# University of Huddersfield Repository

Fenton, Steven and Lee, Hyunkook

Towards a Perceptual Model of 'Punch' in Musical Signals

## Original Citation

Fenton, Steven and Lee, Hyunkook (2015) Towards a Perceptual Model of 'Punch' in Musical Signals. In: 139th International AES Convention, 29th October - 1st November 2015, New York, USA.

This version is available at http://eprints.hud.ac.uk/id/eprint/26288/

# Towards a Perceptual Model Of 'Punch' In Musical Signals

Steven Fenton, Hyunkook Lee, and Jonathan Wakefield

*School Of Computing & Engineering (MTPRG), University of Huddersfield, Huddersfield, UK*
s.m.fenton@hud.ac.uk, h.lee@hud.ac.uk, j.p.wakefield@hud.ac.uk

**ABSTRACT**

This paper proposes a perceptual model for the measurement of 'punch' in musical signals. Punch is an attribute that is often used to characterize music or sound sources that convey a sense of dynamic power or weight to the listener. A methodology is explored that combines signal separation and low level parameter measurement to produce a perceptually weighted 'punch' score. The parameters explored are the onset time and frequency components of the signal across octave bands. The 'punch' score is determined a weighted sum of these parameters using coefficients derived through a large scale listening test. The model may have application in music information retrieval (MIR) and music production tools. The paper concludes by evaluating the perceptual model using commercially released music.

## 1.    INTRODUCTION

This paper proposes a perceptual model for the measurement of 'punch' in musical signals. The model consists of a signal separation stage followed by a weighted stage. Previous work by the authors [1] explored the reverse elicitation of parameters pertaining to the sensation of punch within a music signal. From this work subjectively graded punch samples were obtained after the application of temporal wave shaping. Regression analysis of the high-level control settings chosen by the expert listeners revealed no significant correlation between any singular control setting and the resulting punch score. Parameters modified in the experiment were attack time, decay, sustain and release within three separate frequency bands. The results confirmed the authors belief that rather than any one particular control setting being responsible for punch modification, a number of low level resulting parameters must be attributable. The important factor lies in not the process involved in audio modification but rather the final signal. To explore this concept further a method was proposed that could separate the complex musical signal into its component transient, steady state and residual components [2]. The technique allows the transient portion of the musical signal to be analyzed independently of the other components. By separating the signal, low level parameters can be extracted and a perceptual model can be applied solely to the transient component of the signal.

The multi-resolution technique employed used a combination of quadrature mirror filter bank (QMF), short time Fourier transform (STFT) and median filtering to extract the components of the signal under test. The signals used in testing were complex audio (polyphonic) musical works consisting of drums, bass and guitars. The method was successful in achieving source separation with respect to allowing both the steady state tonal and transient components to be independently measured and described independently and in relation to one another.

The centroid showed correlation to the instrument type in the transient components, thus allowing a possible method for automatic musical transcription. Similarly, in the case of the steady state component, the centroid measure correlated well with both the bass note melody and ascending rising chord structures and pitch bends within the guitar based piece.

This paper aims to continue this work by examining the onset time and frequency components of the signal across octave bands. The purpose of which is to identifying those that show correlation with the perceptual attribute of 'punch'. A model is then proposed which utilizes the weightings obtained from this analysis.

## 2.    NOISE BURST LISTENING TEST

A controlled listening test undertaken by 11 expert listeners. The test involved each listener grading the perceived punch of 45 shaped pink noise bursts. The primary focus of this test was to establish which octave bands were most relevant to punch perception and also how onset time may affect this in each case.

The 45 samples were all 80ms in length, 16bit and presented in mono. The samples were arranged in 9 octave bands and each was shaped with varying onset times of 0ms, 5ms, 10ms, 20ms and 60ms. All the samples had a fixed offset time of 40ms to help negate any offset effects. They were also loudness normalized using a two stage filtering/gain algorithm, as described in section 4.1.

