Audio Engineering Society

Convention e-Brief 497

Presented at the 146th Convention
2019 March 20 – 23, Dublin, Ireland

Automatic Mixing Level Balancing Enhanced Through Source Interference Identification

David Moffat[1] and Mark B. Sandler[1]

[1] *Queen Mary University of London*

Correspondence should be addressed to David Moffat (d.j.moffat@qmul.ac.uk)

ABSTRACT

It has been well established that equal loudness normalisation can produce a perceptually appropriate level balance in an automated mix. Previous work assumes that each captured track represents an individual sound source. In the context of a live drum recording this assumption is incorrect. This paper will demonstrate approach to identify the source interference and adjust the source gains accordingly, to ensure that tracks are all set to equal perceptual loudness. The impact of this interference on the selected gain parameters and resultant mixture is highlighted.

1 Introduction

Automatic mixing is a growing field. Started by [1], the field began to take off when [2] proposed that mixing multitrack audio can be viewed as a constraint problem, which can be solved. From this point, mixing techniques for video game audio [3] and automatic mixing of musical content [4] have been developed. There are many different approaches to automatic mixing. One of the most common approaches the aim of understanding of the mix process and limits of perception [5, 6], and using this to model the intention of mix engineers [7]. This then allows for the creation of rule based mixing systems [8], which more explicitly state the rules. Alternatively, there are data driven approaches, mapping audio features [9] to mixing decisions [10, 11, 12].

One of the primary focuses of many musical mixing applications is level balancing of multiple sources. [13] proposed setting a gain parameter for audio mixing, such that all tracks are normalised to the same perceived loudness, using gated perceptual loudness model. This approach was then developed by [14], where histogram of loudness methods were used to improve the loudness model captured. [15] utilises the equal loudness curves for calculating a perceptual loudness of audio tracks.[16] instead uses the EBU R-128 recommendation [17] and an exponential moving filter. [18] and [19] both used auditory filters from a masking model and the partial loudness measure and apply loudness normalisation. [20] identifies the importance of perceptual loudness of an automatic mixing system. Both [21] and [22] presented reviews of a range of loudness measures used within automatic mixing approaches. It has been well established that level balancing techniques that focus on normalising all tracks to an equal loudness will produce a perceptually preferable balance.

In practice, previous research relies on the assumption that a loudness balance of a source and a signal are equivalent. In the context of live music recordings, such as that of a drum kit, it is common to have multiple microphones recording simultaneously. This multiple microphone situation can lead to signal interference and thus any automatic mixing approach will need to compensate. [23] included source separation techniques in the field of automatic mixing, to remix early recordings of jazz music, which was developed further through the use of neural networks in [24]. [25] proposed the use of source separation technologies, directly in the use of individual audio effects. Analysis of source separation dataset, and how they can be used to inform intelligent mixing practice was performed in [26].

The field of source separation is a large one, with topics from unwanted background noise removal [27, 28], musical structure analysis [29], to the use in automatic transcription [30] or percussive harmonic separation, and an excellent overview is presented in [31].

The aim of this paper is to present an approach for identifying the level of source interference and acknowledge the impact that identifying this interference can have on the resultant automated sound mixture. An approach to separation and a comparison of gain calculation approaches is presented in Section 3. The dataset used for this work is discussed in Section 4, and the resulting source level identification and gain calculations are presented in Section 5. Section 6 will then present the importance of source separation within a live recording automatic mixing context where there may be source interference.

2 Problem Formulation

A sound mixture can be defined as the sum of the nth microphone $X(n)$, and the calculated gain for that microphone $G(n)$. We define a target loudness for a track L_T, and the calculated loudness of the nth microphone is $L(X(n))$. In the typical approach, an audio mixture Y is produced as N microphones combined.

$$Y = \sum_{n=0}^{N} X(n)G(n) \tag{1}$$

$$G_n = \frac{L_T}{L(X(n))} \tag{2}$$

However, in the case where the source and the signal cannot be view as equivalent, such as the case where there is some source interference, the perceived loudness of source $S(n)$ is given as

$$L(S(n)) = \sum_{n=0}^{N} L(S(n,X(n))) \tag{3}$$

Where $S(n,X(n))$ represents the source $S(n)$ found in $X(n)$.

