Audio Engineering Society

Convention Paper

Presented at the 132nd Convention
2012 April 26–29 Budapest, Hungary

OBJECTIVE PROFILING OF PERCEIVED PUNCH AND CLARITY IN PRODUCED MUSIC

Steven Fenton[1] and Jonathan Wakefield[2]

[1] School Of Computing & Engineering (MTRG), University of Huddersfield, Huddersfield, UK
s.m.fenton@hud.ac.uk

[2] School Of Computing & Engineering (MTRG), University of Huddersfield, Huddersfield, UK
j.p.wakefield@hud.ac.uk

ABSTRACT

This paper describes and evaluates an objective measurement that profiles a complex musical signal over time in terms of identification of dynamic content and overall perceived quality. The authors have previously identified a potential correlation between inter-band dynamics and the subjective quality of produced music excerpts. This paper describes the previously presented Inter-Band Relationship (IBR) descriptor and extends this work by conducting more thorough testing of its relationship to perceived punch and clarity over varying time resolutions. A degree of correlation is observed between subjective test scores and the objective IBR descriptor suggesting it could be used as an additional model output variable (MOV) to describe punch and clarity with a piece of music. Limitations have been identified in the measure however and further consideration is required with regards to the choice of threshold adopted based on the range of dynamics detected within the musical extract and the possible inclusion of a gate as utilised in some loudness algorithms.

1. INTRODUCTION

In music production and mastering, there is often a need to achieve recordings that possess certain criteria, for example punch and clarity. Whether these criteria have been met, and as such whether the production has achieved a level suitable for professional release, is often only known through peer assessment and/or comparison to a known good reference. Re-modification of a mix in order to satisfy these criteria is both time consuming and expensive. Even then, further ‘tweaks’ to the mix are often done through a process of trial and error until the desired mix outcome is achieved. On this basis there is clearly a need for research in this area to enable the objective quality assessment of music at the point of production – for example in the recording studio, production or mastering suite.

This paper describes a method for the objective measurement of produced music quality based upon a multiband dynamic analysis technique. The method focuses on the relationship between the dynamics of different frequency bands within a piece of produced music. This is quantified as an Inter-Band Relationship score (IBR). This paper aims to show that this score maps to the perceived listener experience with respect to the punch and clarity of the music under test.

Improving objective measurement tools associated with produced music quality assessment should help to streamline and improve the production process. Current metering tools used within the music production process are described in the following section.

1.1. Common metering tools used in music production

It is common practise in music production to view audio waveforms in their time-domain format i.e. a magnitude vs. time domain plot. Whilst this view allows for the engineer to ascertain relative magnitude levels at points in time of the production thus allowing identification of sections such as verse or chorus, it does not indicate such criteria as relative loudness, clarity, punch and so forth.

Frequency domain format plots are utilised in some audio production tools [4][5], and they are used when analysing the spectra of the audio under test. Due to the nature of their employment in these tools, their usefulness is debatable considering that the display is ever changing and often the time domain is simply represented by the refresh rate of the display.
This means that relative magnitude changes are impossible to identify over a long term window.

Their main use comes to the fore if there is the presence of frequencies introduced by mains hum, room modes etc. In these cases, one would see the relevant harmonic on the display in both the short and medium term and therefore its identification and removal is easily achievable.

Spectrogram plots allow a little more detail with respect to identification of rhythmic content by inspection and if one were to determine the relative harmonic content of each instrument present in the production, the clarity of each. In addition they display the spectral density of the signal over time, therefore, they could be utilised to indicate relative loudness between sections of audio on inspection.

Dynamic range in music production is an important criteria with respect to allowing passages of music to convey expression in addition to allowing the various elements of a mix to have enough headroom to indicate transient behaviour. Metering tools such as the Pleasurize Sound DR meter [6] generally sit on the main mix bus and give the engineer an indication of the dynamic range expressed in decibels. Their use in music production is somewhat limited as they only give an indication of the headroom available at a particular moment in time and don't give an indication of the transient behaviour of the signals under measure.

Loudness meters [1][2] and standards [3] have been proposed and implemented that allow the short and long term loudness of passages of audio (400ms and 3 second windows respectively) to be quantified. The European Broadcasting Union (EBU) have recommended a loudness measurement standard R128 [7] which builds upon the BS.1770-2 standard [3] by adding both short and long term measurement values and a gating mechanism. It defines normalisation levels to try and reduce the problems encountered in differing loudness levels during broadcast. There are a number of manufacturers that produce metering tools that adhere to this standard [8][9], of which the Vis-LM-H from NuGen audio incorporates a loudness history graph.

Whilst loudness normalisation is useful for monitoring and normalising audio material, both short and long term loudness measurements show the user very little in terms of cross spectral dynamic activity. They are in general used to match program loudness in broadcast situations and are not employed in the music production process except perhaps in the instance of final mastering.

Metering tools that enable the detection of punch and clarity in a musical production do not, at present, exist.

1.2. Clarity & Punch in music production

Considering an acoustic space, it is possible to objectively determine the clarity and intelligibility achievable within the space as an alternative to the traditional RT60 measurement [10]. Measures such as C50 (early to late arriving sound ratio) and EDT (early decay time) can all be combined for this purpose however, if one considers a completed music production, the task becomes very difficult for two key reasons.

Firstly, clarity in a musical extract becomes somewhat subjective in context of the production. For example, an ensemble recording may be judged on clarity by considering the tonality, spaciousness and localisation of each individual instrument whilst a contemporary heavy rock recording could be judged on the clarity of vocal and/or drums drum sources and overall bass & guitar frequency interplay.

Secondly, due to the combination of spectral components made up from a number of sources, it's very difficult to utilise traditional acoustic measures to grade the extract. Blind source separation [12][13] could be useful in determining overall clarity of instrumentation contained within a production however, this remains a complex process and often the sources extracted contain residual artefacts. Their use in qualitative measures is therefore limited.

In order for listeners to detect individual notes, instrument timbre and rhythm, it's important that enough elements of the mix conveying this information are clearly audible.

Masking is a phenomena that often occurs during mixdown when harmonic components of one source mask that of another source. Masking can occur in the temporal domain in addition to the frequency domain.

Considering the spectral nature of the individual sound sources during mixdown, it is possible to determine the level of masking taking place and modify the relative balance between sources to minimise this [11]. In a musical context, this anti-masking process will allow the listener to clearly hear the individual sound sources and thus, the overall production could be deemed to have higher clarity.

Due to the varying nature of audio, and moreover the harmonic content within each sound source, this process is not without its difficulties and the process of spectral balance is left to the skill of the engineer.

A reduction in dynamic range, as a result of applying maximisation/compression techniques, has the effect of raising the noise floor whilst at the same time increasing the level of spectral components contained within a source that were previously balanced in relation to their counterparts. Therefore, the process can cause additional masking to occur. This is perhaps one of the reasons lower subjective scores were given to audio samples that had been subjected to high compression levels in initial studies [17].

Whilst anti-masking plays an important role in determining the ability of 'tonality of sources' to be clearly defined in a music production, onset or transient detection is also key [13].