An informal listening test took place prior to the main test to briefly evaluate the effectiveness of the loudness normalization algorithm. Of the 11 participants that took part, 7 agreed that the loudness of the pulses were presented at roughly equal loudness. The remaining 4 suggested that the differences they were hearing were primarily as a result of timbral differences rather than loudness. Adjustments proposed were recorded but not utilized in the main testing as the majority of changes proposed were of a magnitude of 2dB or less.

Each octave band pulse was measured at the listening position using an SPL meter, calibrated to operate without weighting. The SPL measured was 76dB.

Following the informal loudness evaluation, the 45 shaped noise burst were presented using a test interface created using Hulti-Gen test suite for Max [3]. Figure 3 shows the test interface. Each sample was played in a random order and each participant was asked to rank the perceived punch against a 1kHz band reference burst. Scores were collected with 1 and -1 corresponding to the extremely punchier and extremely less punchier limits. A resolution scale of 100 points was used between these limits.
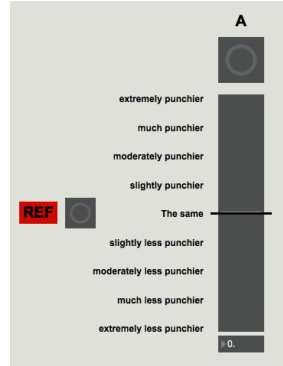


Fig1. Test Interface – Punch Perception Test

## 2.1 Loudness Normalization

In order to present the noise burst samples at equal loudness to the listeners, the samples were pre-processed with both a spectral and temporal weighting coefficients. The spectral weighting curve was based upon that set out in the ITU Recommendation BS. 1770-3 [4]. Modifications to the curve were adopted, namely the gain of the pre-filter was modified to 10dB as opposed to 4dB and the corner frequency of the shelf was adjusted to 1kHz rather than 1.6kHz. Figure 1 shows the resultant octave band and spectral weighting filters. Also shown is the inverse of the spectral weighting filter. These modifications were based on recommendations made by Pestana et al. [5] and through the authors' perceptual observations and testing.

The ITU-R 468-4 filter model [6] was also tested however it was found that the 2kHz-8kHz octave bands were perceptually significantly louder on playback than the lower bands.

Studies have shown that temporal effects, particularly for signals of less that 100mS in duration, must also be accounted for in order to weight the loudness appropriately [7, 8 & 9]. A temporal weighting coefficient (*TWC*) was derived using both the centre frequency of the octave band (*Fc*) and the signal duration (*t*). A calculation of Tau (T) is first derived and this is then used to derive the weighting coefficient.

$$T = [-0.032 * \ln(Fc)] + 0.3095 \qquad (1.)$$

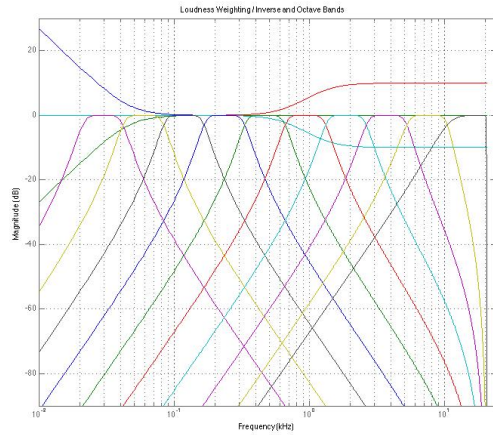$$TWC = \left[ 1 - E^{-t/T} \right]^{-1} \qquad (2.)$$

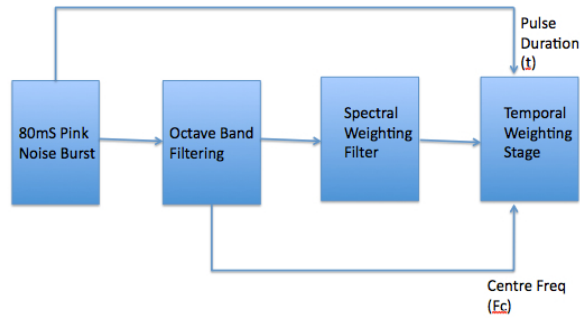Fig 2. Loudness Weighting (and inverse) and Octave Band Filters.