3 Interfering Source Identification

The Blind Source Separation (BSS) Evaluation Toolbox presents a performance measurement of source separation [32]. A recorded signal is defined as a combination of sources such that

$$X(n) = S(n) + e_{interference} + e_{noise} + e_{artifact} \tag{4}$$

Where e represents some additional signal, either interference, noise or artifact. The signals are decomposed into these four relevant aspects and can be used to approximate a set of source separation performance measures. In this work, we are interested in the Source to Distortion Ratio (SDR)

$$SDR := 10\log_{10}\frac{||S(n)||^2}{||e_{interference} + e_{noise} + e_{artifact}||^2} \tag{5}$$

As such, the SDR can be viewed as the ratio of the source audio signal and all other interfering signals. Based on the assumption that each of the close microphones can be used as an approximation for the individual sources, we can calculate the loudness of each sound source in the overhead microphones, using the SDR measure from the BSS Evaluation toolbox.

4 Dataset

Live drum mic recordings was used for demonstration. Analysis was performed using the ENST dataset [33], an audio-visual drum dataset of multichannel recording of drums. All drum recording, except the single hit recordings, were used from the ENST dataset. This provided 210 different drum recordings, from three different drummers, with an average length of 30s and a standard deviation of 21s. A total of 1.7 hours of continuous recording of 8 channels of audio. Only one third of the dataset (one drummer) used the Tom 3, and as such, all results for Tom 3 are taken over the subset of data where it was used.

5 Results

The ENST dataset [33] was used to analyse the level of each signal source in each of the two overhead microphones. Figure 1 presents a boxplot sum total SDR of each sound source in the two overhead microphones. It is possible for the SDR of a single source to be over

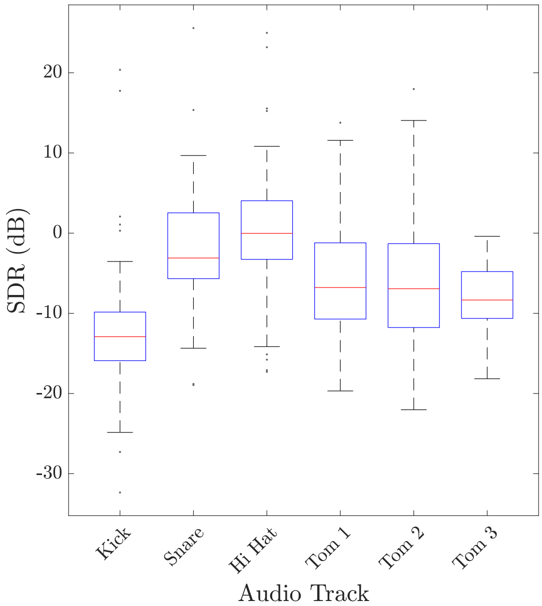

Fig. 1: Box plot of loudness of each source contained within both overhead microphones

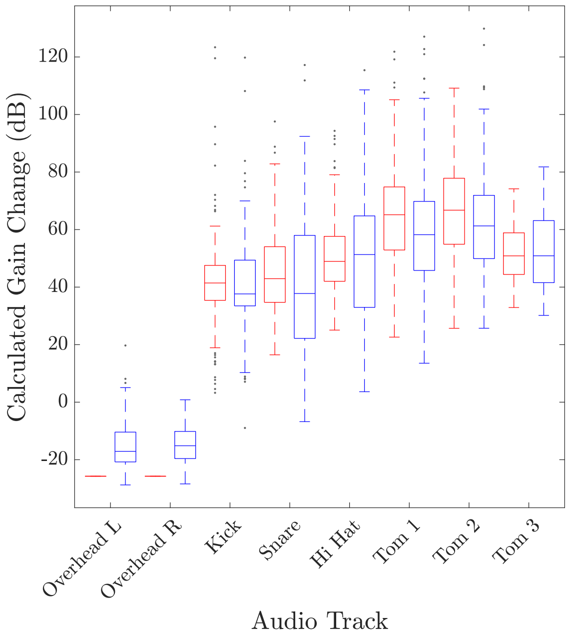

Fig. 2: Box plot of gains calculated between each of the two approaches.
The left item, in red, represents the calculated gains without considering microphone interference. The right box, in blue, represents the calculating gains, compensating for the calculated loudness of each source, using the BSS evaluation toolbox.

0dB, since we are combining the SDR of the two overhead microphones together, to identify the overall loudness of the source in the overhead combination. It is clear from this result, there is a high proportion of each other drum source in the overhead microphones. The impact of this in calculating a gain for an automatic mixing system is presented in Figure 2, and summarised in Figure 3, where the box plots of the calculated gain differences of each of the sources are presented.

