AN ANALYSIS AND EVALUATION OF AUDIO FEATURES FOR MULTITRACK MUSIC MIXTURES

Brecht De Man[1], Brett Leonard[2, 3], Richard King[2, 3], Joshua D. Reiss[1]
[1]Centre for Digital Music, Queen Mary University of London
[2]The Graduate Program in Sound Recording, Schulich School of Music, McGill University
[3]Centre for Interdisciplinary Research in Music Media and Technology
b.deman@qmul.ac.uk, brett.leonard@mail.mcgill.ca,
richard.king@mcgill.ca, joshua.reiss@qmul.ac.uk

ABSTRACT

Mixing multitrack music is an expert task where characteristics of the individual elements and their sum are manipulated in terms of balance, timbre and positioning, to resolve technical issues and to meet the creative vision of the artist or engineer. In this paper we conduct a mixing experiment where eight songs are each mixed by eight different engineers. We consider a range of features describing the dynamic, spatial and spectral characteristics of each track, and perform a multidimensional analysis of variance to assess whether the instrument, song and/or engineer is the determining factor that explains the resulting variance, trend, or consistency in mixing methodology. A number of assumed mixing rules from literature are discussed in the light of this data, and implications regarding the automation of various mixing processes are explored. Part of the data used in this work is published in a new online multitrack dataset through which public domain recordings, mixes, and mix settings (DAW projects) can be shared.

1. INTRODUCTION

The production of recorded music involves a range of expert signal processing techniques applied to recorded musical material. Each instrument or element thereof exists on a separate audio 'track', and this process of manipulating and combining these tracks is normally referred to as mixing. Strictly creative processes aside, each process can generally be classified as manipulating the dynamic (balance and dynamic range compression), spatial (stereo or surround panning and reverberation), and spectral (equalisation) features of the source material, or a combination thereof [1, 4, 8, 15].

Recent years have seen a steep increase in research on automatic mixing, where some of the tedious, routine tasks in audio production are automated to the benefit of the inexperienced amateur or the time constrained professional. Most research is concerned with the validation of a mixing rule based on knowledge derived from practical literature or expert interviews [2, 6, 7, 9], usually through an experiment where a method based on this assumption is compared to a set of alternative methods. Furthermore, some research has been done on machine learning systems for balancing and panning of tracks [13]. In spite of these efforts, the relation between the characteristics of the source material and the chosen processing parameters, as well as the importance of subjective input of the individual versus objective or generally accepted target features, is still poorly understood. Recurring challenges in this field include a lack of research data, such as high-quality mixes in a realistic but sufficiently controlled setting, and tackling the inherently high cross-adaptivity of the mixing problem, as the value of each processing parameter for any given track is usually dependent on features and chosen processing parameters associated with other tracks as well.

In this work, we conduct an experiment where a group of mixing engineers mix the same material in a realistic setting, with relatively few constraints, and analyse the manipulation of the signals and their features. We test the validity of the signal-dependent, instrument-independent model that is often used in automatic mixing research [6, 7], and try to identify which types of processing are largely dependent on instrument type, the song (or source material), or the individual mixing engineer. Consequently, we also identify which types of processing are not clearly defined as a function of these parameters, and thus warrant further research to understand their relation to low-level (readily extracted) features or high-level properties (instrument, genre, desired effect) of the source audio. We discuss the relevance of a number of audio features for the assessment of music production and the underlying processes as described above. This experiment also provides an opportunity to validate some of the most common assumptions in autonomous mixing research.

2. EXPERIMENT

The mixing engineers in this experiment were students of the MMus in Sound Recording at the Schulich School of Music at McGill University. They were divided in two groups of eight, where each group corresponds with a class from a different year in the two-year programme, and each group was assigned a different set of four songs to mix. Each mixing engineer allocated up to 6 hours to each of their four mix assignments, and was allowed to use Avid's Pro Tools including built-in effects (with automation) and the Lexicon PCM Native Reverb Plug-In Bundle, a set of tools they were familiar with.

Four out of eight songs are available on a new multitrack testbed including raw tracks, the rendered mixes and the complete Pro Tools project files, allowing others to reproduce or extend the research. The testbed can be found on `c4dm.eecs.qmul.ac.uk/multitrack`. The authors welcome all appropriately licensed contributions consisting of shareable raw, multitrack audio, DAW project files, rendered mixes, or a subset thereof. Due to copyright restrictions, the other songs could not be shared.

We consider three types of instruments - drums, bass, and lead vocal - as they are featured in all test songs in this research, and as they are common elements in contemporary music in general. Furthermore, we split up the drums in the elements kick drum, snare drum, and 'rest'. Three out of eight songs had a male lead vocalist, and half of the songs featured a double bass (in one case part bowed) while the other half had a bass guitar for the bass part.

