Fenton, Steven, Wakefield, Jonathan P. and Fazenda, Bruno

Objective Measurement of Music Quality Using Inter-Band Relationship Analysis

**Original Citation**

Fenton, Steven, Wakefield, Jonathan P. and Fazenda, Bruno (2011) Objective Measurement of Music Quality Using Inter-Band Relationship Analysis. In: AES 130th Conference, 13th-16th May, Hosted in London, UK., 13-16 May 2011, London.

This version is available at http://eprints.hud.ac.uk/9851/

Audio Engineering Society

# Convention Paper

Presented at the 130th Convention
2011 May 13–16 London, UK



# Objective Measurement of Music Quality Using Inter-Band Relationship Analysis

Steven Fenton[1], Bruno Fazenda[2] and Jonathan Wakefield[3]

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

[2] Acoustics Research Centre, University of Salford, Manchester, UK
b.fazenda@salford.ac.uk

[3] 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 grades the quality of a complex musical signal. 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 testing with real-world music excerpts and a greater number of listening subjects. A high degree of correlation is observed between the Mean Subject Scores (MSS) and the objective IBR descriptor suggesting it could be used as an additional model output variable (MOV) to describe produced music quality. The method lends itself to real-time implementation and therefore can be exploited within mixing, mastering and monitoring tools.

## 1. INTRODUCTION

Audio can have many purposes. In the context of this paper, it is musical performance captured by a recording process (or a programmed sequence) and stored on a medium for later listening and enjoyment. Ever since the very first recordings were made [1] engineers have strived to improve the quality of the recording and playback process.

When considering an audio system, its performance can be measured using different metrics [21]. Generally the measurement of the input signal is compared to that of the output signal and measurements such as Total Harmonic Distortion (THD), Signal to Noise Ratio and frequency response are used to describe the overall system performance.

Whilst this methodology for testing is sufficient, even as a black box approach, to describe individual components and subsequent performance of a 'system', it does not allow the quantification of audio quality if the same techniques are applied to produced music.

Music is a complex signal made up of many different harmonic components of varying phases and magnitudes. Elements of music include pitch, rhythm and the sonic qualities of timbre and texture. In produced music mixes are produced, by the engineer, from individual signal sources that are, generally themselves complex. The complexity of the individual signal sources, often referred to as timbre, is what distinguishes them as different types of voices or musical instruments.

Overall quality can therefore, not be attributed to a single metric or measurement, in fact, many of the metrics relating to audio system performance measurement are not applicable to the measurement of 'overall music quality'. THD for example is referenced to a pure tone at the system input. Whilst it is possible to perform conventional 'system' measurements on produced music to describe aspects of the audio under test, these do not correlate very well with the overall perception of quality by the listener [23].

It is the skill of the engineer in the production stages that often leads to a completed recording being deemed as 'clear', 'defined', 'punchy' or 'highly polished'. A badly engineered and produced recording could be referred to as 'woolly', 'distorted', 'poorly balanced' or 'muddy'.

These descriptors are of course subjective. However, they are frequently used and recognized within the audio industry and for the vast majority of engineers these descriptors are used to categorise the production of a piece of music. They do not, however, allow for a consistent qualitative measure to be established.

This paper is concerned with investigating the influences of dynamic range on the perception of audio quality in produced music. The paper investigates a novel multi-band technique, first proposed by the authors [6] in a pilot study conducted to obtain an objective measure that can be used, in conjunction with other extracted objective measures, to describe one aspect of the basic audio quality (BAQ) of recordings under test.

## 2. BACKGROUND

### 2.1. Subjective & Objective Measures

Formal listening tests are regarded as being the most reliable method for audio quality assessment and a number of methodologies have been established [2]. The proliferation of such tests have in the most part been in response to a need to evaluate the quality of low bit rate CODECS [3][4] due to the wide use of voice over internet, streaming technologies and the dominance of the MP3 format for music distribution.