Within a musical context, temporal changes in frequency component amplitudes within the piece allow us to detect instrumentation [19]. A produced piece of music must contain various elements of information, which include instrumentation and transient content inorder to convey such things as emotion, and energy in the piece.

Dynamic range takes a leading role in allowing these elements to play their role in this process. If dynamic range is reduced, perhaps through excessive use of compression, important information in the piece is detrimentally affected, in particular transient information.

The ability of the listener to detect transients in a piece of music is fundamental to the determination of instrument type, note detection and rhythm. There are a number of automatic methods that have been proposed and evaluated [14][15] that attempt to detect onsets (and subsequent transients). These can include computation in the time, frequency and phase domains. The use of these techniques is often employed in beat extraction to determine rhythm, instrument identification and genre classification.

With respect to polyphonic music productions, where there could be a number of competing audio sources in the overall spectrum and thus overlapping attributes, some onsets events could be promoted as being more important than others [16] and/or wrongly identified. Thus, in order for highly accurate onset detection to take place, a complex algorithm is often required that utilises for example, particular frequency bands for analysis of different onset types.

Even in these cases, current onset detection algorithm performances vary when presented with near simultaneous events and in-distinct spectral signatures.

It has been shown [17] by the authors that a reduction in dynamic range on a piece of music does have an impact on the overall subjective qualitative score given by listeners. Given that transients within a production can be affected by a reduction in headroom it holds true that their associated measurement and detection, both by the listener and by objective measurement will be affected.

By monitoring the temporal changes in dynamic range across three key frequency bands, representing bass, mid and treble from a production viewpoint, it is proposed that the method may yield results that relate to the perception of clarity within a completed musical production.

The hypothesis is that without dynamic content between the frequency bands the listeners will grade the music as lacking clarity.

Punch, is a subjective term often used by engineers to describe a particular moment in a production where there is a degree of change in power in the music. In essence, productions that do not possess transient information cannot posses punch. Thus, punch is both related to transient change and the power spectral density at a particular moment in time and duration.

Further to the above hypothesis, dynamic change in particular frequency bands may contribute to the perception of punch indicated by the listener.

1.3. Inter-Band Relationship

The Inter-Band-Relationship Score (IBR) [17] represents the correlation between the dynamic range of material at particular sections in time. It is a multi-band measurement that looks at the correlation between low, mid and high bands in the audio signal under test.

It therefore could be useful in determining the relative dynamic content *across* frequency bands vs. time and by incorporating it into a metering tool, allow engineers to identify sections of audio that possess differing dynamic attributes.

To determine the IBR score, the audio excerpt is filtered using a 3 band linear phase FIR filter. Three filters were used and their respective cut-off frequencies and Q settings are shown as follows (Table 1).

Filter Type	Fc (Hz)	Fc (Hz)	Q
Low Pass LF	947	-	6.5
Band Pass MF	947	3186	6.5
High Pass HF	-	3186	6.5

Table 1. Filter Corner Frequencies

These frequencies were chosen as they approximate the 1st, 2nd and 3rd set of 8 critical bands in the auditory system.

Following this filtering process, dynamic range analysis was performed. Calculation of the dynamic ranges was derived from the n samples x_i in each band as follows:

$$Srms = \sqrt{\left(\frac{1}{n-1}\sum_{i-1}^{n}\left(x_i-\bar{x}\right)^2\right)} \qquad (1)$$

Where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2)$$

$$Spk = \max\left(x_{1..n}\right) \qquad (4)$$

$$Dr = 20 * \log(Spk / Srms) \qquad (5)$$

The inter-band relationship is derived as follows:

$$IBR = \sqrt{\left(\frac{1}{n-1}\sum_{i-1}^{n}\left(Dr_i-\overline{Dr}\right)^2\right)} \qquad (6)$$

A '75% Overall IBR' was calculated by taking the average of two window size measurements, such that;

Neither IBR above threshold then the Overall Grade = 0

One of the windowed IBR scores exceeds the threshold then the Overall Grade = 0.5

Both Windowed IBR scores exceed threshold then the Overall Grade = 1

Finally an objective profile was produced which represented a moving window average (75% overlap) based on the two IBR averages.

The IBR is a correlation score, such that strong correlation in dynamic range across the bands yields a low IBR score and vice verse. When calculating the IBR score it's also important to consider the dynamic range measurement itself at the time frame in which the IBR is measured. In the case where a high dynamic range is detected across all bands and these are strongly correlated, this may indicate a fast transient within the window size time frame and/or a significant change in loudness level within that time frame. One must state 'time frame' as a point in time cannot be considered to have a dynamic range.

Previous studies identified a correlation between the IBR and the subjective scores given by subjects with respect to overall quality [17]. In these studies excerpts tested were 7 seconds in length and the IBR score for each excerpt was calculated using a 7 second window size.

If one considers the variable nature of musical content over time, it is highly likely that the IBR measurements would vary as different window sizes are utilised and at different points in the audio under test. As such, this paper will extend the previously undertaken work and profile two songs utilising variable window sizes to calculate the IBR. These profiles will be compared to subjective listening test results where the listeners were asked to identify and map the audio with respect to their perceived punch and clarity. The IBR plots will also be compared to EBU R-128 based loudness measurements obtained using utilising a NuGen Vis-LM-H loudness meter [9].

The aim of this paper is to identify trends and relationships between the dynamics contained within a musical signal, its temporal loudness measure and how this relates to the perception of clarity and punch.

2. DESCRIPTION OF TESTING

2.1. Subjective Testing

A listening test was conducted comprising of 8 expert listeners. Each listener was asked to listen to the 6 excerpts and grade them along a time axis with a 400ms resolution. This resolution relates to the short term window integration time defined in the EBU R-128 standard. Resolutions less than 400ms were deemed impractical with respect to how the listeners would enter their grades.

Part of this study into punch and clarity involves allowing the users to grade the audio according to their own perception. However, in order to ensure some consistency in grading to the tests and allow the data to be collated and averaged, the listeners were given a training sheet which detailed the attributes that were to be assessed within the audio. The points detailed on the training sheet were as follows:

> *The audio is clear and punchy - Give a score of 1*
>
> *This can be defined if you can hear a clear vocal that doesn't suffer heavily from masking, clear dynamics are evident, clear drum hits/transients, bass notes, a point whereby dynamic movement is clearly audible.*
>
> *The audio is unfocussed, lacks punch and clarity - Give a score of 0*
>
> *This may consist of a large collection of harmonics or unrelated frequency components, noise, there is evidence of masking, no single element is clear, no dynamics present, distinct lack of transient content.*

The listening test took place in a professional control room environment, commonly found in music studios and the excerpts were auditioned on Genelec 8040 speakers at an average listening level of 74db(A).

The results of the listening tests were collated as an average punch/clarity score (P/C Average) and a profile plot was created which represented the perceived punch and clarity of each excerpt.

2.2. Objective Testing

Matlab was used to calculate the IBR using window sizes of 400ms, and 3s respectively. An initial IBR 'threshold' was chosen based on initial studies [17][18] and IBR scores attained in the 'best' scoring excerpts, in this case 4 or more. This value was based on the average IBR score attained for the two highest scoring excerpts in previous tests. The threshold determines the point at which the IBR score is deemed large enough to relate to a significant change in dynamics.