For the purpose of this investigation, we consider a fragment of the song only, consisting of the second verse and chorus, as all considered sources (drums, bass and lead vocal) are active here.

Whereas the audio was recorded and mixed at a sampling ratio of 96 kHz, we converted all audio to 44.1 kHz to reduce computational cost and to calculate spectral features based on the mostly audible region. The processed tracks are rendered from the digital audio workstation with all other tracks inactive, but with an unchanged signal path including send effects and bus processing [1].

3. FEATURES

The set of features we consider (Table 1) has been tailored to reflect properties relevant to the production of music in the dynamic, spatial and spectral domain. We consider the mean of the feature over all frames of a track fragment.

We use the perceptually informed measure of loudness relative to the loudness of the mix, as a simple RMS level can be strongly influenced by high energy at frequencies the human ear is not very sensitive to. To accurately measure loudness in the context of multitrack content, we use the highest performing modification in [12] (i.e. using a time constant of 280 ms and a pre filter gain of +10 dB) on the most recent ITU standard on measuring audio programme loudness [3].

Category	Feature	Reference
Dynamic	Loudness	[3, 12]
	Crest factor (100 ms and 1 s)	[17]
	Activity	[7]
Spatial	SPS	[16]
	$P_{[band]}$	[16]
	Side/mid ratio	
	Left/right imbalance	
Spectral	Centroid	[5]
	Brightness	
	Spread	
	Skewness	
	Kurtosis	·
	Rolloff (.95 and .85)	·
	Entropy	
	Flatness	
	Roughness	
	Irregularity	
	Zero-crossing rate	
	Low energy	[5]
	Octave band energies	

Table 1: List of extracted features

To reflect the properties of the signal related to dynamic range on the short term, we calculate the crest factor over a window of 100 ms and over a window of 1 s [17].

To quantify gating, muting, and other effects that make the track (in)audible during processing, we measure the percentage of time the track is active, with the activity state indicated by a Schmitt trigger with thresholds at -25 and -30 dB LUFS [7].

To analyse the spatial processing, we use the Stereo Panning Spectrum (SPS), which shows the spatial position of a certain frequency bin in function of time, and the Panning Root Mean Square ($P_{[band]}$), the RMS of the SPS over a number of frequency bins [16]. In this work, we use the absolute value of SPS, averaged over time, and the standard P_{total} (all bins), P_{low} (0-250 Hz), P_{mid} (250-2500 Hz) and P_{high} (2500-22050 Hz), also averaged over time. Furthermore, we propose a simple stereo width measure, the side/mid ratio, calculated as the power of side channel (sum of left and right channel) over the power of the mid channel (average of left channel and polarity-reversed right channel). We also define the left/right imbalance, as $(R - L)/(R + L)$ where L is the total/average power of the left channel, and R is the total/average power of the right channel. A centred track has low imbalance and low side/mid ratio, while a hard panned track has high imbalance and high side/mid ratio. Note that while these features are related, they do not mean the same thing. A source could have uncorrelated signals with the exact same energy in the left and right channel respectively, which would lead to a low left/right imbalance and a high side/mid ratio.

Finally, we use features included in the MIR Toolbox [5] (with the default 50 ms window length) as well as octave band energies to describe the spectral characteristics of the audio.

[1] When disabling the other tracks, non-linear processes on groups of tracks (such as bus dynamic range compression) will result in a different effective effect on the rendered track since the processor may be triggered differently (such as a reduced trigger level). While for the purpose of this experiment, the difference in triggering of bus compression does not affect the considered features significantly, it should be noted that for rigorous extraction of processed tracks, in such a manner that when summed together they result in the final mix, the true, time-varying bus compression gain should be measured and applied on the single tracks.

4. ANALYSIS AND DISCUSSION

4.1 Analysis of variance

Table 2 shows the mean values of the features, as well as the standard deviation between different mixing engineers and the standard deviation between different songs. Most considered features show greater variance for the same engineer across different songs, than for the same song over different engineers. Exceptions to this are the left/right imbalance and spectral roughness, which on average appear to be more dependent on the engineer than on the source content. The change of features (difference before and after processing, where applicable) varies more for different mixing engineers than for different songs, too, for all features. However, when considering the features instrument by instrument, the source material only rarely causes the means of the feature to differ significantly (the means are only significantly different through the effect of source material for the zero-crossing rate of the snare drum track, and for the spectral entropy of the vocal track). This suggests that engineers would disagree on processing values, whereas the source material has less effect.