Three major recommendations with regards to the subjective assessment of audio quality have been established. These are standardized as ITU-R BS.1116, developed primarily to evaluate small impairments in audio quality [5], ITU-R BS.1534-1, commonly referred to as MUSHRA, developed to evaluate intermediate impairments in audio quality [6] and ITU-T P.800, primarily used to evaluate narrowband speech quality [7].

Generally, these testing and measurement techniques are employed to establish audio quality in audio systems (such as CODECs) under test with respect to an original 'untreated' reference signal. The resulting index is named the subjective difference grade (SDG) which attempts to categorize the subjective audio quality. These types of test can be very time consuming and subject to errors through various forms of biasing [8] some of which will be described later. This makes them impractical for evaluating produced music mixes on an ongoing basis.

In order to address the need for automatic quality measurement of audio, a number of objective measures have been proposed. These attempt to predict the BAQ from extracted features of the audio under test. Many of the techniques have been standardized as ITU-R BS.1387-1, otherwise known as PEAQ (Perceptual Evaluation of Audio Quality) [9].

PEAQ combines many different model output variables (MOVs) in order to compute the objective difference grade (ODG). The basic version of PEAQ combines 12 of the MOV's to calculate the ODG whilst the advanced version combines a further 5.

All of the tests, subjective and objective, are full-reference quality indexed, i.e. they compare the audio under test with respect to an original reference signal (uncompressed/unprocessed). Whilst these tests can be used to measure and quantify the BAQ of a piece of audio that has been processed using a codec, they cannot be used to measure and describe the quality of a 'produced' piece of music that has no reference.

### 2.2. Loudness

Many proposals and studies relating to loudness are documented including its measurement; one such standard for measurement is detailed in ITU-R BS.1770 [10]. This loudness model has been extended with further descriptors to allow the effective measurement over time [11]. There have been a number of studies relating to the perception of loudness [11][12][23][24] and their relationship to temporally varying sounds. [24]. Whilst each study attempts to measure and quantify the effectiveness of algorithms relating to listener perception of the sensation of loudness and in some cases dynamic range, they do not relate the measure to the perception of quality.

Loudness, it seems, appears to dominate modern music production. Due mainly to the record label's need to be the loudest on radio, but also because our perception of the production quality appears to be majorly influenced by the demand of the artist for their material to match that of their peers. During loudness maximisation, material is compressed, resulting in a reduced peak to R.M.S level ratio and thus an overall reduction in dynamic range. This peak-level based processing makes material perceptually louder.

Since the mid-1980's, a trend has developed in music production that has resulted in the loudness of completed productions being increased during the mastering process [13]. This increase in loudness has seen the gradual reduction in dynamic range of produced music.

Whilst this continued decrease in dynamic range occurs, it is accepted amongst audio professionals that this is potentially detrimental to the overall audio quality of the music. The push for ever louder recordings has led to the 'loudness wars' [13] and also, in contrast, to movements such as 'Turn Me Up' to promote the opposite [14].

As discussed, development of automatic measurement of audio relating to loudness has been undertaken however, work on measures based on dynamic range and it's correlation with the perception of listener quality need to be extended.

It is the authors' belief that whilst the two measures, loudness and dynamic range are inextricably linked, the latter should be considered as an separate measure when attempting to quantify the music quality.

### 2.3. Dynamic Range

The term dynamic range is often quoted in decibels (dB) when describing the performance of an audio system. The context of measurement is an important factor to consider when the interpretation of the dB value is evaluated. The context can either be categorized as that of a system or signal.

In the context of a system the measurement is used to describe the maximum range that is permissible, before distortion takes place (clipping), measured from the noise floor to the peak level. The Audio Engineering Society specify this measurement as "20 times the logarithm of the ratio of the full-scale signal to the R.M.S. noise floor in the presence of signal, expressed in dB FS" [17]. This value gives an indication of the true headroom of a system and shouldn't be confused with SNR (Signal to Noise Ratio) which is often measured without the presence of a signal and can therefore give an inaccurate system measurement due to muting circuits.