It can be seen that the impact on calculated gains is large. The median difference for the two overhead microphones is 8.7dB and 10.6dB. This is an extreme difference in the calculated gains applied to the overhead microphones, and will have a significant change on the balance of the mix.

The impact on each of the other individual sources can be even greater, with the snare and hi hat microphone gains changing the most (11.0dB). Figure 1 shows the snare and hi hat are the loudest sources identified in the overheads. Furthermore, this intuitively makes sense, as the snare and hi hat sounds will commonly be the loudest sources on a drum kit, and will be clearly heard in the overheads, compared with the kick drum, where only a difference of 2.6dB is made.

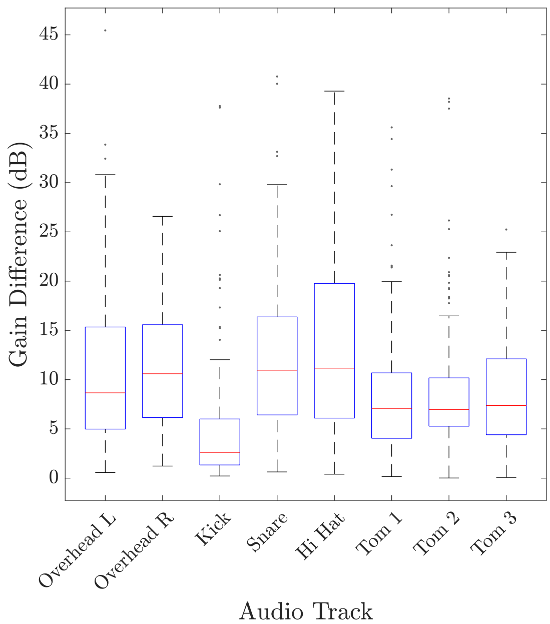

Fig. 3: Difference in gain calculated each source, comparing with and without BSS

6 Conclusion

The existing issues with typical automatic gain balance mixing approaches have been identified. Primarily case where there is natural acoustic signal interference. The use of BSS, to resolve this issue, may not yield perfect results, but will improve the ability to at least identify the level of interference, to allow for the appropriate compensation. It has been shown that, in the case of drums, a gain change of up to 11dB of sources is identified. It is clear that this level of difference will make a significant impact to the perceptual balance of any drum mix

Acknowledgements

This paper is supported by EPSRC Grant EP/L019981/1, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption. Mark B. Sandler acknowledges the support of the Royal Society as a recipient of a Wolfson Research Merit Award.

References

[1] Dugan, D., "Automatic microphone mixing," *Journal of the Audio Engineering Society*, 23(6), pp. 442–449, 1975.

[2] Pachet, F. and Delerue, O., "On-the-fly multi-track mixing," in *Audio Engineering Society Convention 109*, 2000.

[3] Schmidt, B., "Interactive Mixing of Game Audio," in *Audio Engineering Society Convention 115*, 2003.

[4] Perez Gonzalez, E. and Reiss, J. D., "Automatic mixing: live downmixing stereo panner," in *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'07)*, pp. 63–68, 2007.

[5] Bromham, G., Moffat, D., Fazekas, G., Barthet, M., and Sandler, M. B., "The impact of compressor ballistics on the perceived style of music," in *Audio Engineering Society Convention 145*, New York, USA, 2018.

[6] De Man, B., Jillings, N., Moffat, D., Reiss, J. D., and Stables, R., "Subjective comparison of music production practices using the Web Audio Evaluation Tool," in *2nd AES Workshop on Intelligent Music Production (WIMP)*, 2016.

[7] Moffat, D. and Sandler, M. B., "Adaptive Ballistics Control of Dynamic Range Compression for Percussive Tracks," in *Audio Engineering Society Convention 145*, New York, USA, 2018.

[8] Moffat, D., Thalmann, F., and Sandler, M. B., "Towards a Semantic Web Representation and Application of Audio Mixing Rules," in *Proceedings of the 4th Workshop on Intelligent Music Production (WIMP)*, Huddersfield, Uk, 2018.

[9] Moffat, D., Ronan, D., and Reiss, J. D., "An Evaluation of Audio Feature Extraction Toolboxes," in *Proc. 18th International Conference on Digital Audio Effects (DAFx-15)*, 2015.

[10] Martínez Ramírez, M. A. and Reiss, J. D., "Stem audio mixing as a content-based transformation of audio features," in *IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6, 2017.

[11] Martínez Ramírez, M. A. and Reiss, J. D., "Deep Learning and Intelligent Audio Mixing," in *Proceedings of the 3rd Workshop on Intelligent Music Production (WIMP)*, 2017.

