DEVELOPMENT OF A REAL-TIME PUNCH METER PLUGIN

Andrew Parker, Steven Fenton and Hyunkook Lee

Applied Psychoacoustics Lab (*APL*)
University of Huddersfield
Correspondence should be addressed to Andrew.Parker@hud.ac.uk

ABSTRACT

In this paper, a real-time implementation of a punch metering plugin is described. 'Punch' is a perceptual attribute and can be defined by both temporal and frequency characteristics of an audio signal. The metering tool consists of signal separation, onset detection, and perceptual weighting stages. Scores are displayed on both a time graph and a histogram; statistical metrics are derived from the histogram. The output is compared to subjective punch scores obtained from a controlled listening test, and showed a 'strong' correlation with Pearson and Spearman coefficients r=0.840 (p<0.001) and rho=0.937 (p<0.001) respectively.The meter is intended to allow for optimisation and objective control of punch during mixing, mastering, and broadcast.

1. INTRODUCTION

Many different types of meter are used within music production to optimise various perceptual attributes of a given signal. One such attribute is commonly described as 'punch'. Previous work [1] defined 'punch' as relating to the magnitude of transient change within a signal, and the frequency content of these transient events. Following this a perceptually motivated model [2] for punch was proposed.

In this paper a real-time implementation of the model is outlined. The model has been optimised for real-time use by reducing its look-ahead dependency, allowing it to operate on a frame-by-frame basis. Additionally, an interface for the display of the real-time measures is presented, showing a momentary and short-to-long term overview of the signal being measured.

Objective evaluation of the system was performed through correlation analysis of the meter's output against scores gathered from controlled subjective listening tests.

2. REAL-TIME MEASUREMENT SYSTEM

As punch perception is dependent on transient changes in the signal, the transient components of the signal must be isolated. Once isolated, the transient information can be perceptually weighted, and scores can be generated. Figure 1 shows an overview of the processing used to achieve this.

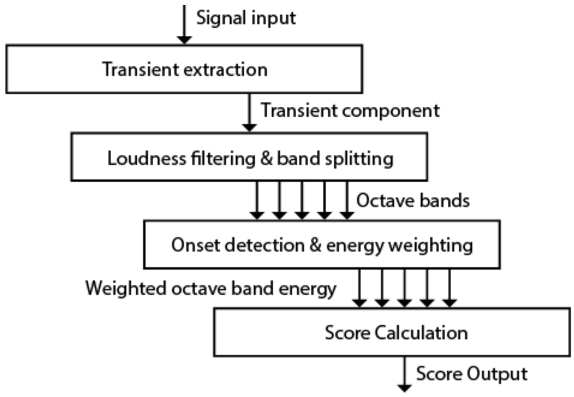


Figure 1 : Block diagram representing the signal flow through the real-time punch meter.

Transient extraction is achieved through the use of a binary mask and short time Fourier transform (STFT) processing. Frequency bins are removed from the spectrogram if they are deemed not to be transient. Mask generation is determined by considering both magnitude and phase in the complex domain resulting in a Euclidean distance deviation coefficient [3]. The deviation is compared between the current and preceding frames for each frequency bin. The deviation value follows the transient envelope of the input signal. An adaptive threshold is used to determine if a given frequency bin will be included in the binary mask for that frame. If the deviation value is below the adaptive threshold, the corresponding frequency bin of the input spectrogram will be nulled at the output. Once masked, the transient signal can be synthesized using the inverse short time Fourier transform (ISTFT).

The transient signal is filtered to account for the influence of perceived loudness. The loudness filtering follows the ITU-R BS.1770-4 recommendation [4], with a modified pre-filter gain and shelf corner frequency which were found to be more relevant in a previous study [2]. Following this, the signal is filtered into five octave bands: 63Hz, 125Hz, 250Hz, 500Hz, 1000Hz.

Onset detection and weighting is performed on each octave band. Energy values are calculated for 100ms frames with a 50% overlap. An adaptive threshold [5] is used to peak-pick from these frames.

If a frame is picked as a peak, the frame's energy is examined in finer resolution to find the onset duration, shown in Figure 3. If more than one onset in a frame is detected, only the onset with the largest energy peak is calculated.