When we describe the signal itself rather than the system under test, the dynamic range can be given as the ratio of the full-scale level of the signal to its lowest level. Given that audio signals under test are generally varying in level, particularly during fade ins-outs, interludes etc, an average level (R.M.S) is generally taken of a section of audio under test as being representative of the 'active' passage of music. This average level is then used to compute the dynamic range in conjunction with the peak level measured during the same passage.

This AES proposal of dynamic range measurement [17] is the one adopted within this paper. However, this method is extended by the inclusion of 3 band filter being applied to the signal under test. This method was explored previously in a pilot study by the authors [6].

Many loudness models implement a multi-band approach [12] to approximate the response of the human ear. The bands are often weighted and the overall loudness is calculated as an average over a period of time, either based on short or long term windows.

## 3. OBJECTIVE ANALYSIS

It is widely accepted that the response of the human ear and listener perception differs across the ears frequency range. It is therefore argued that a single wideband dynamic range figure would be inaccurate in describing the basic audio quality of a signal, although it could be used to represent an overall mean 'figure of merit' score.

The authors' previous proposal of a MOV based upon a multiband dynamic range measurement is now explored [6].

### 3.1. Multiband Analysis

Analysis is made of the dynamic range at three critical bands of hearing and the interaction of each against the combined Mean Subject Score (MSS) is examined. Each excerpt was 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.

| 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 and Q

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

Following this filtering process, dynamic range analysis was performed. Calculations of the dynamic range were derived from the samples in each band as follows:

$$Srms = \sqrt{\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)} \quad (1)$$

Where

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

$$Spk = \max\,(x_{1..n}) \quad (3)$$

$$Dr = 20\log\,(\frac{Spk}{Srms}) \quad (4)$$

The inter-band relationship (IBR) is derived as follows:

$$IBR = \sqrt{\left(\frac{1}{n-1}\sum_{i=1}^{n}(Dr_i - \overline{Dr})^2\right)} \quad (5)$$

Where $Dr_i$ represents the dynamic range calculated at band $i$. $\overline{Dr}$ represents the mean dynamic range.

Effectively, the IBR represents the correlation between the dynamic range existing across the three bands. A low value would represent very little variation in dynamic range across the bands, whereas a high value would represent a high degree of variation across the three bands under test.

It is proposed that music containing very little variation in dynamic range across frequency ranges would in fact result in a lower quality score when auditioned by the listeners. It could also signify a particularly 'dense' section of the arrangement in terms of frequency content.

## 4. LISTENING TEST

This study involved 57 experienced listeners listening to and grading 5 excerpts from commercially available produced music.

### 4.1. Method

A listening test was designed to measure the subjective preference of listeners grading produced music of varying genre. No knowledge of the engineering and production techniques involved with any excerpts was made available to the listeners.

The listeners were asked to listen to each excerpt with respect to analysing their relative punch, clarity, overall tone and balance.

Each listener was then asked to grade each of the excerpts out of 10 with respect to their overall production quality (1 being low quality, 10 being the highest quality).

The listening conditions and equipment varied between subjects however, this was deemed satisfactory for such a wider study. This would in fact, be how the music would be received and auditioned by the general population of listeners. Whilst the listening level between subjects was not maintained, the relative listening levels between excerpts could be assumed to be constant, thus the quality scores given by each of the subjects would be relative between excerpts.

The key point of this paper is to establish whether the objective measure proposed would correlate with the subjective scores given.

A handout was given to each subject also detailing the test and guidelines.

### 4.2. Stimuli

The five excerpts were from the following songs, each of which can be classed as contemporary productions;

Excerpt 1: Metallica – The End Of The Line
Excerpt 2: The Killers – Mr Brightside
Excerpt 3: Sugababes – Freak Like Me
Excerpt 4: Nickleback – Animals
Excerpt 5: Elbow – Seldom Seen Kid

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

The reason for the choice of excerpts was to allow for a varied test set, thus testing the perception of the dynamic range across a number of different stimuli, including transient and harmonically rich material.

## 5. RESULTS AND DISCUSSION

The subjective scores obtained during the test were averaged resulting in a MSS. From this, an order of preference (ranked best quality to worst) of the excerpts was extracted.