Fig 3. Loudness Compensator

## 3.    RESULTS

The results of the noise burst listening test represent the punch coefficients obtained at each octave band and onset. The '0' point on the y-axis indicates the noise burst was perceived to have the same punch level as that of the 1kHz noise burst. A positive value indicates increased punch, negative indicates less.
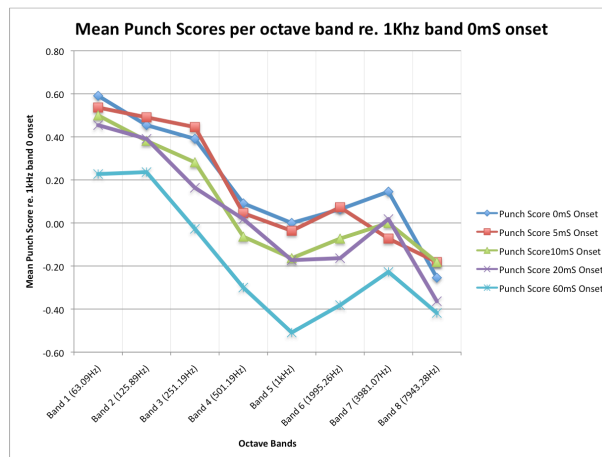


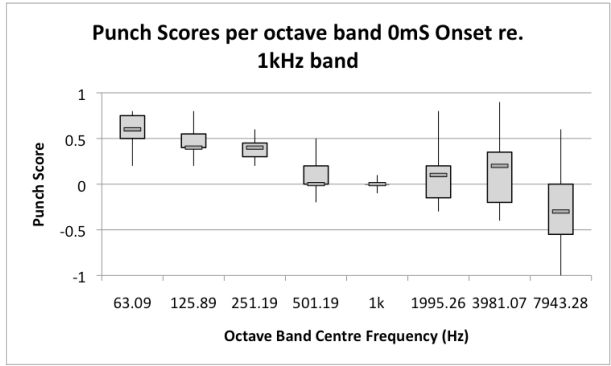Fig 4. Mean Punch Scores relative to 1kHz Octave Band

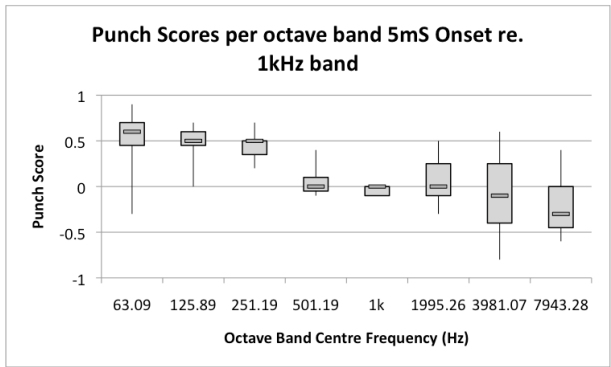Fig 5. Punch Score 0mS Onset Noise Burst
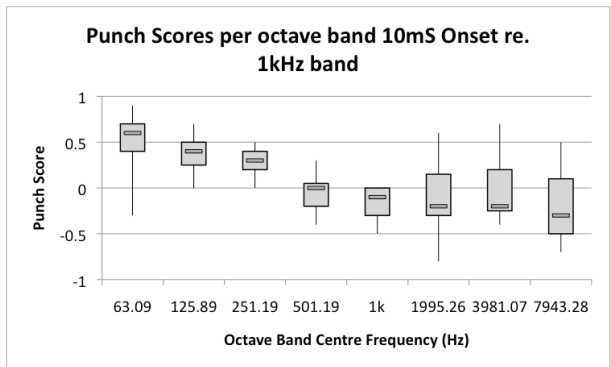