Where there was a significantly low score for the IBR, this threshold was adjusted to accommodate the reduced dynamics within an excerpt.

The IBR values for each window size were measured against time. In addition, measurements were taken which detailed the short and long term loudness and loudness variation against time. The objective IBR scores were then compared to their subjective counterparts.

2.3. Stimuli

2 different audio excerpts were chosen

- Excerpt 1 – Sugababes – Freak Like Me
- Excerpt 2 – Nickelback - Animals

The excerpts were 16bit, 44.1Khz, stereo WAV format.

The reason for choice was to allow for a varied test set based on the initial test carried out in [17]. In that study the Sugababes excerpt was considered to be the worst overall and the Nickelback excerpt as one of the best both subjectively and objectively with an IBR score calculated using a 7 second window size.

The songs were broken down into 3 20 second excerpts which represent key sections in the production: introduction, verse with drums and breakdown. In order to familiarise the reader with the productions the excerpts are now described.

2.3.1. Sugababes Excerpts

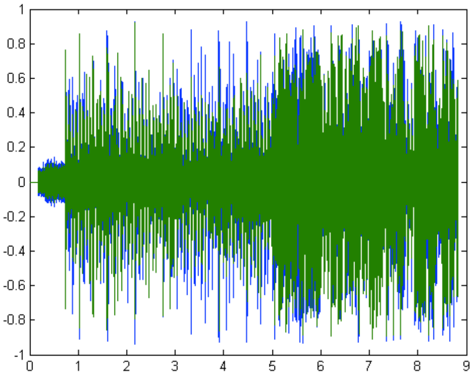

Figure 1. Sugababes Introduction

The Sugababes introduction (Figure 1) commences with a sound effect from the video game 'Frogger'. At approx 1.6 seconds the main Tubeway Army "Are 'Friends' Electric?" synth hook is introduced along with a low pass filtered drum track. This lasts for approximately 11 seconds and during this time the section is heavily processed with a form of sample rate reduction effect (Lo-Fi) and flanging. In addition the main synth hook is panned between left and right channels. At approx 11 seconds the main vocal introduces the first verse with no change to the music backing. Upon examination with a spectrogram, it was observed that up to the 1.6 second point there are no frequency components above 8kHz.

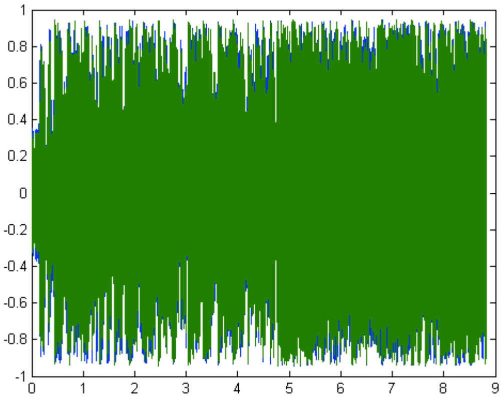

Figure 2. Sugababes Verse With Drums

This section of the song (Figure 2) begins with a small siren effect without drums and backing. The main synth, drums and bass then begin along with the main vocal verse. The verse continues until the 12 second point, at which time a heavy guitar riff is introduced, along with an additional lead synth and the chorus begins.

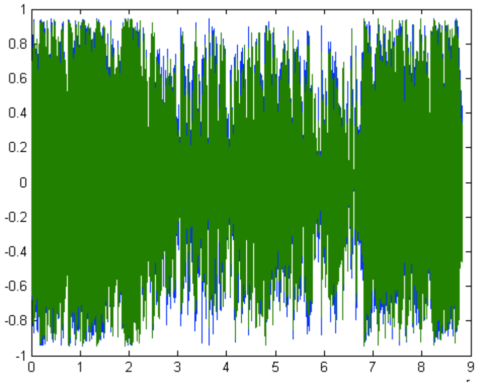

Figure 3. Sugababes Breakdown

The breakdown section (figure 3) begins with the end of the chorus before and at the 6 second point the song drops to the basic vocal and effected backing present in the intro section. This breakdown section continues before the song comes back in with a section identical in construct to that of the verse with drums first 12 second section. This occurs at the 16.5 second point, where there is a significant audio cut out.

2.3.2. Nickelback Excerpts

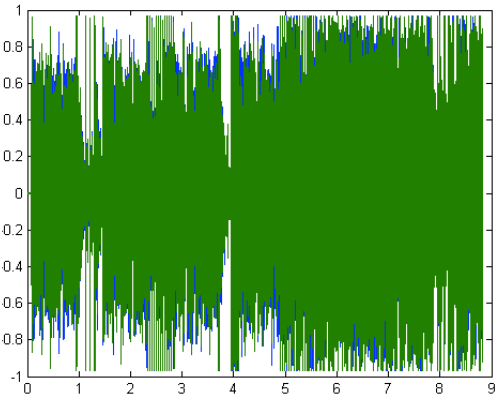

Figure 4. Nickelback Intro

This section (figure 4) opens with a heavy rhythm guitar 4 chord sequence, with strong low-mid frequency components. At the 2 second point a short drum fill occurs lasting until 3.2 seconds. The guitar riff continues including underlying hi-hat quarter note hits. At 8.8 seconds a significant tom fill occurs which includes a brief audio dropout. At 12 seconds a major drum fill occurs and the guitars, bass and drums play the main hook from the 14 second point. At 17.7 seconds a short drum fill / guitar drop out occurs before the main hook continues.

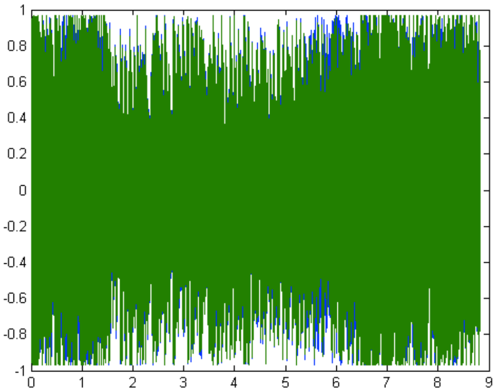

Figure 5. Nickelback Breakdown

This section (figure 5) opens with a heavy rhythm guitar 4 chord sequence, with strong low-mid frequency components. At the 3.5 second point the production drops the guitars out of the mix and features the vocal and hi-hats as a breakdown, a drum fill occurs at 15 seconds, followed by the full drums, bass, vocal and guitar mix at the 17 second point.

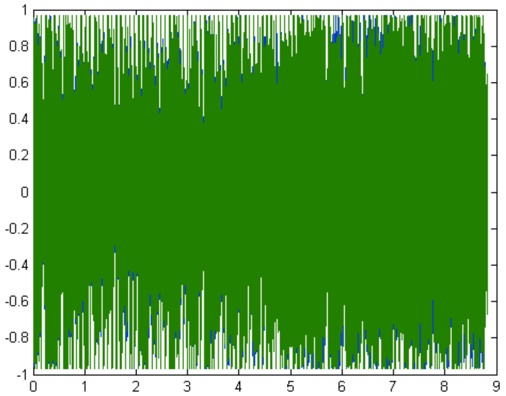

Figure 6. Nickelback Verse With Drums

Figure 6 shows the Verse With Drums section of audio. This section of audio represents the section of the track that contains bass, drums and vocal up to the 16 second point, at which the guitars are re-introduced.