For each feature, we perform an analysis of variance to investigate for which feature we can reject the hypothesis that the different 'treatments' (different source material, mixing engineer or instrument) result in the same feature value. For those features for which there is a significant effect ($p < 0.05$), we perform a multiple comparison of population means using the Bonferroni correction to establish what the mean values of the feature are as a function of the determining factor, and which instruments or songs have a significantly lower or higher mean than others. We discuss the outcome of these tests in the following paragraphs.

As some elements were not used by the mixing engineer, some missing values are dropped when calculating the statistics in the following sections.

4.2 Balance and dynamics processing

In general, the relative loudness of tracks, averaged over all instruments, is dependent on the song ($p < 5 \cdot 10^{-11}$). However, when looking at each instrument individually, the relative loudness of the bass guitar ($p < 0.01$), snare drum ($p < 0.05$) and other drum instruments ('rest', i.e. not snare or kick drum, $p < 5 \cdot 10^{-4}$) is dependent on mixing engineer.

In automatic mixing research, a popular assumption is that the loudness of the different tracks or sources should be equal [7]. A possible exception to this is the main element, usually the vocal, which can be set at a higher loudness [1]. From Figure 1, it is apparent that the vocal is significantly louder than the other elements considered here, whereas no significant difference of the mean relative loudness of the other elements can be shown. Furthermore, the relative loudness of the vocal shows a relative narrow range of values (-2.7 ± 1.6 LU), suggesting an agreement on a 'target loudness' of about -3 LU relative to the overall mix loudness.

It should be noted that due to crosstalk between the drum microphones, the effective loudness of the snare drum and kick drum will differ from the loudness measured from the snare drum and kick drum tracks. As a result, disagreement of the relative loudnesses of snare drum and other drum elements such as overhead and room microphones does not necessarily suggest a significantly different desired loudness of the snare drum, as the snare drum is present in both of these tracks. In this work, however, we are interested in the manipulations of the different tracks as they are available to the engineer.

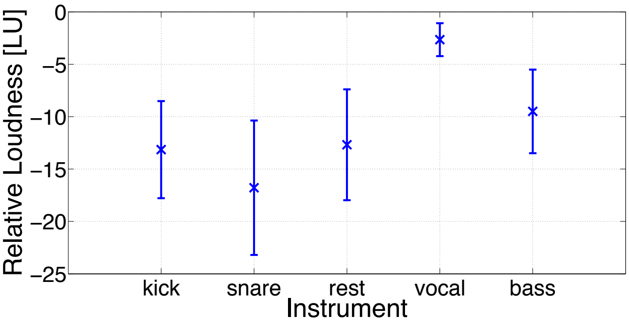


Figure 1: Average and standard deviation of loudness of sources relative to the total loudness of the mix, across songs and mixing engineers.

The crest factor is affected by both the instrument ($p < 5 \cdot 10^{-3}$) and song ($p < 10^{-20}$), and every instrument individually shows significantly different crest factor values for different engineers ($p < 5 \cdot 10^{-3}$). One exception to the latter is the kick drum for a crest factor window size of 1 s, where the hypothesis was not disproved for one group of engineers.

All instruments show an increase in crest factor compared to the raw values (ratio significantly greater than one). This means that the short-term dynamic range is effectively expanded, which can be an effect of dynamic range compression as transients are left unattenuated due to the response time of the compressor, while the rest of the signal is reduced in level.

The percentage of the time the track was active did not meaningfully change under the influence of different source material, individual mixing engineers or instruments. A drop in activity in some instances is due to gating of kick drum, but this is the decision of certain mixing engineers for certain songs, and no consistent trend.

4.3 Stereo panning

Both the average left/right imbalance and average side/mid ratio were significantly higher for the non-pop/rock songs ($p < 10^{-6}$).

The Panning Root Mean Square values $P_{[band]}$ all show a larger value for the total mix and for the 'rest' group. The difference is significant except for the lowest band, where only the bass is significantly more central than the total mix. This can be explained by noting that most of the low frequency sources are panned centre (see further).