Fig 6. Punch Score 5mS Onset Noise Burst
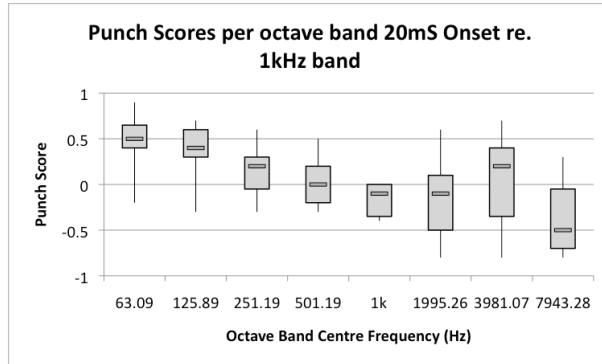


Fig 7. Punch Score 10mS Onset Noise Burst
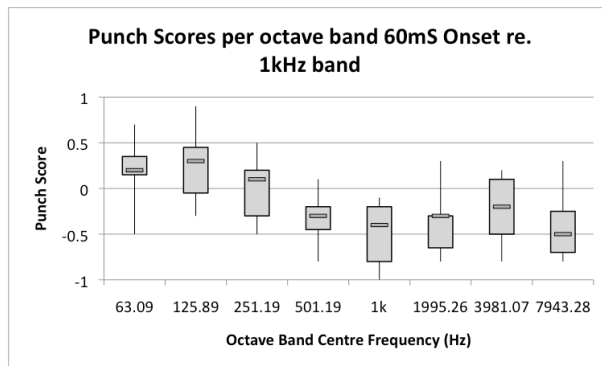
Fig 8. Punch Score 20mS Onset Noise Burst



Fig 9. Punch Score 60mS Onset Noise Burst

## 4.    TEST ANALYSIS AND MODEL PARAMETERS

Initial observations of results, see figure 4, indicated that the mean punch score was related to both the noise burst centre frequency and onset however there appears to be a pivot point around the 1kHz reference band whereby the 2kHz and 4kHz bands have an upward trend. Figures 5-9 show the results in box plot format showing the distributional characteristics of the punch scores as well as the levels.

With reference to the inter-quartile ranges, the greatest degree of variation of punch scores can be seen in the 2kHz, 4kHz and 8kHz bands. Again this trend, in addition to the upward trends shown in figure 4 suggest a greater difference of opinion as to whether the upper bands have a greater punch perception than the 1kHz reference.

Upon interviewing the participants after the experiment, it was found that whilst the upper bands were not necessarily more punchy than the reference, in some cases some had been scored higher as a result of their timbral weight or presence. This being particularly relevant to the 4kHz band.

Multiple linear regression techniques are applied to the results of the noise burst experimental results in order to derive model parameters relating to the noise burst experimental parameters. Due to the pivot point around 1kHz a model utilizing only the 64Hz to 1kHz centre frequencies was employed. Further work to establish the parameters for the 2kHz-8kHz range is on going. The dependent variable in the regression analysis was the Punch Score and the independent variables were Band and Onset.

| R | $R^2$ | Adjusted $R^2$ | Std. Error of Estimation |
|---|---|---|---|
| .952 | .906 | .898 | .09321 |

Table 1    Summary for Punch Score Model

| | SS | df | Mean Squ | F | Sig |
|---|---|---|---|---|---|
| Regression | 1.846 | 2 | .923 | 106.22 | .000 |
| Residual | .191 | 22 | .009 | | |
| Total | 2.037 | 24 | | | |

Table 2    ANOVA summary for Punch Score Model

The Sum of Squares of the residuals is the total deviation of the response values to that of the model prediction values, a value of .191 (see table 2) shows a very tight fit of the model to the data. The resultant coefficient of determination shown in table 1, denoted by the $R^2$, is derived from the sum of squares further indicates that the model is a good fit. The adjusted $R^2$ value is not adversely affected by the low number of data points (11 expert listeners and 5 bands used). The standard error predicted by the model is also very small.