3. RESULTS

The following charts (figures 7-12) compare the punch/clarity average scores with the 75% overlap IBR scores for each excerpt.

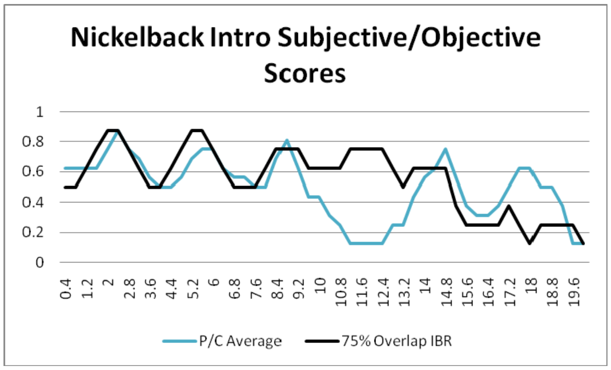

Figure 7. Nickelback Intro – Trend map

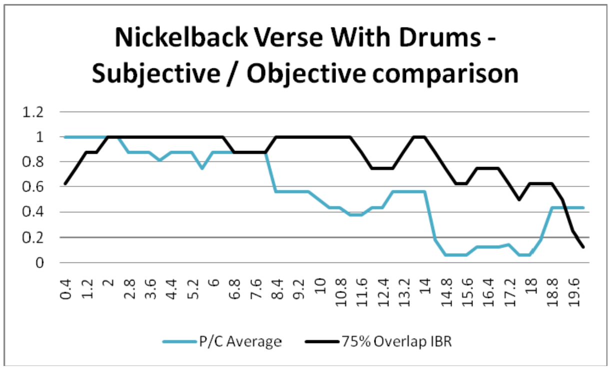

Figure 8. Nickelback Verse With Drums – Trend map

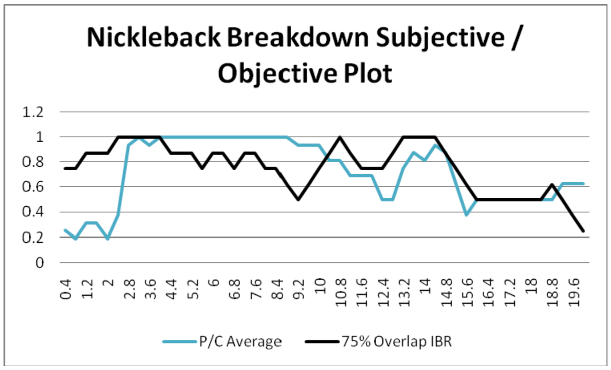

Figure 9. Nickelback Breakdown – Trend map

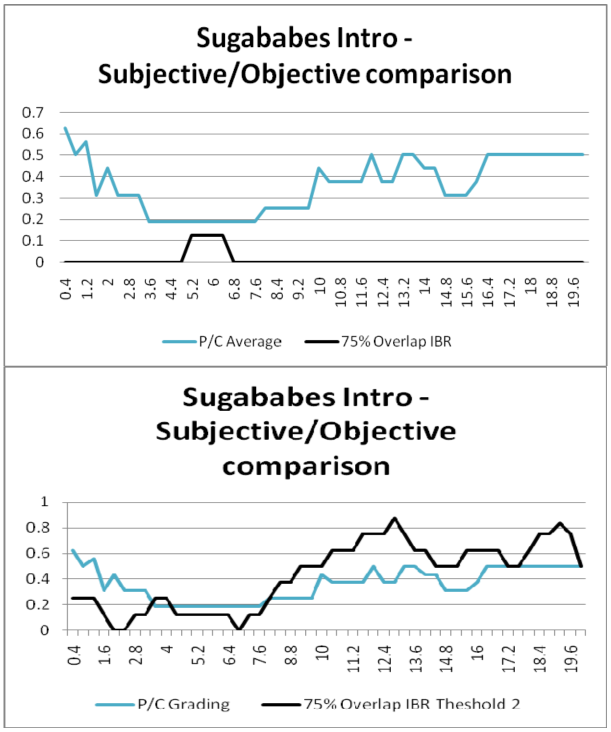

Figure 10a/10b. Sugababes Intro – Trend map

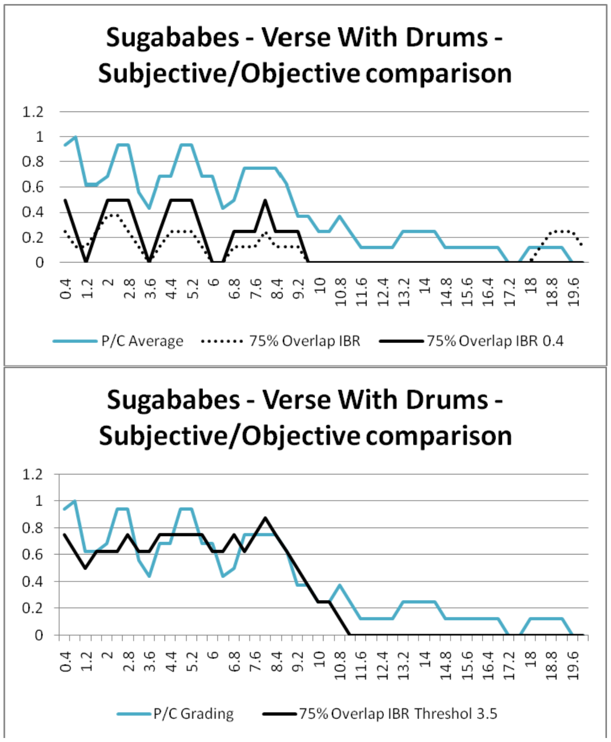

Figure 11a/11b. Sugababes Verse With Drums – Trend map

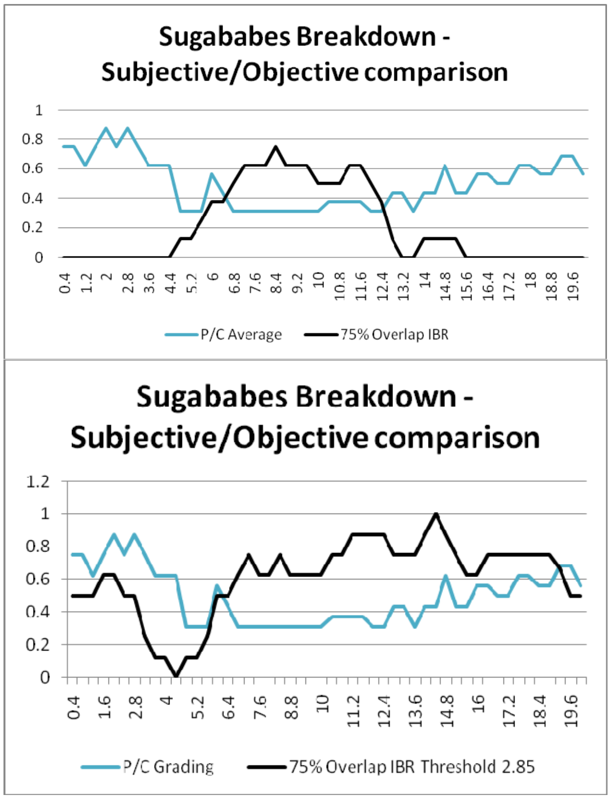

Figure 12a/12b. Sugababes Breakdown – Score comparison and error vs. time.