In literature on automatic mixing and mixing engineering textbooks, it is stated that low-frequency sources as well as lead vocals and snare drums should be panned central [1, 2, 4, 6, 8–10, 14]. To quantify the spatialisation for different frequencies, we display the panning as a function

Feature	**Kick drum**	**Snare drum**	**Rest drums**	**Bass**	**Lead vocal**	**Average**	**Mix**
Loudness [LU]	$-13.15 \pm^{4.05}_{3.89}$	$-16.78 \pm^{6.17}_{4.57}$	$-12.68 \pm^{5.46}_{2.80}$	$-2.65 \pm^{1.52}_{1.31}$	$-9.50 \pm^{3.51}_{2.86}$	$-10.95 \pm^{4.14}_{3.09}$	N/A
Crest (100 ms)	$3.599 \pm^{0.603}_{0.330}$	$4.968 \pm^{0.998}_{0.469}$	$4.510 \pm^{1.065}_{0.354}$	$2.565 \pm^{0.443}_{0.166}$	$3.315 \pm^{0.403}_{0.208}$	$3.791 \pm^{0.634}_{0.274}$	$3.332 \pm^{0.294}_{0.116}$
Crest (1 s)	$9.824 \pm^{3.074}_{1.911}$	$16.724 \pm^{6.458}_{3.135}$	$12.472 \pm^{4.710}_{1.823}$	$4.339 \pm^{1.098}_{0.449}$	$5.283 \pm^{1.102}_{0.514}$	$9.728 \pm^{2.907}_{1.398}$	$5.315 \pm^{0.997}_{0.554}$
Activity	$0.676 \pm^{0.250}_{0.122}$	$0.861 \pm^{0.161}_{0.078}$	$0.909 \pm^{0.115}_{0.029}$	$0.958 \pm^{0.076}_{0.009}$	$0.844 \pm^{0.089}_{0.044}$	$0.850 \pm^{0.117}_{0.048}$	$0.995 \pm^{0.009}_{0.004}$
L/R imbalance	$\mathbf{0.075} \pm^{\mathbf{0.094}}_{\mathbf{0.137}}$	$\mathbf{0.144} \pm^{\mathbf{0.153}}_{\mathbf{0.227}}$	$0.361 \pm^{0.303}_{0.213}$	$\mathbf{0.107} \pm^{\mathbf{0.135}}_{\mathbf{0.176}}$	$\mathbf{0.045} \pm^{\mathbf{0.072}}_{\mathbf{0.085}}$	$\mathbf{0.146} \pm^{\mathbf{0.139}}_{\mathbf{0.152}}$	$0.088 \pm^{0.075}_{0.074}$
Side/mid ratio	$\mathbf{0.036} \pm^{\mathbf{0.055}}_{\mathbf{0.076}}$	$\mathbf{0.036} \pm^{\mathbf{0.040}}_{\mathbf{0.043}}$	$0.242 \pm^{0.183}_{0.154}$	$\mathbf{0.009} \pm^{\mathbf{0.013}}_{\mathbf{0.015}}$	$\mathbf{0.022} \pm^{\mathbf{0.018}}_{\mathbf{0.022}}$	$0.069 \pm^{0.060}_{0.059}$	$0.101 \pm^{0.049}_{0.046}$
P_{total}	$0.104 \pm^{0.102}_{0.090}$	$0.108 \pm^{0.082}_{0.059}$	$0.307 \pm^{0.028}_{0.027}$	$0.075 \pm^{0.093}_{0.083}$	$\mathbf{0.134} \pm^{\mathbf{0.022}}_{\mathbf{0.027}}$	$0.145 \pm^{0.060}_{0.052}$	$0.234 \pm^{0.030}_{0.027}$
P_{low}	$\mathbf{0.066} \pm^{\mathbf{0.078}}_{\mathbf{0.087}}$	$0.122 \pm^{0.102}_{0.073}$	$0.243 \pm^{0.045}_{0.041}$	$0.040 \pm^{0.063}_{0.059}$	$\mathbf{0.147} \pm^{\mathbf{0.034}}_{\mathbf{0.042}}$	$0.123 \pm^{0.061}_{0.056}$	$0.188 \pm^{0.042}_{0.034}$
P_{mid}	$\mathbf{0.066} \pm^{\mathbf{0.074}}_{\mathbf{0.076}}$	$0.114 \pm^{0.090}_{0.064}$	$0.290 \pm^{0.023}_{0.023}$	$0.052 \pm^{0.082}_{0.067}$	$\mathbf{0.177} \pm^{\mathbf{0.027}}_{\mathbf{0.035}}$	$0.140 \pm^{0.054}_{0.048}$	$0.248 \pm^{0.027}_{0.023}$
P_{high}	$0.106 \pm^{0.104}_{0.091}$	$0.105 \pm^{0.081}_{0.058}$	$0.309 \pm^{0.029}_{0.028}$	$0.076 \pm^{0.094}_{0.085}$	$\mathbf{0.124} \pm^{\mathbf{0.022}}_{\mathbf{0.028}}$	$0.144 \pm^{0.061}_{0.053}$	$0.231 \pm^{0.033}_{0.029}$