[12] Ronan, D., Moffat, D., Gunes, H., and Reiss, J. D., "Automatic Subgrouping of Multitrack Audio," in *Proc. 18th International Conference on Digital Audio Effects (DAFx-15)*, 2015.

[13] Perez Gonzalez, E. and Reiss, J. D., "An automatic maximum gain normalization technique with applications to audio mixing." in *Audio Engineering Society Convention 124*, 2008.

[14] Perez Gonzalez, E. and Reiss, J. D., "Automatic gain and fader control for live mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'09)*, pp. 1–4, 2009.

[15] Scott, J. J. and Kim, Y. E., "Instrument Identification Informed Multi-Track Mixing." in *Proc. 14th International Society for Music Information Retrieval Conference (ISMIR13)*, pp. 305–310, 2013.

[16] Mansbridge, S., Finn, S., and Reiss, J. D., "Implementation and evaluation of autonomous multitrack fader control," in *Audio Engineering Society Convention 132*, 2012.

[17] ITU-R, "BS. 1770-2, Algorithms to measure audio programme loudness and true-peak audio level," Technical report, International Telecommunications Union, Geneva, Switzerland, 2011.

[18] Ma, Z., Reiss, J. D., and Black, D. A., "Partial loudness in multitrack mixing," in *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*, 2014.

[19] Terrell, M., Simpson, A., and Sandler, M., "The mathematics of mixing," *Journal of the Audio Engineering Society*, 62(1/2), pp. 4–13, 2014.

[20] Ward, D., Reiss, J. D., and Athwal, C., "Multitrack mixing using a model of loudness and partial loudness," in *Audio Engineering Society Convention 133*, 2012.

[21] Wichern, G., Wishnick, A., Lukin, A., and Robertson, H., "Comparison of loudness features for automatic level adjustment in mixing," in *Audio Engineering Society Convention 139*, 2015.

[22] Ward, D. and Reiss, J. D., "Loudness algorithms for automatic mixing," in *Proceedings of the 2nd AES Workshop on Intelligent Music Production (WIMP)*, 2016.

[23] Matz, D., Cano, E., and Abeßer, J., "New Sonorities for Early Jazz Recordings Using Sound Source Separation and Automatic Mixing Tools." in *Proc. 16th International Society for Music Information Retrieval Conference (ISMIR-15)*, pp. 749–755, 2015.

[24] Mimilakis, S. I., Cano, E., Abeßer, J., and Schuller, G., "New sonorities for jazz recordings: Separation and mixing using deep neural networks," in *2nd AES Workshop on Intelligent Music Production (WIMP)*, 2016.

[25] Pardo, B., Rafii, Z., and Duan, Z., "Audio Source Separation in a Musical Context," in *Springer Handbook of Systematic Musicology*, pp. 285–298, Springer, 2018.

[26] Ward, D., Wierstorf, H., Mason, R., Plumbley, M., and Hummersone, C., "Estimating the loudness balance of musical mixtures using audio source separation," in *Proceedings of the 3rd Workshop on Intelligent Music Production (WIMP)*, 2017.

[27] Fano Yela, D., Ewert, S., Fitzgerald, D., and Sandler, M., "Interference reduction in music recordings combining kernel additive modelling and non-negative matrix factorization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 51–55, 2017.

[28] Moffat, D. and Reiss, J. D., "Dereverberation and its application to the blind source separation problem," in *Proc. Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*, 2016.

[29] Quinton, E., O'Hanlon, K., Dixon, S., and Sandler, M., "Tracking metrical structure changes with sparse-NMF," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 41–45, 2017.

[30] Dittmar, C. and Gärtner, D., "Real-Time Transcription and Separation of Drum Recordings Based on NMF Decomposition." in *Proc. 17th International Conference on Digital Audio Effects (DAFx-14)*, pp. 187–194, 2014.

[31] Cano, E., Fitzgerald, D., Liutkus, A., Plumbley, M. D., and Stoter, F.-R., "Musical Source Separation: An Introduction," *IEEE Signal Processing Magazine*, 36(1), pp. 31–40, 2019.

[32] Vincent, E., Gribonval, R., and Févotte, C., "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, 14(4), pp. 1462–1469, 2006.

[33] Gillet, O. and Richard, G., "ENST-Drums: an extensive audio-visual database for drum signals processing." in *Proc. of the 7th International Society for Music Information Retrieval Conference (ISMIR06)*, pp. 156–159, 2006.