### 5.1. Listening Test Results

These subjective results are summarized as follows

Excerpt 5: 1st Place - Elbow – Seldom Seen Kid
Excerpt 4: 2nd Place - Nickleback – Animals
Excerpt 1: 3rd Place - Metallica – The End Of The Line
Excerpt 2: 4th Place - The Killers – Mr Brightside
Excerpt 3: 5th Place – Sugababes – Freak Like Me

The results of the test are shown in figure 1.

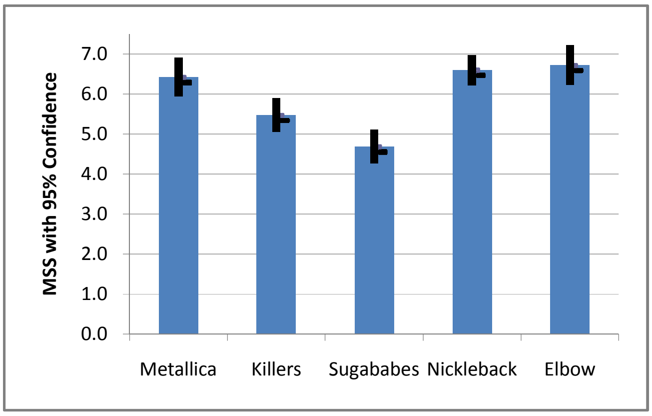


Figure 1. Mean Subject Scores

It should be noted that whilst an order of preference was extracted, the scores given for the top three placed excerpts were indeed very closely grouped. Indeed, if the 95% confidence intervals are examined (figure 1) one can see that there is a high degree of overlap between the scores obtained by the top three placed excerpts.

Reasons for this could be due to biasing factors in genre preference in addition to all excerpts being professionally produced and mastered, thus making their relative ranking scores group together.

At this stage we could consider relaxing the ranking and suggest that this test has extracted the top three, 2nd place and last place excerpts in terms of audio quality. Indeed, if we consider the 95% confidence intervals shown in figure 1, there is a clear differentiation between the tiers identified.

If one considers that the excerpts are all 'release' quality one would expect that all the excerpts would achieve a relatively high MSS. However, if one considers all five excerpts, there is clearly some differentiation between their perceived audio qualities.

In order to establish whether the variation of preference scores achieved were due to the excerpts under test or simply down to chance an analysis of variance was undertaken with the following null hypothesis:

$H_0$ = There is no difference in quality between excerpts.
$H_1$ = There is a difference in audio quality between the excerpts.

The results of an ANOVA test for this (See Table 2) show that the F value is highly significant and the resultant p-value obtained is <<0.05, suggesting that the differences found in the MSS across the samples is more than would be expected by chance alone.

The *F* value obtained is significantly larger than the *Fcrit* value therefore we can reject the null hypothesis with a large degree of confidence.

| *ANOVA* | | | | | | |
|---|---|---|---|---|---|---|
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 174.22456 14 | 4 | 43.556140 35 | 15.100 26935 | 3.43334E-11 | 2.40388 4714 |
| Within Groups | 807.64912 28 | 280 | 2.8844611 53 | | | |
| | | | | | | |
| Total | 981.87368 42 | 284 | | | | |

Table 2. ANOVA

## 5.2. Objective Results

The excerpts were each analyzed with respect to the IBR as detailed in section 3.1.

The results of the analysis are shown in figure 2.

The order of the excerpts, based on the largest to smallest IBR measurement, was extracted. The order can be summarized as follows;

Excerpt 5: 1st Place - Elbow – Seldom Seen Kid
Excerpt 1: 2nd Place - Metallica – The End Of The Line
Excerpt 4: 3rd Place - Nickleback – Animals
Excerpt 2: 4th Place - The Killers – Mr Brightside
Excerpt 3: 5th Place - Sugababes – Freak Like Me

A second measurement, based on calculating the IBR based only on the Low-Mid bands was extracted. This is shown in figure 2 as LMIBR.

