Intelligent Multitrack Dynamic Range Compression

ZHENG MA[1], BRECHT DE MAN,[1] *AES Student Member*, PEDRO D. L. PESTANA,[2] *AES Associate Member*,

(z.ma@qmul.ac.uk)   (b.deman@qmul.ac.uk)   (ppestana@porto.ucp.pt)

DAWN A. A. BLACK[1], AND JOSHUA D. REISS,[1] *AES Member*

(dawn.black@qmul.ac.uk)   (joshua.reiss@qmul.ac.uk)

[1]*Queen Mary University of London, London, UK*
[2]*The Catholic University of Portugal, Oporto, Portugal*

We present an intelligent approach to multitrack dynamic range compression where all parameters are configured automatically based on side-chain feature extraction from the input signals. A method of adjustment experiment to explore how audio engineers set the ratio and threshold is described. We use multiple linear regression to model the relationship between different features and the experimental results. Parameter automations incorporate control assumptions based on this experiment and those derived from mixing literature and analysis. Subjective evaluation of the intelligent system is provided in the form of a multiple stimulus listening test where the system is compared against a no-compression mix, two human mixes, and an alternative approach. Results showed that mixes devised by our system are able to compete with or outperform manual mixes by semi-professionals under a variety of subjective criteria.

1 INTRODUCTION

Dynamic range compression (DRC) is a nonlinear audio effect that reshapes the dynamic range of an audio signal resulting in a reduced amplitude range [1, 2]. DRC is commonly used in audio production, noise management, broadcasting, and live performance applications. To a large extent, DRC defines much of the sound of contemporary mixes. However, it is arguably the most misused and overused effect in audio mixing [1]. If used excessively, the dynamic range compressor suppresses musical dynamics, producing lifeless recordings deprived of their natural character. Inappropriate parameter settings also produce artifacts such as pumping and breathing. Furthermore, conventional use of a static set of compressor parameters might not be optimal when the dynamic characteristics of the signal vary significantly over time.

Parameter automation of a dynamic range compressor using computerized signal analysis can provide advantages to audio amateurs or musicians who lack expert knowledge in signal processing. Such tools are capable of producing intelligent mixing decisions [3] that speed up the routine work and the trial-and-error process of avoiding inappropriate sonic artifacts.

Automatic DRC is a relatively unexplored field in academia. An RMS measurement was used to scale the release time constant in [4]. In [5] the time constants were automated by observing the difference between the peak and RMS levels of the signal fed into the side-chain. More relevant research can be found in [6, 7] where a series of

methods to automate most of the parameters of a digital dynamic range compressor based on side-chain feature extraction from the input signal were presented. However, in this system, the threshold was still manually chosen, with ratio set to infinity and an automated soft knee determining the amount of compression based on spectral flux. A new linear DRC technique that reduced the peak amplitude of transient signals using golden ratio allpass filters was introduced in [8]. In [9] a new class of adaptive digital audio effects that mapped semantic metadata to control parameters was proposed. However, the system assumed that the metadata already exists either from a prior process or manual configuration and might be invoked on demand. The automation was performed using a fairly simple mapping between different selection of metadata and static compression presets. No subjective evaluation was provided.

Perhaps the most relevant previous work is [10], which described an off-line method for automating multitrack DRC based on loudness and loudness range. The control strategy was to reduce the difference between the highest and lowest loudness range of the multitracks and sound sources where a higher loudness range requires greater amounts of DRC. However, the parameter automation of transforming the three controls (threshold, ratio, and knee) into a single control could have a significant effect on the final result. The evaluation results in [10] were inconclusive regarding the sonic improvement of the mixes.

In this paper we propose a fully automated multitrack DRC algorithm exploiting the interdependence of the input audio features in order to produce the appropriate amount of
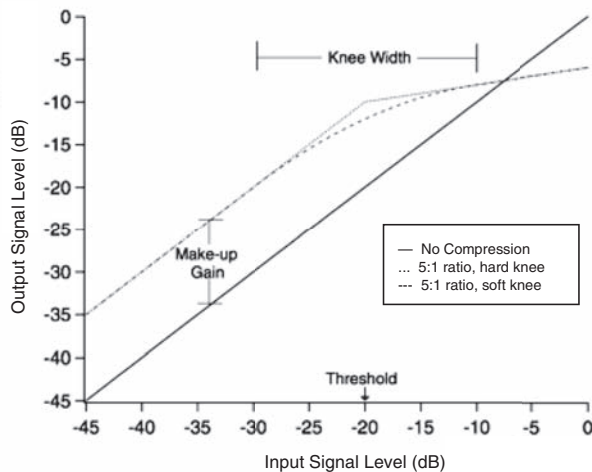
Fig. 1. Static DRC characteristic with make-up gain and hard or soft knee