The following table represents the Pearson coefficent calculations for each of the excerpts.

Excerpt	Pearson Coefficient
Sugababes – Intro	Fig 10a 0.118 Fig 10b 0.354
Sugababes – VerseWDrums	Fig 11a 0.605 Fig 11b 0.917
Sugababes – Breakdown	Fig 12a -0.684 Fig 12b -0.240
Nickelback – Intro	0.330
Nickelback - VerseWDrums	0.485
Nickelback - Breakdown	0.334

Table 2. Pearson Correlation per excerpt

4. DISCUSSION OF RESULTS

As can be seen in the results, there is some degree of correlation between the overlapped IBR scores and the subjective scores given during testing. However, the correlation varies across different sections of the audio under test. What follows is a discussion and analysis based on the best and worst correlation scores measured, that of Nickelback Intro and Sugababes Breakdown excerpts respectively.

With reference to Figure 7 and Table 2, one can observe a very strong correlation between the subjective test scores and the IBR for the Nickelback Intro excerpt. The excerpt has an overall Pearson coefficient of 0.330. Visual inspection of Figure 7 shows , a high correlation between the IBR and subjective scores up to approx 8.8 seconds at which point the two trends deviate. This error margin begins to decrease around the 13 second point of the audio.

During the initial 8.8 second period of this excerpt the elements that are prominent are those of the drums and guitars at different times in a call and response pattern. Major drum fills occur centered around the 2, 5.2 and 8.4 second mark. These fills relate to the points at which the listeners have graded the audio as punchy and clear. These points also correlate well with the objective IBR score. Further points where the error is minimised between the two measures are at the 12 second point, again a point at which a major drum fill occurs

Significant errors begin to occur at 8.8 seconds where there is a major tom fill and an audio drop out. This could explain the de-correlation in subjective and objective scores within this period. The tom fill and audio drop out is seen by the algorithm as a highly transient event and therefore a high IBR score is calculated whilst the loss of audio could be considered by the listeners as lacking both as punch and clarity. The loudness plot at this point (see Figure 21) shows a loudness range increase of approximately 14dB, this lasts for around 3 seconds which relates to the point at which the error margin begins to decrease. This period of de-correlation is caused by a period of low level audio and this issue may be overcome by making use of a gate in the same way as used in EBU R-128.

Whilst the IBR score indicates transient behaviour is occurring during the 8.8-13.2 second period, this is graded with a low score by the listeners. The IBR 400ms plot for this excerpt (Figure 13) shows the 400ms IBR score falling below the threshold value during this time period. The calculated IBR compared to the subjective measures is based on a moving average (75% overlap) window and also incorporates the 3 second IBR score. This suggest that in order to increase the accuracy of the IBR score, one might consider a calculation based solely on the 400ms windowed IBR.

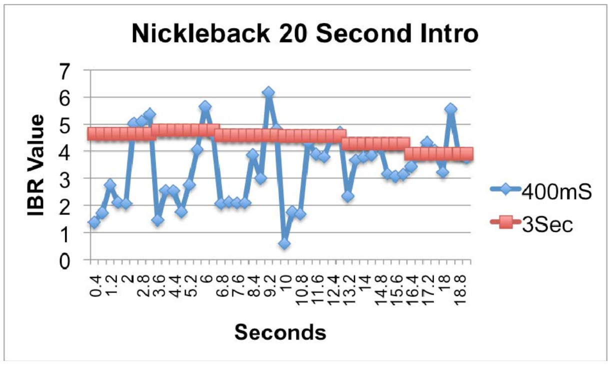

Figure 13. Nickleback Intro IBR Scores for 400ms and 3 Second Window Sizes

With reference to figure 11a, this technique was applied to the Sugababes Verse With Drum excerpt as can be seen, the trends begin to map more closely.

Overall, when comparing all six of the Nickelback and Sugababes excerpts, it was noted that despite the Nickelback exceprts having the lowest overall loudness ranges (Figures 21, 23 & 25), they possessed on average, higher instances of dynamic content, this is shown by the IBR scores that exceed the threshold limit of ‘4’.

The Nickelback excerpts regularly exceed the threshold whilst the Sugababes excerpts don’t. If one examines the associated loudness plots (Figures 21, 23, 25, 27, 29 & 31) both excerpts are well in excess of the proposed -23 LUFS loudness level. Of interest, is that the Nickleback samples indicate a smaller average loudness variation but still posses enough dynamic fluctuation to assert a high IBR score, which suggests frequency band de-correlation.

The Sugababes excerpts, whilst not having the reduced loudness range of the Nickelback tracks do not exhibit the same frequency band de-correlation, hence the lower average IBR scores obtained.

The nature of the subjective test was non-comparative i.e the listeners were asked to grade each excerpt separately. Therefore, it’s likely that the listeners were grading ‘punch and clarity’ by comparing points in time of the audio they were listening to for that particular grading phase. The original threshold grading of 4 was perhaps wrongly chosen as it represents a comparative value which originally distinguished between good and bad quality excerpts[18].

Table 2 shows that with a Pearson coefficient of -0.684, the Sugababes Breakdown excerpt was the excerpt which showed the worst correlation. With reference to Figure 12a, the error margin decreases between the 6 and 16 second point. This in fact corresponds to the point at which the vocal line is prominent in the mix. This equates to a high IBR score both in the 3s and 400ms window time frame due to varying dynamics in the mix i.e de-correlation in the low, mid and high frequency bands.

The listeners appear to grade the excerpt differently, giving higher scores to the sections of audio outside this region. This could be due to them gauging drums, bass and synth as punchy and a lone vocal as neither punchy or clear. A point to raise here is that whilst there is clearly dynamic content within the piece signified by the IBR scores, there is both a drop in loudness level by approx 13.5dB which could also correspond to the listener perception of punch and clarity. This might suggest that a combination of loudness and dynamics measure could improve the accuracy of the overall punch and clarity score. For example, weighting of the IBR with respect to overall loudness could improve the accuracy

By varying the threshold used in the IBR calculation, it is possible to profile the excerpts but with a greater *dynamic sensitivity.* Figures 10b, 11b and 12b show an IBR profile that has been calculated using threshold values of 3, 3.5 and 2.85 respectively. Looking at table 2, the Pearson coefficients indicate that this change in sensitivity improves the trend correlation. The Sugababes – Verse With Drums excerpt shows an improvement from 0.605 to 0.917 which could be considered highly correlated.

5. CONCLUSIONS

This paper presents and begins to quantify the performance of an objective measure that can be used to assess audio quality with respect to punch and clarity.