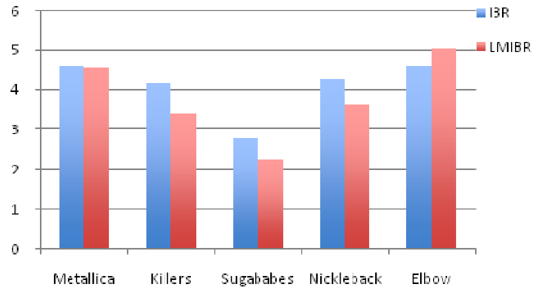


Figure 2. Inter-band Relationship

### 5.3. Discussion of Results

If the extracted subjective and objective order of preference is compared in figure 3, one can see that there is a large degree of correlation between the two.

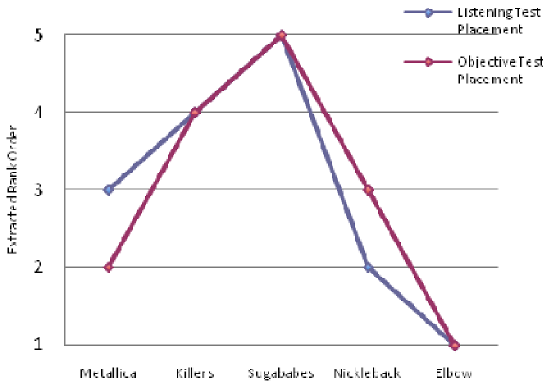


Figure 3. Subjective & Objective Order of Preference Comparison

Running a Spearman correlation test between the subjective and objectively extracted rank scores a correlation coefficient of 0.9 is obtained along with a significance value of 0.037, this is significant at the 0.05 level (2- tailed)

Elbow, Killers & Sugababes were all ranked identically with placements of 1st, 4th and 5th in order of quality. Thus, the objective score based on IBR successfully identified the excerpt which was ranked as being the best in the subjective test. It also identified the 2 excerpts that were graded as having the lowest quality subjectively by the listeners.

The top three excerpt ranking based on the IBR score and the MSS differed with Metallica and Nickleback being reversed in order. This was the only difference in ranking score.

Differences in the placement of Nickleback and Metallica shown in figure 3 could be attributed to biasing due to personal preference in genre. The excerpts themselves contained variation in transient content and tempo, again, these could be affecting the overall rankings given by the subjects. The order of playback of excerpts was left up to the subjects therefore allowing continuous and multiple comparisons to be made between each excerpt.

If one normalizes both the mean subjective scores and inter-band relationship score for each of the excerpts tested, the correlation between the two can be observed, see figure 4.

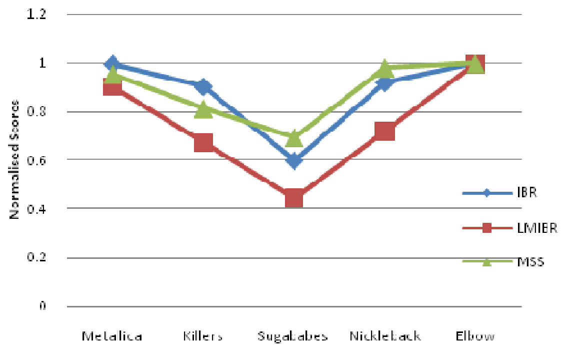


Figure 4. Normalised Subjective & Objective scores vs. Excerpt

If we consider the deviation between low and mid range only, shown on the plots as LMIBR, the objective quality ranking becomes more pronounced.

The low and mid range frequency bands (as observed in this particular test) could therefore be attributable to having a greater effect on the perception of quality in music production.

