaluating the quality of an unknown audio stream.

6.1. Further work

6.1.1. Application to greater bit depth

These findings apply to digital audio with a bit-depth of 16 bits and a sampling rate of 44.1kHz. The ratio of quantisation levels to samples per second is 1.49:1 and this allows the PMF to be sufficiently non-sparse. For 24-bit resolution it would take a sampling rate of 11.25MHz to achieve the same ratio of quantisation levels to sampling rate. Thus, many distortion artefacts present in the 16-bit PMF will take a different form in systems with a greater bit-depth. To the authors knowledge, there have not yet been published studies analysing 24 or 32-bit digital audio in this manner, where even at the highest sampling rates commonly used, the PMF would be highly sparse. Further work would involve testing the methods described in this paper on 24-bit audio where discrete sample level counts are close to zero or equal to zero. The PMF of 32-bit audio, with 2^{32} levels, is likely to be prohibitively large to be able to study a large dataset of audio samples with the techniques described. New techniques are currently under trial.

6.1.2. Modelling distortion in mastered music

While clipping is well defined and evident in waveforms of mastered audio, soft distortions vary in their complexity, with varying dynamic and harmonic stability [30]. Further work is required to determine whether such analytical models can be used to describe how soft distortion appears in mastered commercial releases.

6.1.3. Quality-prediction

This study indicates that distortion character may not contribute greatly to a solely objective model of audio quality but does indicate that subjective elements, such as perceived punch, clarity and harshness, can provide useful information. Regarding the study of perceived audio quality in commercial music, the results related to dynamic range compression will be added to findings in other areas, such as overall balance of instruments and emotional response of the listener, to give a wider picture of how we understand audio quality in this context.

7. REFERENCES

[1] 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, 1997.

[2] ITU-T P.800, "Methods for objective and subjective assessment of quality," Tech. Rep., International Telecommunications Union, 1996.

[3] Amandine Pras, Rachel Zimmerman, Daniel Levitin, and Catherine Guastavino, "Subjective Evaluation of mp3 compression for different musical genres," *Audio Engineering Society Convention 127*, 2009.

[4] 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.

[5] Alex Wilson and Bruno Fazenda, "Perception & evaluation of audio quality in music production," in *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013, pp. 1–6.

[6] Eric Benjamin, "Characteristics of musical signals," *Audio Engineering Society Convention 97*, vol. 4, 1994.

[7] Miomir Mijić, Draŝko Maŝović, Dragana Ŝumarac Pavlović, and Milan Petrović, "Statistical properties of music signals," in *Audio Engineering Society Convention 126*, May 2009.

[8] Joan Serrà, Alvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll Arcos, "Measuring the evolution of contemporary western popular music.," *Scientific reports*, vol. 2, pp. 521, Jan. 2012.

[9] D Tardieu, E Deruty, C Charbuillet, and G Peeters, "Production effect: audio features for recording techniques description and decade prediction," in *Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11)*, 2011, pp. 441–446.

[10] Emmanuel Deruty and Damien Tardieu, "About Dynamic Processing in Mainstream Music," *Journal of the Audio Engineering Society*, vol. 62, no. 1, 2014.

[11] Earl Vickers, "The loudness war: Background, speculation, and recommendations," *Audio Engineering Society Convention 129*, pp. 1–27, 2010.

[12] Naomi B H Croghan, Kathryn H Arehart, and James M Kates, "Quality and loudness judgments for music subjected to compression limiting.," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 1177–88, Aug. 2012.

[13] T.J. Cox, B.M. Fazenda, S. Groves-Kirkby, I.R. Jackson, P. Kendrick, and F. Li, "Quality, timbre and distortion: perceived quality of clipped music," in *Proc. Institute of Acoustics - Conference on Reproduced Sound 2013. Manchester, UK*. November 2013, Institute of Acoustics (UK).

[14] Søren H Nielsen and Thomas Lund, "Level control in digital mastering," in *Audio Engineering Society Convention 107*. Audio Engineering Society, 1999.

[15] Ethan Smith, "Even heavy-metal fans complain that today's music is too loud!!!," *Wall Street Journal*, 25 September 2008. Available: http://online.wsj.com/news/articles/SB122228767729272339, accessed 18 March 2014.

[16] S. Michaels, "Death magnetic 'loudness war' rages on," *The Guardian*, 1 October 2008. Available: http://www.theguardian.com/music/2008/oct/01/metallica.popandrock, accessed 18 March 2014.

[17] Gary Galo, "A De-Emphasis Test CD," http://www.audioxpress.com/files/2009/02/galo3025.pdf, accessed December 27, 2013.

[18] Olivier Lartillot and Petri Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, Sept. 10-15, 2007, pp. 237–244.

[19] ITU-R BS.1770-3, "Algorithms to measure audio programme loudness and true-peak audio level," Tech. Rep., International Telecommunications Union, 2012.

[20] George Tzanetakis and Perry R. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[21] Tuomas Eerola, Olivier Lartillot, and Petri Toiviainen, "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models," in *International Conference on Music Information Retrieval*, Kobe, Japan, Oct. 26-30, 2009, pp. 621–626.

[22] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan, "Orange: Data mining toolbox in python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.

[23] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[24] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, Mar. 2002.

[25] Raimund Schatz, Sebastian Egger, and Kathrin Masuch, "The impact of test duration on user fatigue and reliability of subjective quality ratings," *Journal of the Audio Engineering Society*, pp. 63–73, 2012.

[26] George Kingsley Zipf, *Human behaviour and the principle of least effort*, addison-wesley press, 1949.

[27] Ingo Feinerer, Kurt Hornik, and David Meyer, "Text mining infrastructure in R," *Journal of Statistical Software*, vol. 25, no. 5, pp. 1–54, March 2008.

[28] S Fenton and J Wakefield, "Objective profiling of perceived punch and clarity in produced music," *Audio Engineering Society Convention 132*, pp. 1–15, 2012.

[29] PD Pestana, Z Ma, and JD Reiss, "Spectral Characteristics of Popular Commercial Recordings 1950-2010," *Audio Engineering Society Convention 135*, pp. 1–7, 2013.

[30] Sean Enderby and Zlatko Baracskai, "Harmonic instability of digital soft clipping algorithms," in *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, York, UK, September 2012.

A. APPENDIX 1 - LIST OF WORDS PROVIDED TO PARTICIPANTS

Bright, dark, loud, quiet, mellow, clear, clean, punchy, dull, bland, dense, exciting, weak, strong, sweet, shiny, fuzzy, wet, dry, distorted, realistic, spacious, narrow, wide, deep, shallow, aggressive, light, gentle, cold, hard, synthetic, crunchy, hot, rough, harsh, smooth, thin, full, airy, big