LOUDNESS ALGORITHMS FOR AUTOMATIC MIXING

Dominic Ward

DMT Lab
Birmingham City University
dominic.ward@bcu.ac.uk

Joshua D. Reiss

Centre for Digital Music
Queen Mary University of London
joshua.reiss@qmul.ac.uk

ABSTRACT

Accurate loudness measurement is imperative for intelligent music mixing systems, where one of the most fundamental tasks is to automate the fader balance. The goal of this short paper is to highlight state-of-the-art loudness algorithms to the automatic mixing community, and give insight into their differences when applied to multi-track audio.

1. INTRODUCTION

Loudness models of varying computational complexity have been used to automatically balance the levels of multi-track audio [1–3], yet little is known about how well they measure the relative loudness of individual instruments. For example, although [2] reported success using the EBU short-term loudness measure [4], [5] revealed a tendency for the metric to underestimate the subjective loudness of percussive material. This paper explores the predictions of both multiband and single band loudness models using different descriptors to quantify the overall loudness of a sound, and suggests directions for future work in the field.

2. METHODOLOGY

Three multiband models: GM02 [6], CF02 [7], and CH12 [8]; and three single band models: LARM [9], EBU [4] and V01 [10]; were compared. Given a waveform, each algorithm outputs a loudness time-function, from which a single value representing overall loudness must be determined. Developers generally suggest a statistic that quantifies central tendency, e.g. mean long-term loudness for the GM02. We use the term 'mean' to denote some form of temporal average, and 'peak' to denote the maximum. The algorithms were instructed to equalise the loudness of 110 short segments (RMS level = 73 dB SPL) of multi-track audio spanning a range of genres. The target loudness was taken as the average loudness of all segments, and an iterative procedure was used for the non-linear models. The GM02 and CH12 were configured to run at a lower complexity following suggestions given in [11]. Resulting level balances were centred on zero.

3. RESULTS

Figure 1 (A) shows the distribution of RMS levels after loudness equalisation, using peak loudness in the case of the multiband models. The spread of levels is markedly wide for the multiband procedures, indicating a greater sensitivity to the physical characteristics of the stimuli. The highest 5% of positive gains within each model were predominantly applied to bass instruments, demonstrating a common strategy across the algorithms to attenuate low frequencies when measuring loudness. The EBU programme loudness gives the narrowest spread, with 50% of the segment levels within 1 dB of the input level. The EBU, followed by LARM, was therefore the most consistent with a simple energy measurement. In contrast, the GM02 (mean loudness) applied gains as high as 31.6 dB for equal loudness (a bass drum segment). Thus, for projects involving a range of instruments, very different mixes can be expected from the algorithms. Subplot (B) shows the RMS errors between pairwise combinations of level balances. The single band models show greater agreement with one another compared to the multiband devices. The GM02 and CF02 yield notably different balances compared with those generated by the single band algorithms, especially when using the mean loudness descriptor.

Table 1 gives the RMSEs between the balances obtained using the two global loudness descriptors (mean or peak). The type of descriptor influenced the level balance most for the CF02, followed by the GM02. Our findings indicate that for the multiband algorithms, peak loudness is more appropriate when the sound corpus involves transient instruments, since averaging the loudness time series tends to underestimate salient peaks, unless specific envelope detectors or temporal weightings designed to emphasise them are incorporated as done by LARM and the V01, respectively. Interestingly, the predicted gains obtained using the EBU programme loudness and maximum EBU momentary loudness differ by only 1.6 dB, on average.

Model	RMSE (dB)	CI_{95} (dB)
GM02	5.6	[4.7, 6.4]
CF02	6.9	[5.7, 8.2]
CH12	2.7	[2.2, 3.2]
LARM	2.2	[1.9, 2.6]
EBU	1.6	[1.4, 1.7]
V01	2.0	[1.7, 2.2]

Table 1: RMSE (and bootstrapped 95% confidence intervals) between the loudness balances obtained using mean and peak loudness.