Interestingly, if we examine the dynamic range measurements made on each of the excerpts across the three bands, see Figure 5, one can observe that whilst the lowest placed excerpt has a greater degree of dynamic range in its low frequency band, it's IBR score is equated as the lowest due to its mid and high band correlation.

| Ex | LF Dr | MF Dr | HF Dr | IBR | LMIBR |
|---|---|---|---|---|---|
| **Metallica** | 8.1577 | 14.6563 | 17.0584 | 4.6048 | 4.5952 |
| **Killers** | 10.3824 | 15.2067 | 18.7134 | 4.1828 | 3.4113 |
| **Sugababes** | 11.7525 | 14.9279 | 17.2871 | 2.7773 | 2.2453 |
| **Nickleback** | 9.8965 | 15.0646 | 18.3508 | 4.2619 | 3.6544 |
| **Elbow** | 8.7291 | 15.8984 | 17.3616 | 4.6199 | 5.0695 |

Figure 5. DR across bands and Inter-band Relationship

One can see from figure 5 that there is a close correlation between mid and high band dynamic ranges extracted for the excerpts placing in the top 3 (that of Elbow, Nickleback & Metallica). However, if one examines the low frequency band dynamic ranges and compare them to those of the lowest two ranked excerpts, one can observe that these are smaller. This suggests that whilst the top three excerpts have a reduced dynamic range in their low frequency band, the differential between their respective mid-high frequency band ranges is greater, resulting in a greater IBR score.

This could suggest the importance of controlling the dynamics of the low frequency bands, often referred to as 'tightening' resulting in a subjectively powerful / punchy mix. This technique, common to the Rock/Metal genre, would probably have been applied to Metallica & Nickleback. Interestingly, the elbow sample, despite not being in the same genre, was a production that attempted to adhere to the 'Turn It Up' movement by keeping as much of the dynamic range intact during mastering.

Low frequency content of produced pieces of music contribute greatly to the spectral energy of the piece, therefore a loss in this energy could result in a perceptual loss of audio quality by the subject.

For each excerpt a single IBR measure was extracted, this utilized a fixed widow size based on the sample length. It is the authors' intension to extend this work by examining the IBR value with respect to time, thus showing the variation in dynamics temporally.

A temporal based IBR score may identify particular 'dense' moments in a musical arrangement. These would relate to very low IBR values.

Previous work by Lund proposed a measure of 'consistency' [11] which measures *the variation of loudness on a macroscopic timescale.* He describes the application of a loudness-correction processor increasing the consistency of the musical material. The authors agree that this is the typical approach applied to music during the mastering stage and often results in loss of dynamic range across the frequency range. Lunds' measure of consistency is based up statistical distribution of measured loudness utilising the BS.1770 loudness algorithm as a starting point.

It uses a statistical distribution to prevent the measure from being skewed by short but loud musical sequences and/or fadeouts.

The IBR method does not employ any method of statistical distribution or weighting at this stage. Short loud sequences would, if they were loud due to a large average correlation of frequency bands, give a low IBR score. As this method is based upon frequency band correlation, it can be assumed that during fade ins and outs the relative measure between peak and R.M.S levels across all frequency bands used in the calculation of the IBR score would remain constant. Therefore, the IBR score would be unaffected.

In addition, since the IBR score is derived by measuring the correlation between relative dynamic ranges across frequency bands, it is unaffected by the overall playback level selected by the listener. Thus allowing for a qualitative measure to be made prior to amplification taking place.

.

## 6. CONCLUSIONS

This paper represents a wider study into the IBR model output variable. The comparison of the objective IBR measurement with subjective results suggest that the IBR shows potential as a measure to assess audio quality with respect to dynamic range.

A reduction in dynamic range in a single frequency band does not necessarily result in a perception of low quality by the listener, rather, the relationship between the dynamic ranges in bands has been shown to correlate to this score.

Due to the wide variation in spectral content between pieces of produced music, in addition to fade outs and fade ins a wideband figure of dynamic range is not accurate enough to be utilised as a figure of merit score of quality. However, the use of an IBR score such as that proposed could be utilised as a more accurate measure than its wideband counterpart.

## 7. FURTHER DEVELOPMENTS

Detailed analysis is required to study the relationship between critical bands with respect to their dynamic range, both in their short term and long term measurement, and how this relates to our perception of audio quality in terms of MSS.