Centroid [Hz]	$2253.8 \pm^{1065.6}_{729.8}$	$4395.3 \pm^{1448.6}_{554.2}$	$4130.8 \pm^{1228.1}_{483.2}$	$1046.5 \pm^{520.1}_{232.4}$	$2920.2 \pm^{452.1}_{264.7}$	$2949.3 \pm^{872.1}_{418.6}$	$2478.8 \pm^{517.9}_{247.1}$
Brightness	$0.306 \pm^{0.105}_{0.103}$	$0.598 \pm^{0.156}_{0.069}$	$0.557 \pm^{0.115}_{0.058}$	$0.135 \pm^{0.082}_{0.031}$	$0.455 \pm^{0.071}_{0.040}$	$0.410 \pm^{0.100}_{0.056}$	$0.362 \pm^{0.070}_{0.034}$
Spread	$3250.1 \pm^{783.2}_{447.5}$	$4363.6 \pm^{701.9}_{335.9}$	$4422.1 \pm^{734.6}_{292.3}$	$2426.6 \pm^{559.2}_{320.4}$	$3369.9 \pm^{324.6}_{191.3}$	$3566.5 \pm^{587.5}_{298.0}$	$3453.2 \pm^{421.7}_{200.6}$
Skewness	$3.649 \pm^{1.068}_{0.886}$	$1.492 \pm^{0.663}_{0.301}$	$1.665 \pm^{0.682}_{0.246}$	$6.234 \pm^{1.885}_{0.630}$	$2.470 \pm^{0.573}_{0.243}$	$3.102 \pm^{0.912}_{0.427}$	$2.779 \pm^{0.600}_{0.257}$
Kurtosis	$23.847 \pm^{11.997}_{9.164}$	$5.965 \pm^{2.905}_{1.474}$	$7.053 \pm^{3.449}_{1.263}$	$58.870 \pm^{31.874}_{11.107}$	$11.579 \pm^{4.267}_{1.784}$	$21.463 \pm^{9.834}_{4.477}$	$13.646 \pm^{4.511}_{2.073}$
Rolloff .95 [Hz]	$8880.1 \pm^{3679.2}_{2151.2}$	$13450.9 \pm^{3100.6}_{1582.2}$	$13373.4 \pm^{2594.1}_{1007.4}$	$4389.4 \pm^{2714.7}_{1244.5}$	$9879.0 \pm^{1335.7}_{725.3}$	$9994.5 \pm^{2498.0}_{1240.8}$	$9679.0 \pm^{1563.8}_{734.3}$
Rolloff .85 [Hz]	$4513.7 \pm^{2736.6}_{1788.8}$	$8984.3 \pm^{3139.7}_{1348.5}$	$8755.3 \pm^{2742.5}_{975.6}$	$1625.5 \pm^{1205.0}_{594.3}$	$5595.8 \pm^{1121.4}_{609.7}$	$5894.9 \pm^{2047.2}_{986.1}$	$5026.2 \pm^{1337.8}_{599.8}$
Entropy	$0.655 \pm^{0.104}_{0.090}$	$0.840 \pm^{0.084}_{0.057}$	$0.832 \pm^{0.051}_{0.025}$	$0.552 \pm^{0.073}_{0.026}$	$0.735 \pm^{0.043}_{0.016}$	$0.723 \pm^{0.066}_{0.038}$	$0.744 \pm^{0.043}_{0.015}$
Flatness	$0.148 \pm^{0.072}_{0.051}$	$0.350 \pm^{0.142}_{0.056}$	$0.337 \pm^{0.118}_{0.045}$	$0.073 \pm^{0.035}_{0.020}$	$0.167 \pm^{0.030}_{0.018}$	$0.215 \pm^{0.074}_{0.035}$	$0.174 \pm^{0.046}_{0.020}$
Roughness	$\mathbf{84.72} \pm^{\mathbf{84.85}}_{\mathbf{98.32}}$	$\mathbf{36.30} \pm^{\mathbf{41.16}}_{\mathbf{43.32}}$	$67.57 \pm^{71.76}_{46.28}$	$\mathbf{236.04} \pm^{\mathbf{160.38}}_{\mathbf{176.05}}$	$\mathbf{247.00} \pm^{\mathbf{216.15}}_{\mathbf{247.36}}$	$\mathbf{134.33} \pm^{\mathbf{319.30}}_{\mathbf{338.44}}$	$\mathbf{1843.31} \pm^{\mathbf{1341.50}}_{\mathbf{1419.35}}$
Irregularity	$0.158 \pm^{0.098}_{0.063}$	$0.235 \pm^{0.151}_{0.079}$	$0.297 \pm^{0.135}_{0.069}$	$0.502 \pm^{0.176}_{0.065}$	$0.540 \pm^{0.165}_{0.094}$	$0.346 \pm^{0.136}_{0.075}$	$0.705 \pm^{0.090}_{0.078}$
Zero-crossing	$584.7 \pm^{509.5}_{409.4}$	$2222.0 \pm^{1183.3}_{604.7}$	$1988.9 \pm^{944.1}_{466.1}$	$246.6 \pm^{217.8}_{89.6}$	$1177.5 \pm^{233.7}_{143.6}$	$1243.9 \pm^{554.3}_{305.4}$	$905.2 \pm^{237.4}_{118.8}$
Low energy	$0.752 \pm^{0.113}_{0.081}$	$0.723 \pm^{0.084}_{0.055}$	$0.682 \pm^{0.047}_{0.034}$	$0.507 \pm^{0.096}_{0.033}$	$0.544 \pm^{0.065}_{0.048}$	$0.641 \pm^{0.073}_{0.048}$	$\mathbf{0.541} \pm^{\mathbf{0.035}}_{\mathbf{0.038}}$