compression, incorporating best practices as control rules. Sec. 2 provides control assumptions to automate the system and the rationale as to why the proposed features explored in Sec. 3 are relevant. A DRC method of adjustment experiment is described in Sec. 3 and multiple linear regression models are then applied to the results to derive the ratio and threshold automations. Finally, the intelligent multitrack DRC algorithms are presented in Sec. 4 followed by a subjective evaluation in the form of listening test in Sec. 5 and conclusion in Sec. 6.

2 BEST PRACTICES IN USE OF DRC

2.1 Compressor Model and Parameters

The fundamental digital compressor model design employed in our approach is a feed-forward compressor with smoothed branching peak detector [6, 7]. A typical set of parameters of a dynamic range compressor can be briefly described as follows:

Threshold denotes the level above which gain reduction starts. *Ratio* determines the input/output ratio for signals exceeding the threshold level. *Attack* and *Release* time, also known as time constants, determine how fast the compressor acts. *Knee width* controls whether the threshold-determined point in the transfer characteristics of a compressor has a sharp (hard-knee) or smoothed transition (soft-knee). As the compressor reduces the level of the signal, a *Make-up Gain* can be added at the output to compensate for level loss. The static transfer characteristic of DRC is depicted in Fig. 1. A dynamic range compressor may have many additional controls such as hold, side-chain filtering or look-ahead [1].

2.2 Control Assumptions

In this section we list and discuss control assumptions derived from the literature and analysis to automate the compressor parameters.

List of Assumptions:

1. A signal with a high degree of level fluctuations should have more compression.
2. A signal with more low frequency content should have more compression.
3. Attack and release time should be dependent on the transient nature of the signal.
4. Knee width should depend on the amount of compression applied.
5. Make-up gain should be set so that output loudness equals input loudness.
6. There is a maximum and optimal amount of DRC that depends on sound source features.

Regarding Assumption 1, in a survey about the main reasons to apply DRC [11, 12], most professional mixing engineers who participated stated that their main intention was to "stabilize erratic loudness range." They often compress instruments that have high note-to-note level variations, such as vocals or drum tracks, so that their relative levels are more consistent. A number of dynamics features have been proposed recently that measure the degree of level fluctuation, including EBU loudness range [13] and dynamic spread [14], which is simply the p-norm of the signal. Yet subjective listening test results in [15] suggested none of the metrics accurately predict the perceived dynamic range of a musical track. In [16], the authors proposed some parameter alterations of [13] that might yield better results for multitrack material. Alternatively, the crest factor, calculated as the peak amplitude of an audio waveform divided by its RMS value, can also be a coarse measurement of dynamic range.

Assumption 2 is based on analysis of mixes in [11, 12], which showed that "Compression takes place whenever headroom is at stake and the low-end is usually more critical." Thus spectral features of the source audio signal such as spectral centroid, spectral spread, and brightness are worth exploring to reveal the degree of frequency dependence and low-end sensitivity of DRC.

As for Assumption 3, attack times usually span between 5 ms and 250 ms and release times are often within the 5 – 3000 ms range. It is generally accepted that attack and release time parameters are employed to catch the transient nature of the sound [1, 17]. Some commercial compressors offer a switchable auto-attack or auto-release, which are mostly based on measuring the difference between the peak and RMS levels of the side-chain signal. In academia, [4] proposed to scale the release time based on RMS values, [5] scaled attack and release times based on the crest factor. More recently, [6] revisited the subject and scaled both parameters based on either modified crest factor or modified spectral flux, also used in [10]. Previous research shares a general idea: if a signal is highly transient or percussive, shorter time constants are preferred.

Regarding Assumption 4, a soft knee enables smoother transition between non-compressed and compressed parts of the signal and thus yields a more transparent compression effect. In order to produce a natural compression

effect in an automatic mixing system, the knee width should be adaptively configured based on the estimated amount of compression applied on the signal [3]. The amount of compression applied largely depends on the relationship between threshold and ratio.