The effect of the regression is statistically significant, thus the effect on the punch scores is mainly predictable through variation of the independent variables rather than chance.

| | Unstand Coeff | | Stand Coeff | | |
|---|---|---|---|---|---|
| | B | Std.Err | Beta | t | Sig |
| Const | .954 | .059 | | 16.185 | .000 |
| Band | -.171 | .013 | -.845 | -12.941 | .000 |
| Onset | -.088 | .013 | -.438 | -6.706 | .000 |

Table 3    Coefficients for Punch Score Model

Table 3, showing the raw coefficient score (partial regression coefficient scores B), indicate that the Band would have both the largest effect and be detrimental to the punch score as it is increased. This trend can be observed in figures 4-9. The effect of the Onset is somewhat smaller, again having a detrimental effect as it is increased. One might expect, and indeed it can be observed particularly from figure 9, that the longer onsets have more of an impact in the upper bands with respect to a lowering of the punch score.

The resultant estimated model is as follows, where $X$ & $Y$ are Band and Onset respectively. Figure 10 shows the model output compared to the subjective results obtained.

$$EstPunch = .954 - (X * 0.171) - (Y * 0.088) \qquad (3.)$$
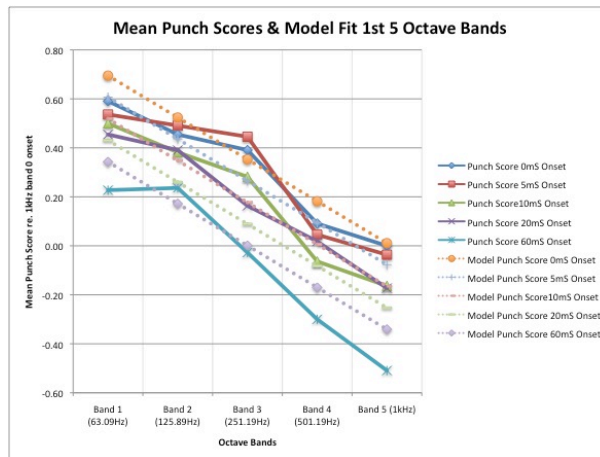


Fig 10. Model vs Subjective results

## 5.  MODEL IMPLEMENTATION

The model employs a multi-stage approach as detailed in figure 15. Firstly, the complex musical signal is separated into its component transient, steady state and residual components. The technique allows the transient portion of the musical signal to be analyzed independently of the other components which has significant advantages over approaches that consider the signal in its mixed state

The separation process offers good time and frequency resolution by employing a combination of quadrature mirror filter bank (QMF), short time Fourier transform (STFT) and median filtering. This pre-processing is explained in detail in [3] and was based on work by Fitzgerald et al. [10, 11].

The transient component is then fed into a modified loudness filter, as described in section 3.1. Each octave band is then independently weighted based on coefficients outlined in Table 3 and equation 3 and the energy of each is summed together to provide and overall punch indicator.

The summing process is based on the block based momentary loudness model [4] however a smaller block size was incorporated to lower the level of signal integration taking place and also to allow for inclusion of onset detection (shown below as Temporal Data) into the model in the future.
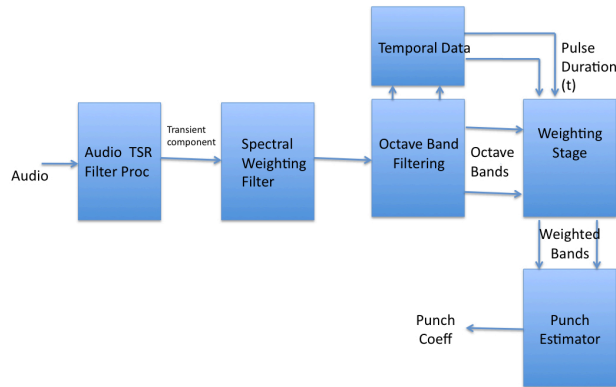
Fig 15. Punch Model Diagram

The model was implemented using Matlab.