The results show that despite a musical piece having a smaller loudness range, it is the transient content and dynamic range de-correlation between frequency bands that relate to higher subjective scores given by the listeners. The excerpts tested were from the rock and pop genre. Additional testing is therefore required to measure the effectiveness of the IBR measure within other genres.

Further work is required to define clearly the distinction between punch and clarity in a musical context. Despite the relative success in the mapping of the IBR score to the subjective trends, with the benefit of hindsight the authors should have asked the listeners to score the excerpts based on a single measure, i.e. either punch or clarity.

Further work could investigate the inclusion of a gating system and an adaptive dynamics threshold which might lead to a stronger correlation between the objective measures and subjective scores.

Due to the simplicity of the algorithm, it lends itself to real-time implementation and therefore can be exploited within mixing, mastering and monitoring tools.

6. APPENDIX

6.1. Subjective Results

The following graphs show the individual subject scores and the average punch/clarity scores (P/C) received with respect to time for each sample.

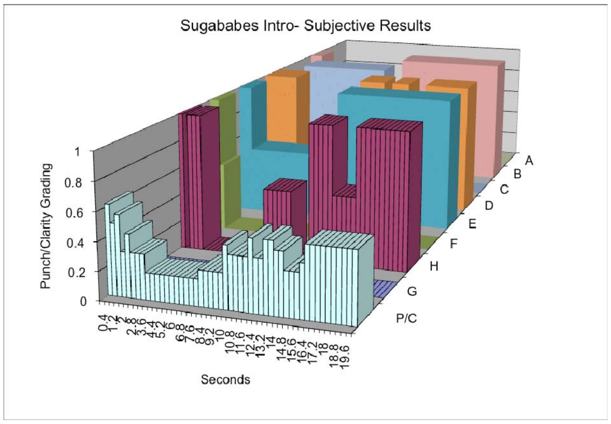

Figure 14. Sugababes Intro Subjective Scores

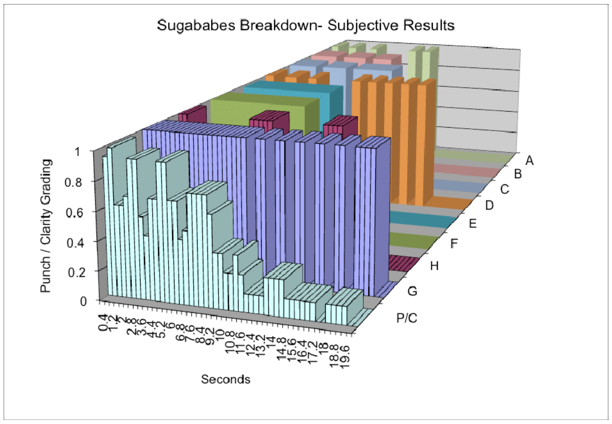

Figure 15. Sugababes Breakdown Subjective Scores

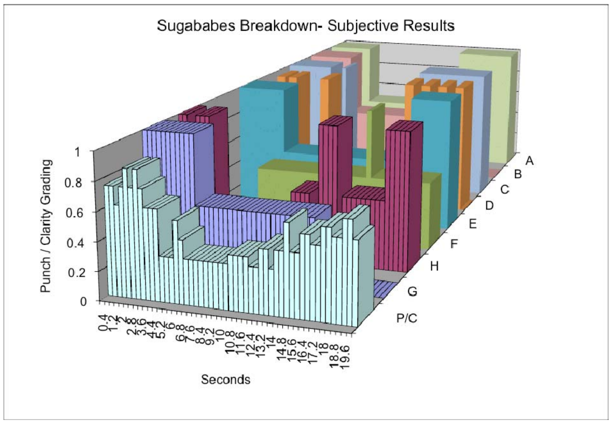

Figure 16. Sugababes Breakdown Subjective Scores

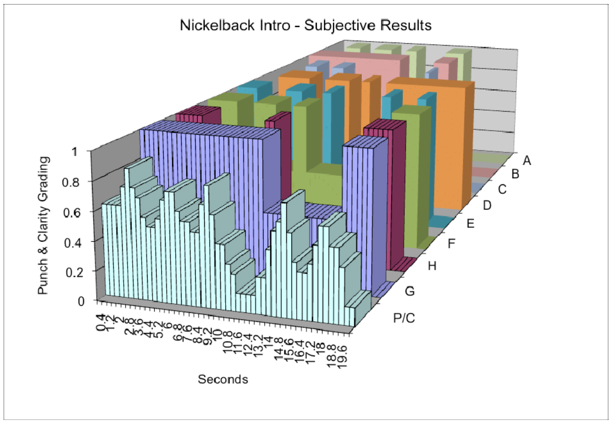

Figure 17. Nickelback Intro Subjective Scores

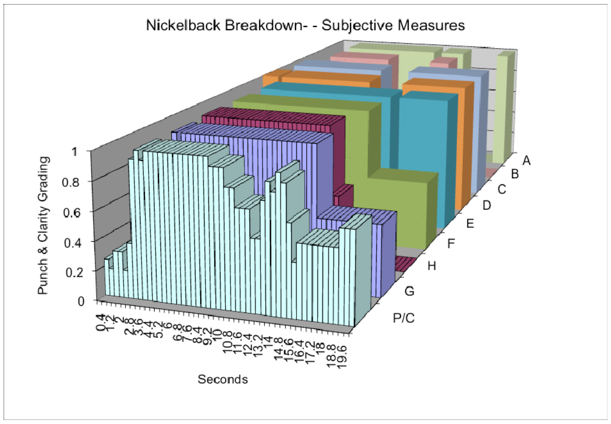

Figure 18. Nickelback Breakdown Subjective Scores

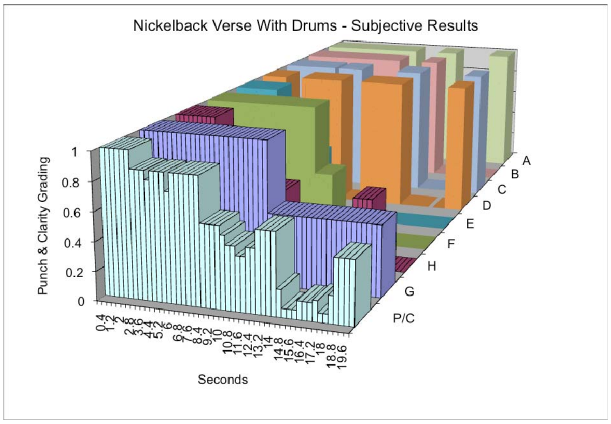

Figure 19. Nickelback Verse With Drums Subjective Scores

6.2. Objective Results

The following graphs were plotted representing the points in time where the calculated IBR score exceeded the threshold of 4. The graphs show the value based on 3 second and 400ms window sizes and the average of them both. In addition the loudness level plots of each are shown.