Due to its simplicity, the calculation of the IBR variable could lend itself well to real time implementation. Thus, it could give an indication of projected perception of quality by the end user, when used as a tool in music production.

A more accurate model of the basilar membrane will be utilised to separate out and measure the dynamic range across all 24 critical bands.

Additional study of 'produced' music will be undertaken to establish a mean dynamic range across these critical bands and map this to a quality score.

An IBR model will be developed that will adapt to temporal changes in the audio under test and utilise a dynamic window size.

## 8. REFERENCES

[1] Feaster P, Edouard-Leon Scott de Martinville's – Principes De Phonoautographie (1857)" , Firstsounds.org, U.S.A. (2008) [Accessed 12th September, 2009]

[2] S.Bech and N.Zacharov, Perceptual audio evaluation, theory, method and application, J.Wiley, Chichester, 2006

[3] G. Stoll, F.Kozamernik, EBU listening tests on internet audio codecs, EBU Technical Review, 2000

[4] D.Marston and A.Mason, Cascaded audio coding, EBU Technical Review 304, Geneva, Switzerland, 1994

[5] ITU-R BS.1116-1, Methods for the subjective assessment of audio systems including multichannel sound systems, International Telecommunications Union, Geneva, Switzerland, 1994

[6] ITU-R BS.1534-1, Method for the subjective assessment of intermediate quality level of coding systems, International Telecommunications Union, Geneva, Switzerland, 2003

[7] ITU-T. P800, Methods for subjective determination of transmission quality, International Telecommunications Union, Geneva, Switzerland, 1990

[8] S.Zielinski and F.Rumsey, On some biases encountered in modern audio quality listening tests-A Review, Journal AES Vol 56, No 6, June 2008.

[9] ITU-R BS.1387-1, Method for objective measurement of perceived audio quality', International Telecommunications Union, Geneva, Switzerland, 1998

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

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

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

[13] Loudness Wars, http://en.wikipedia.org/wiki/Loudness_war [Accessed 24th September, 2009]

[14] Turn Me Up, http://www.turnmeup.org/ [Accessed 12th September, 2009]

[15] E.Vincent, MUSHRAM 1.0, Centre for Digital Music, Queens Mary, University of London, November 2005.

[16] ITU-R BS.1284-1, General methods for the subjective assessment of sound quality International Telecommunications Union, Geneva, Switzerland, 2003

[17] AES-6id-2006, AES information document for digital audio – Personal computer audio quality measurements, Audio Engineering Society, Inc, 2006

[18] SQAM Test CD, Sound Quality Assessment Material, Recordings for subjective tests – Cat. No. 422 204-2, EBU 1988,2005.

[19] Pleasurize Sound Foundation, http://www.dynamicrange.de/node/1 [Accessed 24th September, 2009]

[20] S.Fenton, B.Fazenda & J.Wakefield, Objective Quality Measurement of Audio Using Dynamic Range Analysis, IOA-25 Conference paper, November 2009

[21] AES-6id-2006, AES information document for digital audio – Personal computer audio quality measurements, Audio Engineering Society, Inc, 2006.

[22] J.Villegas, M.Cohen, "Roughometer": Realtime Roughness Calculation and Profiling 125th Convention, Convention Paper 7516, Audio Engineering Society,San Francisco, CA, USA, October 2008.

[23] J.Boley, M.Lester, C.Danner, "Measuring Dynamics: Comparing and Contrasting Algorithms for the Computation of Dynamic Range" 129th Convention, Convention Paper 8178, Audio Engineering Society,San Francisco, CA, USA, November 2010.

[24] S.Ferguson, D.Cabrera, E.Schubert " Comparing continuous subjective loudness responses and computational models of loudness for temporally varying sounds", 129th Convention, Convention Paper 8177, Audio Engineering Society,San Francisco, CA, USA, November 2010.

[25] Moore, B.C.J., Glasberg, B.R. & Stone, M.A "Why Are Commercials so Loud? -- Perception and Modeling of the Loudness of Amplitude-Compressed Speech", Journal of the Audio Engineering Society, vol.51:12, 2003