## 6.    MODEL OUTPUT

The following plots represent the model output with respect to varying input stimuli. Figures 11 and 12 represent the output of the model when presented with the test stimuli used in the subjective tests, these are the octave spaced progressively increasing centre frequency noise bursts. Figure 11 simply shows the outputs of the model without weightings being applied. This is to allow comparison with the standard momentary loudness model output shown in figure 13.

The 0dB point on the figures represents full scale. Level output is similar to that of the standard loudness model such that if the input stimulus is a full scale digital broad band pink noise burst, the output of the model would be -3dB. This is shown in figure 14.

Comparing figures 11 and 13, the outputs of the models are very similar. However, due to the smaller integration window used in the punch model, the output is able to clearly differentiate the individual noise bursts and associated level. Figure 12 is showing the output of the model with weightings enabled and consideration of the first 5 bands being summed. As the noise burst centre frequency is increased, one can see that the output of the model drops, as expected when compared to figure 10.

Figures 15, 16, 17 & 18 compare the output of punch model with commercially available music samples with the standard momentary loudness model.

Figure 15 shows the momentary loudness of an excerpt from Michael Jackson - 'Billie Jean'. Using loudness as a metric, it can be seen that there is little that can be obtained from this plot in terms of signal dynamic. Comparing this to the model output, figure 16, the signal dynamic with respect to the transient components is clearly visible. More punch is evident at 1 second intervals starting from 0s, less punch is evident starting at 0.5s using the same interval.

Figures 17 and 18 are measures of Def Leppards – 'Animal'. As with the previous sample, more signal dynamic is evident with the punch model in addition to indication of relative levels with respect to the transient components. Comparing the output of the model in these two commercial examples (figures 16 & 18), the punch model indicates that Billie Jean has both a higher level of punch and more dynamic fluctuation within it. When listening to the two samples, in the authors opinion this is the case.
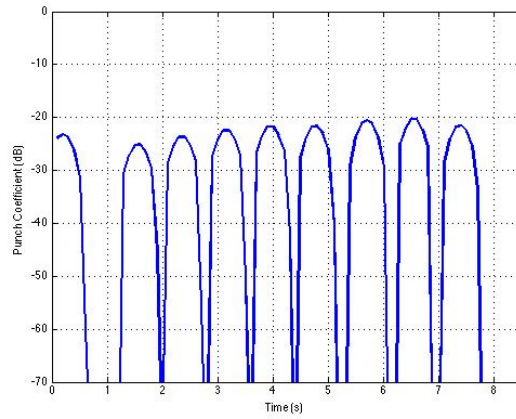
Fig 11. Measurement Of Noise Bursts, progressive octave bands using 280mS Momentary Octave model.
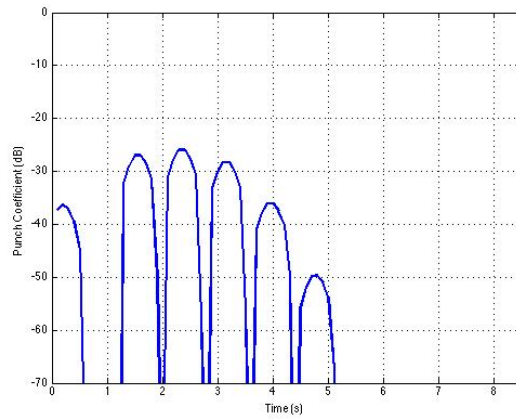