Assumption 5 can be regarded as a direct consequence of the definition of make-up gain. Automatic make-up gain based on the average control-voltage is commonly used in commercial DRC products. However, [1] pointed out that this often produces a perceived loudness variation in practice. Subjective evaluation in [6] showed that the EBU loudness-based make-up gain produced a better approximation of how professional mixing engineers would set the make-up gain.

As for Assumption 6, quantitative descriptions about the amount of compression that should be applied on different instruments can be found in the literature [18, 19]. [6] performed a discrete separation between transient and steady state signals and allowed the former to have larger variability of the knee width in order to accommodate for transient peaks. In [20], percussive tracks are manually separated from sustained sounds, as they are assumed to need a different treatment.

3 RATIO AND THRESHOLD ADJUSTMENT EXPERIMENT

Ratio and threshold are the most crucial parameters in determining the amount of DRC. Assumptions 1 and 2 (Sec. 2.2) suggest that audio features that describe the dynamic and spectral content of the signal might have a high degree of correlation with the preferred amount of DRC. We propose a method of adjustment experiment to uncover how subjects set the ratio and threshold. Several feature candidates are proposed and their correlations with the subjective results are analyzed. We apply a least-squares based multiple linear regression model to formulate the relationship between the identified features and the test results and finally to derive the ratio and threshold parameter automation.

3.1 Method of Adjustment Experiment

Fifteen participants, all of whom have audio engineering experience, two of whom are professional mixing engineers, were recruited to perform a DRC adjustment task on four multitrack recordings in various genres (Song 1: Rock; Song 2: Pop; Song 3: Alternative; Song 4: Folk). Twenty-second excerpts were extracted from the chorus of each song for use in the test. The test subjects were asked to set the ratio and threshold for each instrument track of each song until they were satisfied with the amount of DRC applied to the mix. A solo function to play back an individual track with or without compression was provided in the experiments. However, subjects were advised to listen to the mix when setting the parameters for each individual track. Other compressor parameters were automated as described in Sec. 4. The ratio and threshold values were hidden from participants to prevent bias resulting from common practices.

The average results of the 15 participants for ratio and threshold for each track, along with 95% confidence intervals, are shown in Fig. 2. The small variations in results were unexpected since dynamic range compression is often assumed to be an art, with varying tastes in its application. However, we can also see from Fig. 2 that different tracks have differing variation sizes, suggesting that DRC parameter setting is track dependent. We further note that half of the participants are from the same UK research group and thus might share a similar taste in compression that could potentially bias the results.

3.2 Feature Correlations

Several dynamic and spectral features are proposed, extracted, and analyzed based on the subjective results.

First, the RMS level as a rough dynamic feature is defined by Eq. (1),

$$x_{RMS} = 20 \log_{10} \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} |x^2(n)|} \qquad (1)$$

where N is the window length.

EBU loudness range (LRA) is defined as the difference between the 10th percentile and the 95th percentile on the histogram after a dual gating process [13].

Dynamic spread [14] is given by Eq. (2),

$$d = \frac{1}{N} \sum_{n=0}^{N-1} |x_{dB}(n) - x_{RMS}| \qquad (2)$$

where x_{dB} is the input signal in digital full scale (dBFS).

The spectral centroid [21] is the barycenter of the spectrum, calculated by Eq. (3),

$$\mu = \frac{\sum_{k=0}^{K-1} f(k)X(k)}{\sum_{k=0}^{K-1} X(k)} \qquad (3)$$

where $X(k)$ represents the spectral magnitude of signal $x(n)$, of bin number k, and $f(k)$ represents the center frequency at that bin.

Spectral spread [21] represents the spread of the spectrum around its mean value is defined as

$$\sigma_s^2 = \sum_{k=0}^{K-1} (X(k) - \mu)^2 f(k) \qquad (4)$$

The brightness [21] of the signal is defined as the amount of energy above a defined cut-off frequency of 1500 Hz.

The practical calculation of the features mentioned before can be found in [21]. We also propose two new, cross-adaptive audio features called percussivity weighting and low-frequency weighting.