6.3. Nickelback – Animals

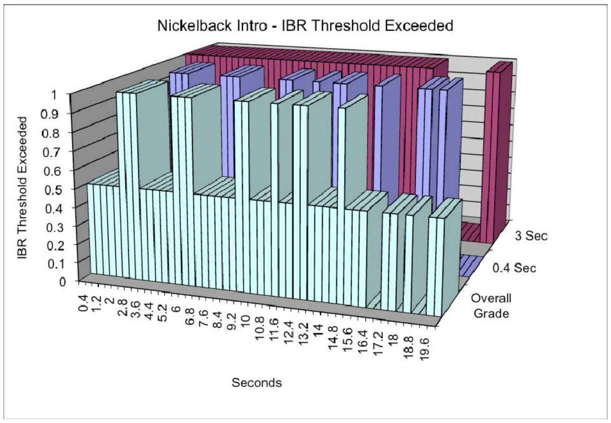

Figure 20. Nickelback Intro IBR threshold Scores

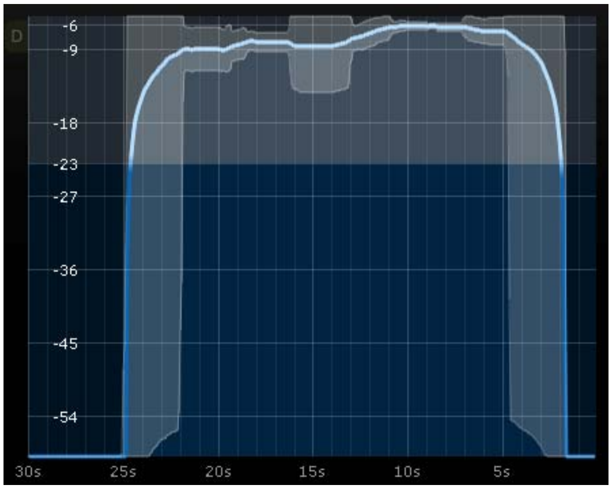

Figure 21. Nickelback Intro LUFS Loudness

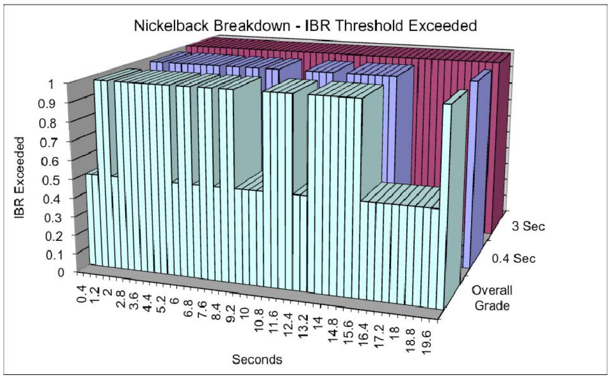

Figure 22. Nickelback Breakdown IBR threshold Scores

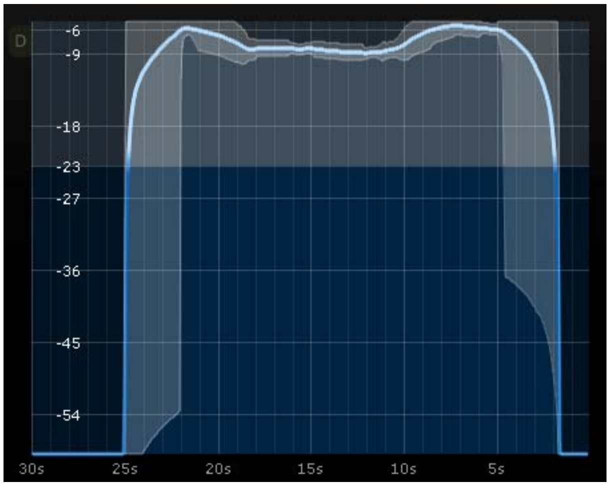

Figure 23. Nickelback Breakdown LUFS Loudness

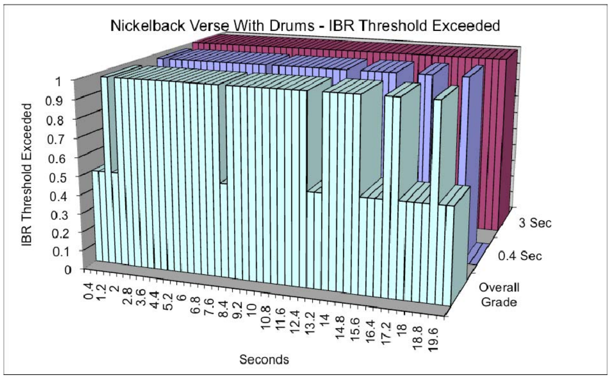

Figure 24. Nickelback Verse With Drums IBR threshold Scores

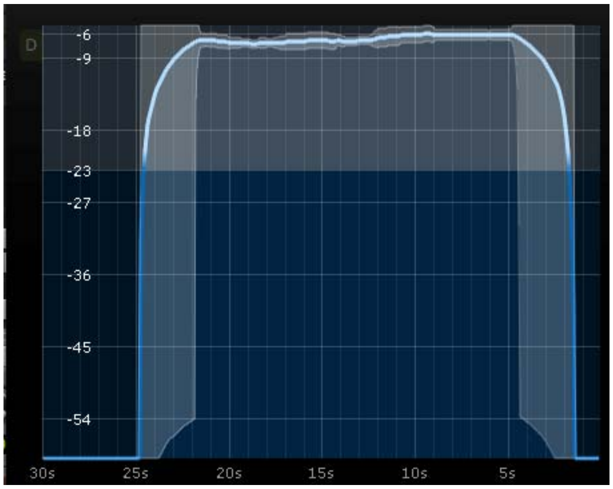

Figure 25. Nickelback Verse With Drums LUFS Loudness

6.4. Sugababes – Freak Like Me

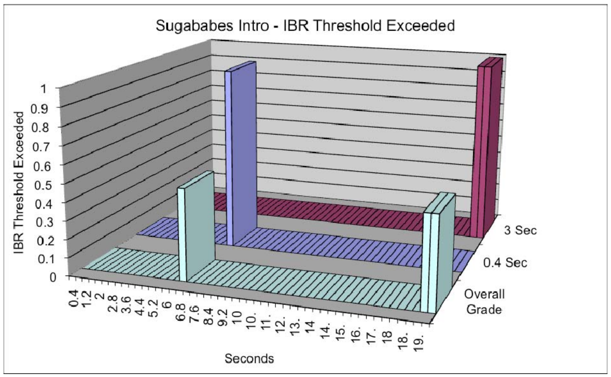

Figure 26. Sugababes Intro IBR threshold Scores

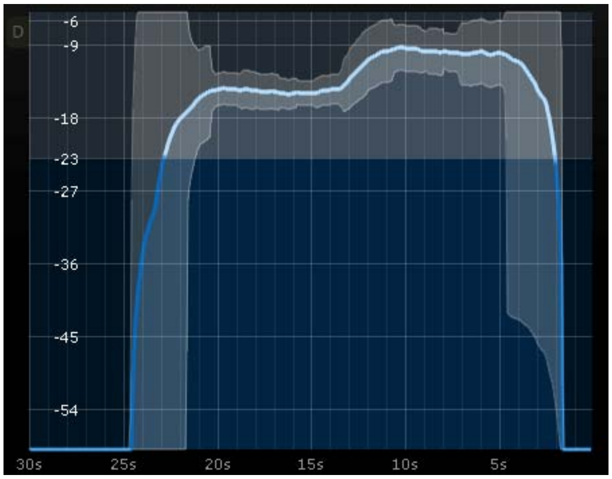

Figure 27. Sugababes Intro LUFS Loudness

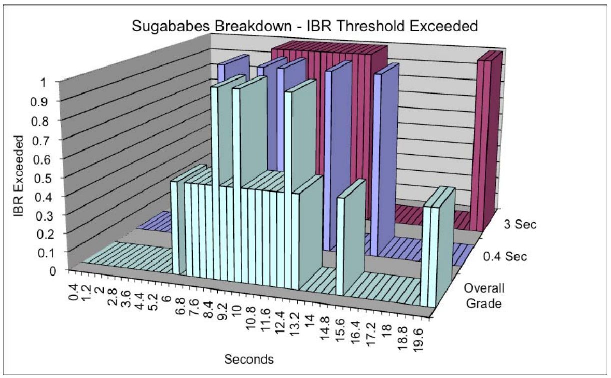

Figure 28. Sugababes Breakdown IBR threshold Scores

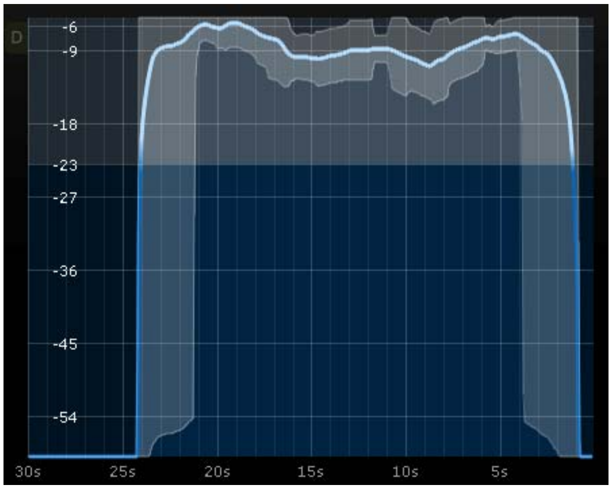

Figure 29. Sugababes Breakdown LUFS Loudness

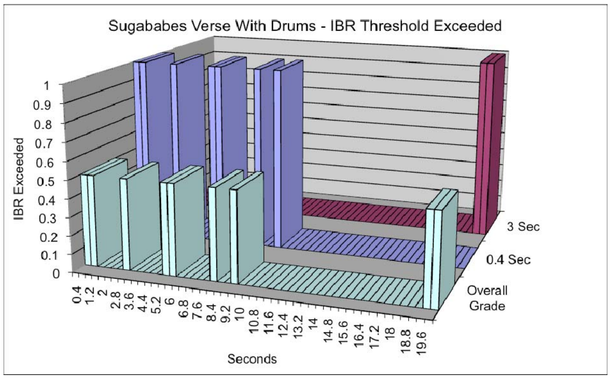

Figure 30. Sugababes Verse With Drums IBR threshold Scores

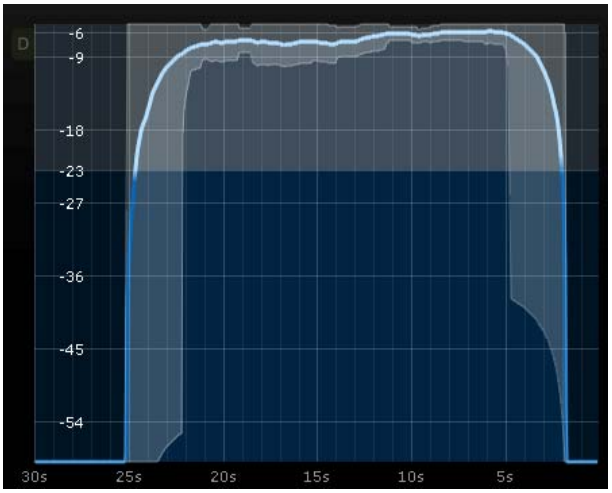

Figure 31. Sugababes Verse With Drums LUFS Loudness

7. REFERENCES

[1] E.Skovenborg and T.Lund, Loudness descriptors to characterize programs and music tracks, AES Convention Paper 7514, October 2008

[2] E.Skovenborg and S.H.Nielsen, Evaluation of Different Loudness Models with Music and Speech Material, AES 117th Convention, San Francisco, CA, USA, October 2004

[3] ITU-R BS.1770, Algorithms to measure audio programme loudness and true-peak audio level, International Telecommunications Union, Geneva, Switzerland, 2006

[4] Voxengo Gliss EQ, http://www.voxengo.com/product/glisseq/ [accessed 3rd January, 2012]

[5] Fab Filter, Pro Q Equalizer, http://www.fabfilter.com/products/pro-q.php [accessed 8th January 2012]

[6] Pleasurize sound meter, http://www.dynamicrange.de [accessed 20th November 2011]