Table 2: Average values of features per instrument, including average over instrument and value of total mix, with standard deviation between different songs by the same mixing engineer (top), and between different mixes of the same song (bottom). Values for which the variation across different mixes for the same song is greater than the variation across different songs for the same engineer are displayed in bold.

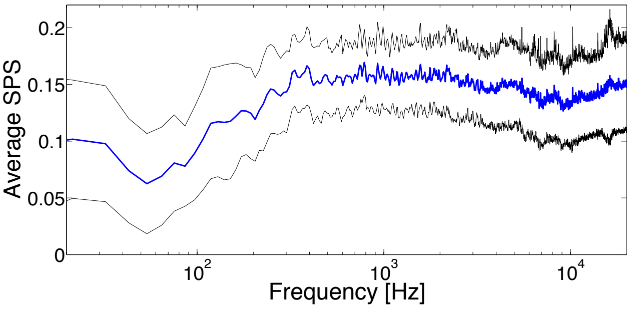


Figure 2: Mean Stereo Panning Spectrum (with standard deviation) over all mixes and songs

of frequency in Figure 2, using the average Stereo Panning Spectrum over all mixes and songs. From this figure a clear increase in SPS with increasing frequency is apparent between 50 Hz and 400 Hz. However, this trend does not extend to the very low frequencies (20-50 Hz) or higher frequencies (>400 Hz).

4.4 Equalisation

To assess the spectral processing of sources, mostly equalisation in this context, we consider both the absolute values of the spectral features (showing the desired features of the processed audio) as well as the change in features (showing common manipulations of the tracks). When only taking the manipulations into account, and not the features of the source audio, similar to blindly applying a software equaliser's presets, the results would be less translatable to situations where the source material's spectral characteristics differs from that featured in this work [2]. However, considering the change in features could reveal common practices that are less dependent on the features of the source material. Therefore, we investigate both.

The spectral centroid of the whole mix varies strongly depending on the mixing engineer ($p < 5 \cdot 10^{-6}$). The centroid of the snare drum track is consistently increased through processing, due to a reduction of the low energy content as well as spill of instruments like kick drum (see further regarding the reduction of low energy) and/or the emphasis of a frequency range above the original centroid.

The brightness of each track except bass guitar and kick drum (the sources with the highest amount of low energy) is increased.

For a large set of spectral features (spectral centroid, brightness, skewness, roll-off, flatness, zero-crossing, and roughness), the engineers disagree on the preferred value for all instruments except kick drum. In other words, the values describing the spectrum of a kick drum across engineers are overlapping, implying a consistent spectral target (a certain range of appropriate values). For other features (spread, kurtosis and irregularity) the value corresponding with the kick drum track is also significantly different across engineers. The roughness shows no significantly different means for any instrument except the 'rest' bus.

The low energy of each track is reduced for each instrument, with significantly more reduction for snare drum than for kick drum and bass guitar. Its absolute value for bass and vocal is significantly different across engineers, whereas there is a general overlap for all other instruments including the mix. As the variation in the resulting value of low energy is higher than the variation for the unprocessed versions, no target value is apparent for any instrument, nor for the total mix.

Analysis of the octave band energies reveals definite trends across songs and mixing engineers, for a certain instrument as well as the mix. The standard deviation does not consistently decrease or increase over the octave bands for any instrument when compared to the raw audio. The suggested 'mix target spectrum' is in agreement with [11], which derived a 'target spectrum' based on average spectra of number one hits from various genres and over several decades. Figure 4 shows the measured average mix spectrum against the octave band values of the average spectrum of a number one hit after 2000 from that work, which lies within a standard deviation from our result with the exception of the highest band. The average relative change in energies is not significantly different from zero (no bands are consistently boosted or cut for certain instruments), but taking each song individually in consideration, a strong agreement of reasonably drastic boosts or cuts is shown for some songs. This confirms that the equalisation is highly dependent on the source material, and engineers largely agree on the necessary treatment for source tracks showing spectral anomalies.