Fig 12. Measurement Of Noise Bursts, progressive octave bands using 280mS Momentary Punch Weighted model.
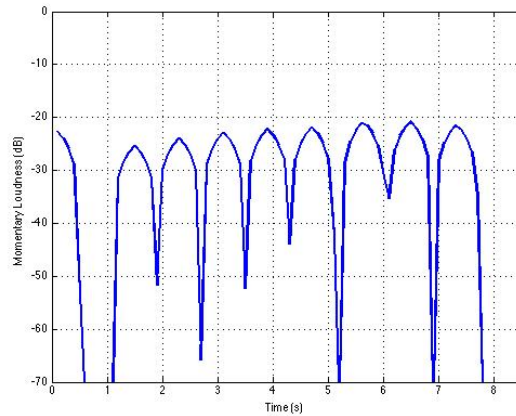


Fig 13. Measurement Of Noise Bursts, progressive octave bands using standard 400mS Momentary Loudness model
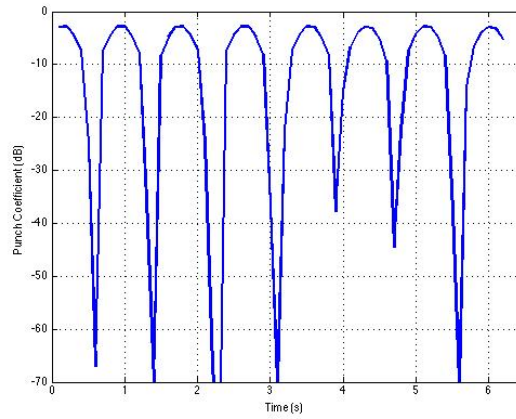
Fig 14. Measurement Of Full Scale, Broadband Pink Noise Bursts using the model.
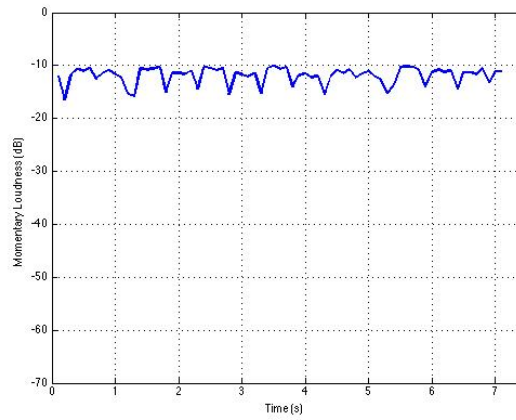


Fig 15. Measurement Of Billy Jean Sample, using standard 400mS Momentary Loudness model
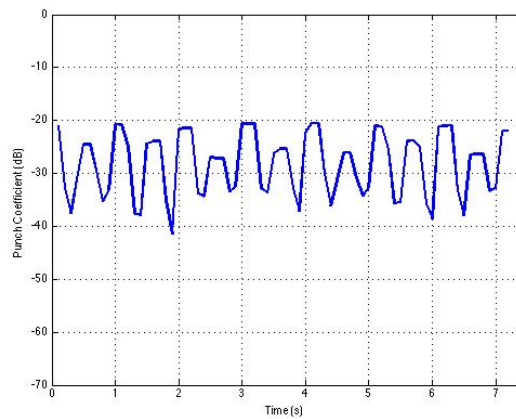