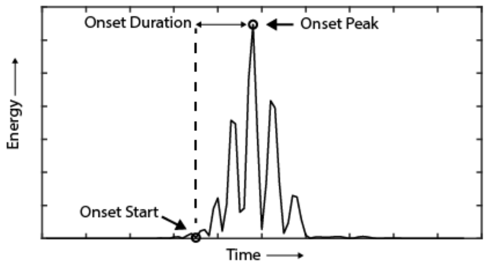


Figure 2 : Graph of a picked peak and its calculated onset location, peak location, and duration.

Based on the octave band and onset duration, the energy of the frame is weighted. The weighted energy is summed across corresponding octave bands in a logarithmic power equation. This power value is the ‘punch score’ for the given frame. These scores can be analysed momentarily or statistically over a time period.

2.1. Statistical Metrics

A number of metrics are statistically derived from a histogram of the punch scores. The histogram of punch scores collected over the period of time represented by the time graph. This was arbitrarily 141 frames or 7.05 seconds at 44.1kHz sampling frequency.

- *Mean* - The running mean of the punch scores in decibels.
- *P95* - The ninety-fifth percentile of punch scores in dB; the border of the five percent of frames with the highest punch scores.
- *P95M* - The P95 value divided by the mean of the punch scores. This is a ratio similar to the peak-to-loudness ratio in loudness metering. In this case, it shows a ratio of average punch scores to the P95 value. For display purposes, this value can be multiplied by 100 and inverted (as is the case shown in Figure 3).

3. INTERFACE

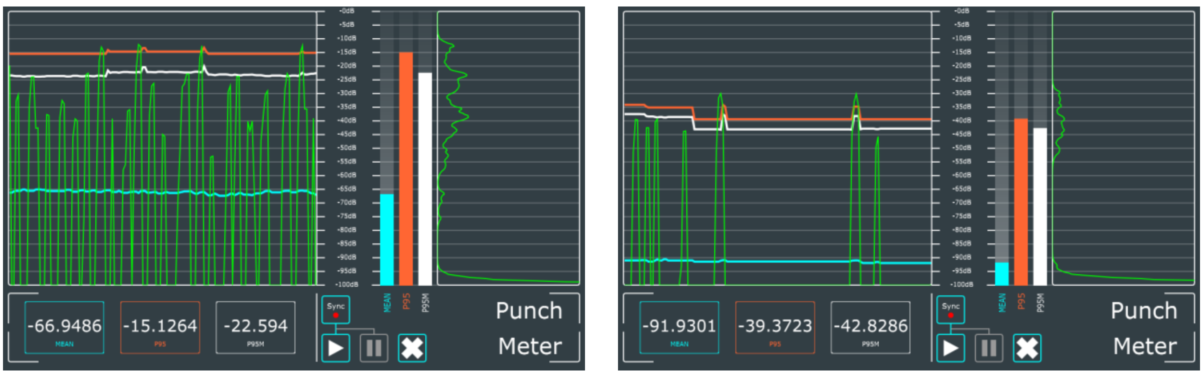
Figure 3 demonstrates a comparison between the measurements of two different tracks. Both tracks in this comparison were loudness normalized to -23LU. Figure 3a shows the response to an excerpt from ‘The Real Slim Shady’ by Eminem, which was shown to have a high level of perceived punch in a subjective listening test [6]. Figure 3b shows the response to an excerpt from ‘Allegro Con Brio’ by Beethoven, which has comparatively fewer and softer onsets and which had a low level of perceived punch in the same test [6].

The most recent 141 punch frame scores are displayed as a right to left scrolling history. There are four lines representing the punch frame score, mean, P95, and P95M values. This display allows for simultaneous views of momentary and statistical scores.

A histogram is shown to the right of the interface which is derived from the historical punch scores. A smoothed line is shown representing punch score (vertical axis) versus relative frequency density (horizontal axis). This gives a longer-term overview of how punch is evolving.

The statistical measures are calculated from the histogram data. They are shown as numerical values, and also on their respective bar graphs to the left of the histogram. The value readouts allow for a more measurable view of the results, and the bar graphs visually highlight the focal points of both the histogram and time graph data.

Controls are provided which allow the user to play or pause score generation. If ‘Sync’ is enabled, this manual control is relinquished, and the generation of new scores follows the host’s playback. The reset button, represented by the ‘X’ symbol, can be used at any time to clear the history of scores, resetting them to -100.