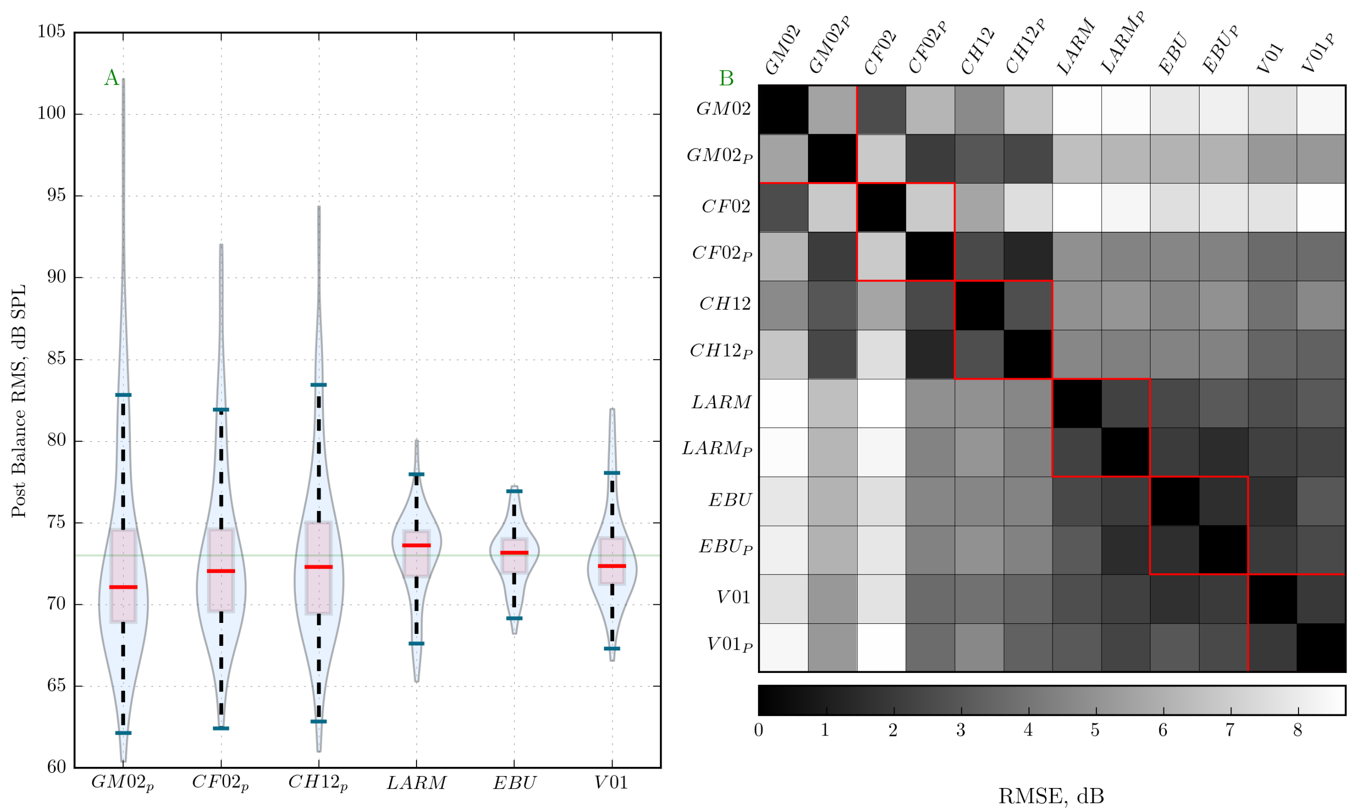

Figure 1: (A) Violin plots of the stimulus RMS levels after loudness equalisation, and (B) RMSE matrix for assessing balance similarity. The green horizontal line in (A) shows the input level of all segments. The subscript p denotes peak loudness.

4. CONCLUSION

Listening tests conducted by the authors suggest that the reference single band algorithms (mean descriptor) are robust to a broad range of content, and such large gains predicted by some of the complex auditory models may not be realistic. However, single band devices do not model auditory masking, a perceptual phenomenon that complicates many music production tasks. In this case, *partial* loudness calculation is more important, but further research into the generalisation of auditory models is needed first. In line with [5], future work should concentrate on fitting loudness models to a subjective reference dataset involving multi-track content, rather than programme material. Although at the present time the needed empirical data are unavailable, the audio segments, loudness predictions, level balances and details of model configurations are freely available at:
https://code.soundsoftware.ac.uk/hg/wimp16-ward-reiss.

5. REFERENCES

[1] E. P. Gonzalez and J. D. Reiss, "Automatic Gain and Fader Control For Live Mixing," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009.

[2] S. Mansbridge, S. Finn, and J. D. Reiss, "Implementation and Evaluation of Autonomous Multi-Track Fader Control," in *Proc. 132nd AES Convention*, 2012.

[3] D. Ward, J. D. Reiss, and C. Athwal, "Multi-Track Mixing Using a Model of Loudness and Partial Loudness," in *Proc. 133rd AES Convention.*, 2012.

[4] EBU, "EBU-R 128: Loudness Normalisation and Permitted Maximum Level of Audio Signals," tech. rep., 2014.

[5] P. D. Pestana, J. D. Reiss, and A. Barbosa, "Loudness Measurement of Multitrack Audio Content using Modifications of ITU-R BS.1770," in *Proc. 134th AES Convention*, 2013.

[6] B. R. Glasberg and B. C. J. Moore, "A Model of Loudness Applicable to Time-Varying Sounds," *J. Audio. Eng. Soc.*, vol. 50, no. 5, 2002.

[7] J. Chalupper and H. Fastl, "Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners," *Acustica – Acta Acustica*, vol. 88, 2002.

[8] Z. Chen and G. Hu, "A Revised Method of Calculating Auditory Excitation Patterns and Loudness for Time-Varying Sounds," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2012.

[9] E. Skovenborg and S. H. Nielsen, "Evaluation of Different Loudness Models with Music and Speech Material," in *Proc. 117th AES Convention*, 2004.

[10] E. Vickers, "Automatic Long-Term Loudness and Dynamics Matching," in *Proc. 111th AES Convention*, 2001.

[11] D. Ward, S. Enderby, C. Athwal, and J. D. Reiss, "Real-Time Excitation Based Binaural Loudness Meters," in *18th Int. Conf. DAFx*, 2015.