[7] EBU-R128, Loudness normalisation and permitted maximum level of audio signals, EBU PLoud Group, August 2011

[8] AC-R128 Loudness meter, https://www.audiocation.de/en/plugin [accessed 8th January 2012]

[9] Vis-LM-H, NuGen Audio Loudness Meter with history, http://www.nugenaudio.com/visLM_loudness-meter_VST_AU_RTAS.php [accessed 2nd February, 2012]

[10] G.Ballou “Handbook for sound engineers”, 3rd Edition, Gulf Professional Publishing, April 2005

[11] E.Gonzalez and J. D. Reiss, "Automatic equalization of multi-channel audio using cross-adaptive methods", Proceedings of the 127th AES Convention, New York, October 2009

[12] D.Barry, D.Fitzgerald, E.Coyle and B. Lawlor, Single Channel Source Separation using Short-time Independent Component Analysis, AES 119th conference, October 2005.

[13] M.Every, Discriminating Between Pitched Sources in Music Audio, IEEE Transactions, Feb 2008

[14] S.Hainsworth and M.Macleod. Onset detection in musical audio signals. In Proc. Int. Computer Music Conference, pages 163–6, 2003

[15] JP Bello, L Daudet, S Abdallah, C Duxbury, M Davies, M Sandler, A Tutorial on Onset Detection in Music Signals. IEEE Transactions on Speech and Audio Processing. 13, 1035–1047, Sept. 2005

[16] N.Collins, A Comparison of Sound Onset Detection Algorithms with Emphasis on Psychoacoustically Motivated Detection Functions, AES Convention Paper Presented at the 118th Convention 2005 May 28–31 Barcelona, Spain

[17] Fenton, S., Fazenda, B. and Wakefield, J.‘Objective quality measurement of audio using multiband dynamic range analysis’. Institute of Acoustics (IOA) Conference-November 2009

[18] Fenton, S., Wakefield, J. and Fazenda, B. ‘Objective Measurement of Music Quality Using Inter-Band Relationship Analysis’.AES 130th Conference, London, UK., 13th-16th May 2011.

[19] M.Lagrange, M.Raspaud, R.Badeau, and G.Richard, Explicit Modeling of Temporal Dynamics within Musical Signals for Acoustical Unit Similarity, June 2009