Percussivity weighting describes the cross-adaptive relationship among all the input signals regarding the degree of level fluctuations and is based on the crest factor values.
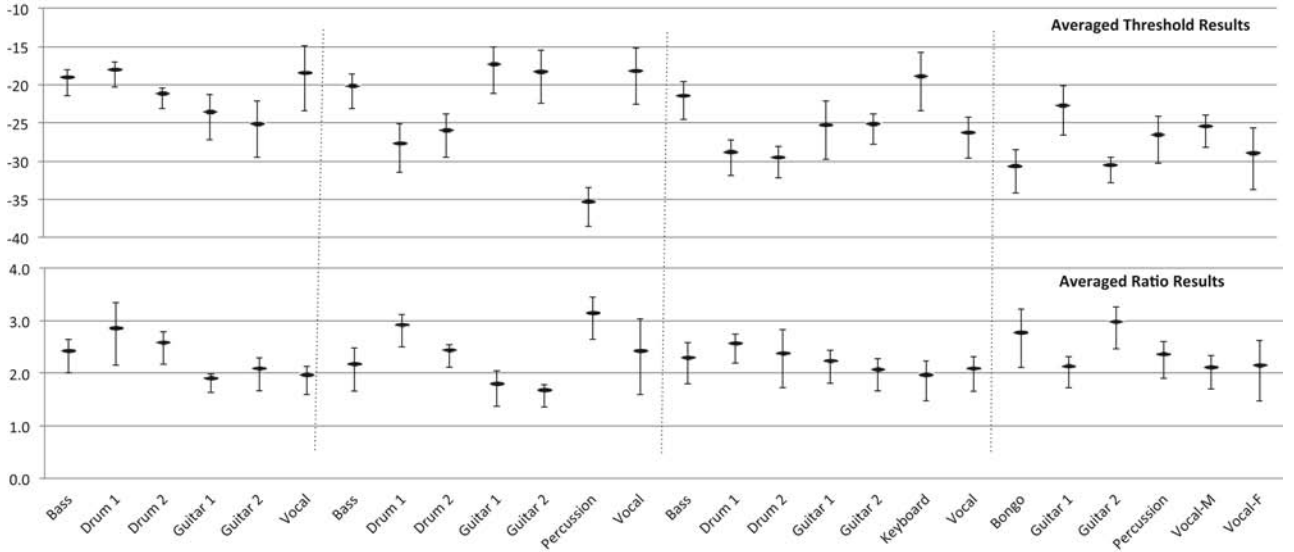
Fig. 2.   Ratio and threshold mixing results with 95% confidence intervals

First, the average value of the crest factor over all tracks is computed as

$$\bar{x}_{crest} = \frac{1}{M} \sum_{m=1}^{M} x_{crest}(m) \qquad (5)$$

where m is the index of the track number and M is the total number of input tracks. The average crest factor $\bar{x}_{crest}$ is then used as an adaptive threshold for the percussivity weighting $w_p(m)$ calculation.

The mapping between $w_p(m)$ and $\bar{x}_{crest}$ is formulated using a modified Gaussian distribution centered around $\bar{x}_{crest}$ by Eq. (6),

$$g(x) = a e^{\frac{(x_{crest} - \bar{x}_{crest})^2}{2\sigma^2}} \qquad (6)$$

where σ is the standard deviation controlling the width of the 'bell' shape. $w_p(m)$ is formulated empirically as follows,

$$w_p(m) = \begin{cases} e^{\frac{(x_{crest}(m) - \bar{x}_{crest})^2}{2\sigma^2}}, & x_{crest}(m) \le \bar{x}_{crest} \\ 2 - e^{\frac{(x_{crest}(m) - \bar{x}_{crest})^2}{2\sigma^2}}, & x_{crest}(m) > \bar{x}_{crest} \end{cases} \qquad (7)$$

σ is set to 2 based on informal testing. Eq. (7) shows that $w_p(m) \in (0, 2)$. The larger the $w_p(m)$ value, the more percussive the mth track is. Eq. (7) guarantees that most values of $w_p(m)$ are centered on the adaptive reference $\bar{x}_{crest}$.