Fig 16. Measurement Of Billy Jean Sample, using 280mS Momentary Punch Weighted model.
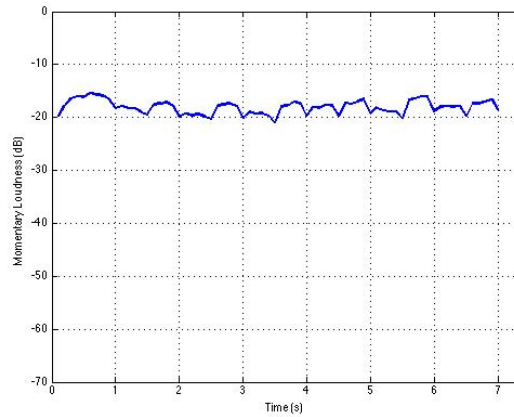
Fig 17. Measurement Of Animal Sample, using standard 400mS Momentary Loudness model.
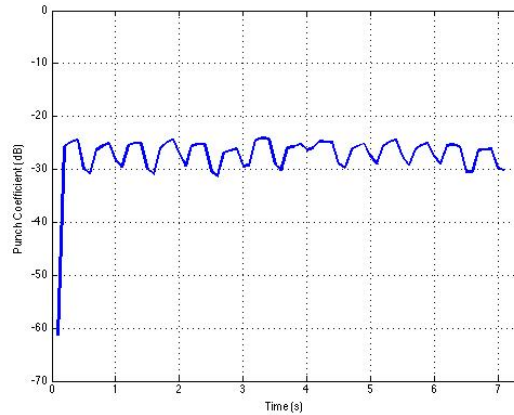


Fig 18. Measurement Of Animal Sample, using 280mS Momentary Punch Weighted model.

## 7.   CONCLUSION

The model yields the possibility to perceptually weight the transient components of an audio signal. In doing so, output relative to the perception of punch in the signal is possible. Analysis of these components may yield parameters that could be of use in both mixing/mastering and also audio transcription and retrieval.

## 8.   FURTHER WORK

Currently the models sidechain for temporal data extraction is being developed. It is hoped that the testing of which will yield further interesting results and improve the model.

Further subjective tests are planned with a panel of expert listeners. The listening test will attempt to evaluate the effectiveness of the punch measure across differing genres in commercial music.

Comparisons of the model to measures such as LRA [12] will be  investigated along with the expansion of the model with respect to other low level descriptors that have been investigated previsouly.

## 9. REFERENCES

[1] S.Fenton, H. Lee and J.Wakefield, "Elicitation and Objective Grading of 'Punch' Within Produced Music. 136th Audio Engineering Society Convention, Berlin, April 2014.

[2] S.Fenton, H.Lee and J.Wakefield,, "Hybrid Multiresolution Analysis Of 'Punch' In Musical Signals", AES 138 Convention, Warsaw, May 2015.

[3] C.Gribben and H.Lee "Hulti-Gen - Huddersfield Universal Listening Test Interface Generator", 2015, available from : http://www.hud.ac.uk/research/researchcentres/mtprg/projects/apl/

[4] ITU-R BS.1770-3, Algorithms to measure audio programme loudness and true-peak audio level, International Telecommunications Union, Geneva, Switzerland, 2012

[5] P Pestana, J Reiss and A Barbosa, "Loudness Measurement of Multitrack Audio Content using Modifications of ITU-R BS.", 134th Audio Engineering Society Convention, Rome, May 2013.

[6] ITU-R 468-4 Measurement of audio-frequency noise voltage level in sound broadcasting, International Telecommunications Union, Geneva, Switzerland, 1986.

[7] C.Watson and R.Gengel, "Signal Duration and Signal Frequency in Relation to Auditory Sensitivity", JASA, Vol. 46, No 4 (Part 2), 1969

[8] J.Verhey and B.Kollmeier, "Spectral Loudness Summation as a Function of Duration", JASA, Vol. 111, No 3, March 2002.

[9] J.V and R.A. Rasch, "The Perceptual Onset of Musical Tones", Perception and Psychphys, vol 29, 1981.

[10] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audiosignal into harmonic/percussive components by complemen- tary diffusion on spectrogram," in *Proc. of the EU- SIPCO 2008*, Lausanne, Switzerland, Aug. 2008.

[11] D. Fitzgerald, "Harmonic/percussive separation using median filtering." Proc. of the DAFx-10, Graz, Austria, Sept. 2010.

[12] E.Skoveborg, "Measures Of Microdynamics", AES 137 Convention Paper, Los Angeles, October 2014.