5. CONCLUSION

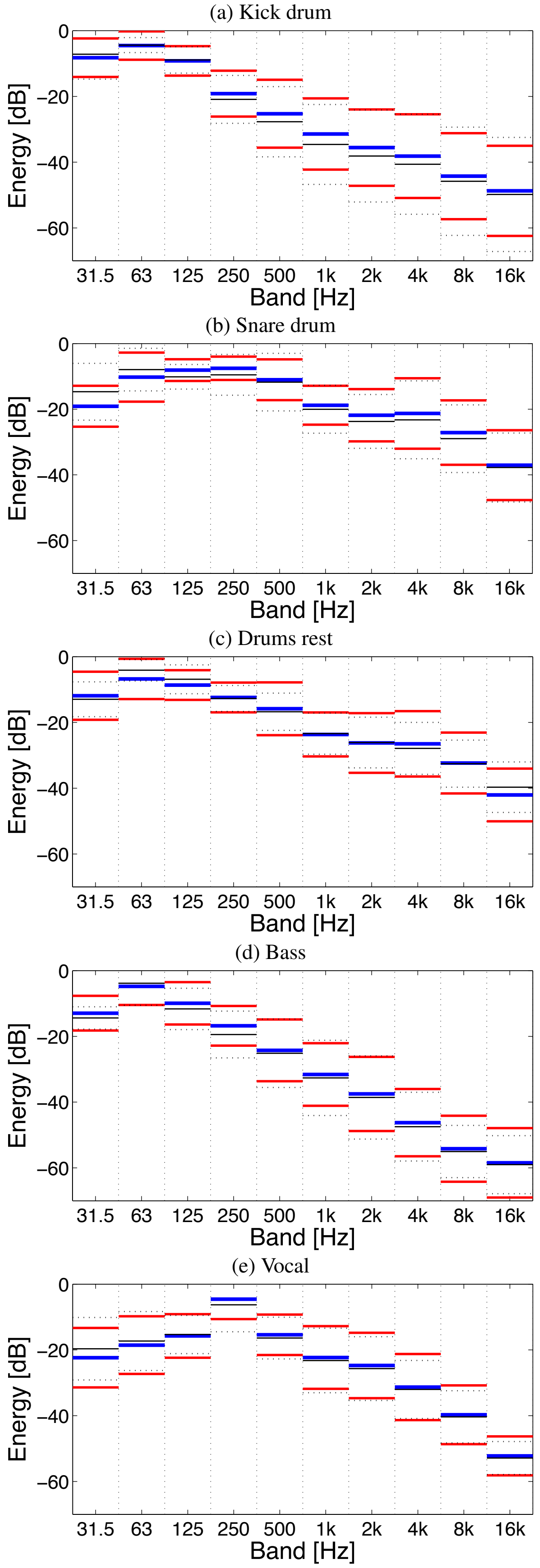
We conducted a controlled experiment where eight multitrack recordings mixed by eight mixing engineers were analysed in terms of dynamic, spatial and spectral processing of common key elements.

We measured a greater variance of features across songs than across engineers, for each considered instrument and for the total mix, whereas the mean values corresponding to the different engineers were more often statistically different from each other.

The relative loudness of the lead vocal track was found to be significantly louder than all other tracks, with an average value of -3 LU relative to the total mix loudness.

The amount of panning as a function of frequency was investigated, and found to be increasing with frequency up to about 400 Hz, above which it stays more or less constant.

We measured a consistent decrease of low frequency energy and an increase of crest factor for all instruments, and an increase of the spectral centroid of the snare drum track. Spectral analysis has shown a definite target spectrum that agrees with the average spectrum of recent commercial recordings.


(a) Kick drum

(b) Snare drum

(c) Drums rest

(d) Bass

(e) Vocal

Figure 3: Average octave band energies (blue) with standard deviation (red) for different instruments after processing, compared to the raw signal (black).