(a) 'The Real Slim Shady' by Eminem. (b) ‘Allegro Con Brio’ by Beethoven.

Figure 3: Example of the proposed real-time punch meter display.

4. OBJECTIVE EVALUATION

Subjective data from Fenton et al.'s study [6] was used to evaluate the system. In this study, 14 listeners graded normalised (-23LU), monophonic musical audio stimuli from various genres for punch. A forced pairwise comparison test was used to rank the stimuli based on the level of perceived punch. Scaled response coefficients from the pairwise comparison data were then derived using a Bradley-Terry-Luce (BTL) model resulting in the overall subjective scores.

The real-time system's P95 and P95M outputs were recorded for the same audio files. Pearson (*r*) and Spearman (*rho*) correlation coefficients were then calculated for the objective metrics against the subjective scores. This showed how closely correlated the system's output and rank order of the audio files was to the subjective scores. The correlation results are shown in Table 1 and Figure 4. All of the coefficients resulted in a confidence of $p<0.001$.

Measure	Pearson Coefficient (*r*)	Spearman Coefficient (*rho*)
P95	0.840	0.937
P95M	-0.824	-0.958

Table 1 : Correlation of punch scores to subjective punch data

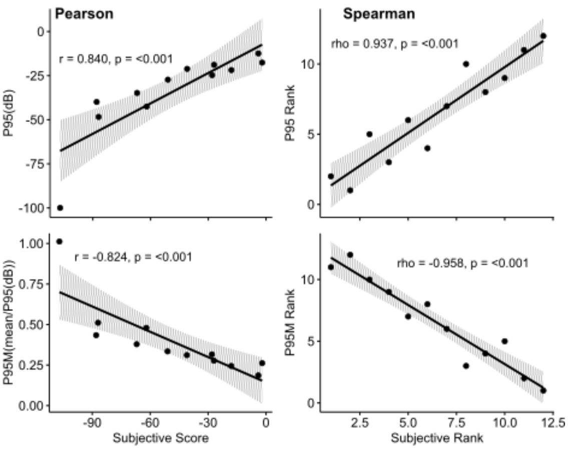


Figure 4 : Graphs to show Pearson (left column) and Spearman (right column) correlations between objective meter output and subjective scores. The shaded area shows the 95% confidence interval.

A 'strong' correlation was shown between the P95 and subjective scores, with Pearson and Spearman coefficients r=0.840 (p<0.001) and rho= 0.937 (p<0.001) respectively. The P95M score also showed a 'strong' correlation to the subjective scores, with Pearson and Spearman coefficients r=-0.824 (p<0.001) and rho=-0.958 (p<0.001). The P95M score's correlation coefficients are inverted due to inclusion of the mean in the calculation of this score.

5. CONCLUSION

A system for the real-time measurement of punch has been described. Scores are a result of transient isolation, and perceptual energy weighting based on octave band and transient onset time. A number of metrics are output.

The system was tested for its correlation to subjective punch scores using both Pearson and Spearman analysis. Both P95 and P95M metrics showed a 'strong' correlation to the subjective scores with both analysis methods, showing they good objective indicators of perceived punch.

Further work will be undertaken in the form of a controlled listening test using a larger number of stimuli, and a larger number of participants. Usability data will also be collected. There are several parameters within the system that may be optimised through further testing, which may increase the correlation of the system's output to the subjective scores. Metering ballistics for the tool will also be investigated.

6. REFERENCES

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

[2] S. Fenton and H. Lee, "Towards a Perceptual Model of 'Punch' in Musical Signals," *139th International AES Convention*, New York, USA (2015).

[3] J. P. Bello, C. Duxbury, M. Davies, M. Sandler, "On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain," *IEEE Signal Processing Letters* vol. 11, no. 6, pp 553—556 (2004).

[4] ITU-R BS.1770-4, Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level, International Telecommunications Union, Geneva, Switzerland (2015).

[5] J. Glover, V. Lazzarini, J. Timoney, "Real-Time Detection of Musical Onsets with Linear Prediction and Sinusoidal Modeling," *EURASIP Journal on Advances in Signal Processing* (2011).

[6] S. Fenton, H. Lee, J. Wakefield, "Evaluation of a Perceptually Based Model of 'Punch' with Music Material," *141st Audio Engineering Society Convention*, Los Angeles, USA (2016).