Low-frequency weighting is introduced to describe the relative amount of low-frequency energy of each signal compared to the average low frequency ratio. A Fast Fourier Transform (FFT) with Hanning window is performed on each signal frame to obtain the spectral distribution, $X(m,k)$ of track m at frequency bin k. $X_{low}(m,k)$ is the spectral distribution of low-pass filtered version of input signals

with cut-off frequency set to 1 kHz, the cross-adaptive low-frequency weighting, $w_f(m)$ is defined by Eq. (8),

$$w_f = \frac{\sum_{k=0}^{K-1} \frac{X_{low}(m,k)}{X(m,k)}}{\frac{1}{M} \sum_{m-1}^{M} \sum_{k=0}^{K-1} \frac{X_{low}(m,k)}{X(m,k)}} \qquad (8)$$

The values of each described feature are extracted from each multitrack and shown in Table 1. The cross-correlation coefficient between each feature and the averaged ratio and threshold values across 15 participants is calculated as follows

$$r_{xy} = \frac{\sum_{i=1}^{M} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{M} (x_i - \bar{x})^2 \sum_{i=1}^{M} (y_i - \bar{y})^2}} \qquad (9)$$

where x_i is the feature value, y_i is the observed ratio or threshold value of each multitrack, and $\bar{x}$, $\bar{y}$ are the respective means. The coefficients are listed in Table 2.

As Table 2 shows, spectral features generally exhibited higher correlation with ratio parameter than dynamic features. This is in agreement with Assumption 2. The proposed low-frequency weighting shows the highest correlation with the ratio parameter.

The RMS level shows the strongest correlation with threshold. However, in the spectral feature subgroup, all correlations are relatively weak, indicating that dynamic features play a more significant role in setting the threshold parameter than spectral features.

Notice that the perception-based EBU LRA has the lowest correlation coefficients with both ratio and threshold. First, EBU loudness is designed for broadcast material rather than individual tracks in multitrack content. Second, the 3s integration window length is too long to capture small level fluctuations in terms of dynamics.

Table 1. Dynamic and spectral feature values

Song	Track	Percussivity weighting	EBU loudness range (LU)	Dynamic Spread	RMS (dB)	Low-Frequency weighting	Brightness	Spectral Centroid (Hz)	Spectral spread (Hz)
1	Bass	0.36	1.17	1.31	−15	1.37	0.026	373.1	1315.8
	Drum1	0.86	1.69	8.09	−17.5	1.38	0.539	3479.7	3906.4
	Drum2	1.52	2.09	3.39	−18.4	0.85	0.694	4394.7	3973.4
	Guitar1	1.15	0.47	0.88	−14.6	0.75	0.458	1762.1	1697.7
	Guitar2	1.07	0.77	0.97	−17.2	0.56	0.549	2140.9	1987.3
	Vocal	0.67	3.54	8.08	−19.8	0.72	0.592	4313.2	4219.6
2	Bass	0.52	1.88	2.47	−16.8	1.13	0.049	476.5	1331.6
	Drum1	0.93	3.91	10.65	−26.3	1.35	0.360	1927.8	2834.3
	Drum2	0.99	4.85	9.79	−25.6	0.93	0.440	2051.3	2665.5
	Guitar1	0.34	7.80	3.88	−13.6	1.03	0.196	836.5	1358.3
	Guitar2	0.36	0.65	2.02	−12.4	0.58	0.257	1087.3	1378.8
	Percussion	1.98	0.77	3.19	−31.9	1.29	0.997	3082.4	4115.7
	Vocal	0.63	6.08	8.38	−21.2	0.62	0.429	3354.7	4376.5
3	Bass	0.6	3.02	6.14	−17.2	1.38	0.105	683.7	1922.6
	Drum1	0.64	5.41	12.04	−25.1	1.63	0.488	3512.2	4245.3
	Drum2	0.97	6.83	10.58	−24.3	0.94	0.525	3563.6	4396.2
	Guitar1	0.99	2.00	1.26	−22.1	0.88	0.365	1458.2	1648.4
	Guitar2	1.13	3.67	1.96	−21.2	0.48	0.597	1987.3	1778.5
	Keyboard	1.01	3.91	2.67	−15.6	0.41	0.317	1573.1	2258.3
	Vocal	0.96	10.20	17.92	−21.6	0.79	0.305	1908.5	2628.4
4	Bongo	0.98	1.61	13.93	−26.2	0.92	0.222	1473.9	2406.1
	Guitar1	1.01	2.97	2.50	−30.5	0.93	0.170	1216.7	2691.3
	Guitar2	1.31	4.52	6.33	−18.3	1.25	0.390	2459.7	3695.1
	Percussion	0.76	3.22	26.98	−30.5	1.22	0.388	3082.4	4412.0
	Vocal-M	1.12	4.94	2.80	−25.4	0.88	0.224	1507.0	2798.0
	Vocal-F	0.7	9.30	10.44	−20.1	0.72	0.340	2581.7	4061.4