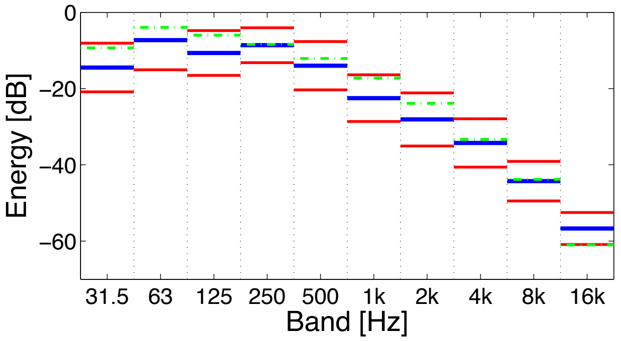


Figure 4: Average octave band energies for total mix, compared to ‘After 2000’ curve from [11] (green dashed line)

6. FUTURE WORK

Future work will be concerned with perceptual evaluation of mixes and its relation to features, using both qualitative (‘which sonic descriptors correspond with which features?’) and quantitative analysis (‘which manipulation of audio is preferred?’).

Further research is needed to establish the desired loudness of sources, as opposed to loudness of tracks, and its variance throughout songs, genres, and mixing engineers.

An extrapolation of the analysis described in this paper to other instruments is needed to validate the generality of the conclusions regarding the processing of drums, bass and lead vocal at the mixing stage, and to further explore laws underpinning the processing of different instruments.

Based on the findings of this work, which showed trends and variances of different relevant features, we can inform knowledge engineered or machine learning based systems that automate certain mixing tasks (balancing, panning, equalising and compression).

This work was based on a still relatively limited set of mixes, for which the engineers came from the same institution. Through initiatives such as the public multitrack testbed presented in this paper, it will be possible to analyse larger corpora of mixes, where more parameters can be investigated with more significance.

7. ACKNOWLEDGEMENTS

The authors would like to thank George Fazekas for assistance with launching the multitrack testbed.

This study was funded in part by the Engineering and Physical Sciences Research Council (EPSRC) Semantic Media grant (EP/J010375/1).

8. REFERENCES

[1] Alex Case. *Mix Smart: Professional Techniques for the Home Studio*. Focal Press. Taylor & Francis, 2011.

[2] Brecht De Man and Joshua D. Reiss. A knowledge-engineered autonomous mixing system. In *135th Convention of the Audio Engineering Society*, 2013.

[3] ITU. Recommendation ITU-R BS.1770-3 Algorithms to measure audio programme loudness and true-peak audio level. Technical report, Radiocommunication Sector of the International Telecommunication Union, 2012.

[4] Roey Izhaki. *Mixing audio: concepts, practices and tools*. Focal Press, 2008.

[5] Olivier Lartillot and Petri Toiviainen. MIR in Matlab (II): A toolbox for musical feature extraction from audio. In *Proceedings of the 8th International Society for Music Information Retrieval Conference*, 2007.

[6] Stuart Mansbridge, Saoirse Finn, and Joshua D. Reiss. An autonomous system for multi-track stereo pan positioning. In *133rd Convention of the Audio Engineering Society*, 2012.

[7] Stuart Mansbridge, Saoirse Finn, and Joshua D. Reiss. Implementation and evaluation of autonomous multi-track fader control. In *132nd Convention of the Audio Engineering Society*, 2012.

[8] Bobby Owsinski. *The Mixing Engineer’s Handbook*. Course Technology, 2nd edition, 2006.

[9] Enrique Perez-Gonzalez and Joshua D. Reiss. Automatic mixing: Live downmixing stereo panner. In *10th International Conference on Digital Audio Effects (DAFx-10)*, 2007.

[10] Pedro Pestana. *Automatic mixing systems using adaptive digital audio effects*. PhD thesis, Catholic University of Portugal, 2013.

[11] Pedro Duarte Pestana, Zheng Ma, Joshua D. Reiss, Alvaro Barbosa, and Dawn A. A. Black. Spectral characteristics of popular commercial recordings 1950-2010. In *135th Convention of the Audio Engineering Society*, 2013.

[12] Pedro Duarte Pestana, Joshua D. Reiss, and Alvaro Barbosa. Loudness measurement of multitrack audio content using modifications of ITU-R BS.1770. In *Audio Engineering Society Convention 134*, 2013.

[13] Jeff Scott and Youngmoo E. Kim. Analysis of acoustic features for automated multi-track mixing. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011.

[14] Jeff Scott and Youngmoo E. Kim. Instrument identification informed multi-track mixing. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 2013.

[15] M. Senior. *Mixing Secrets*. Taylor & Francis, 2012.

[16] George Tzanetakis, Randy Jones, and Kirk McNally. Stereo panning features for classifying recording production style. In *Proceedings of the 8th International Society for Music Information Retrieval Conference*, 2007.

[17] Earl Vickers. The loudness war: Background, speculation, and recommendations. *129th Convention of the Audio Engineering Society*, 2010.