3.3 Curve Fitting: Modeling Ratio and Threshold

Multiple linear regression techniques are applied to model the relationship between the proposed features and the ratio and threshold experiment results [22]. Combinations of different audio features and various modeling functions are investigated to obtain the best fit by assessing their Goodness-Of-Fit (GOF) statistics, confidence intervals, and residual plots with validation data.

We investigate various modeling functions with all the feature combinations considered. Ratio curve fits with significant Goodness-Of-Fit are presented in Table 3. Insignificant fits are not depicted in the table.

The Sum of Squares due to Error (SSE) is the total deviation of the response values from the fit to the response values, or simply the sum of squares of residuals, calculated as

$$SSE = \sum_{i=0}^{n} (y_i - \bar{y}_i) \qquad (10)$$

where y_i is the ith response value from the fit, $\hat{y}_i$ is response value and n is the number of observations. It is a measure of the discrepancy between the data and an estimation model. A small SSE indicates a tight fit of the model to the data.

The coefficient of determination, denoted R^2 provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model. R^2 ranges from 0 to 1, with a value closer to 1 indicating that the model accounts for a

Table 2. Feature correlations

	Feature	Ratio Correlation	Threshold Correlation
Dynamic Feature	Percussivity	0.4954	−0.6019
	LRA	−0.1499	−0.1275
	Dynamic Spread	0.2486	0.3294
	RMS level	−0.4871	0.6659
Spectral Feature	Low-Frequency	0.6351	−0.248
	Spectral Centroid	0.3592	−0.2031
	Spectral Spread	0.4996	−0.3571
	Brightness	0.3791	−0.3926

Table 3. Ratio curve Fitting results Goodness-Of-Fit

Feature Selection				Goodness-Of-Fit			
X data	Y data	Modeling Functions $f(x, y) = \ldots$	Coefficients	SSE	R-square	Adjusted R-square	RMSE
Low-Frequency		$p_1 x + p_2$	$p_1 = 0.7411; p_2 = 1.51$	2.101	0.4034	0.3785	0.2959
Percussivity		$p_1 x + p_2$	$p_1 = 0.5077; p_2 = 1.762$	2.657	0.2454	0.214	0.3327
Percussivity	Low-Frequency	$p_{10} x + p_{01} y$	$p_{10} = 0.969; p_{01} = 1.342$	2.45	0.3043	0.2753	0.3195
Percussivity	Low-Frequency	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = 0.968; p_{10} = 0.554$ $p_{01} = 0.783$	1.078	0.6939	0.6673	0.2165
Percussivity	Low-Frequency	$1 + p_{10} x + p_{01} y$	$p_{10} = 0.540; p_{01} = 0.764$	1.079	0.6935	0.6807	0.2121
Percussivity	Low-Frequency	$p_{00} + p_{10} x + p_{01} y + p_{11} xy$	$p_{00} = 1.108; p_{10} = 0.122$ $p_{01} = 1.257; p_{11} = 1.108$	1.84	0.4776	0.4063	0.2892
Percussivity	Low-Frequency	$p_{00} + p_{10} x + p_{01} y + p_{20} x^2 \rightleftharpoons + p_{11} xy + p_{02} y^2$	$p_{00} = 0.933; p_{10} = 0.222$ $p_{01} = 1.325; p_{11} = -0.182$ $p_{20} = 0.008; p_{02} = 0.008$	0.915	0.7401	0.6752	0.2139
Percussivity	Low-Frequency	$p_1 2^x + p_2 2^y$	$p_1 = 0.4508; p_2 = 0.6694$	1.337	0.6202	0.6044	0.2361
LRA	Low-Frequency	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = 1.57; p_{10} = -0.013$ $p_{01} = 0.731$	2.07	0.4121	0.361	0.3
Dynamic spread	Low-Frequency	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = 1.5; p_{10} = 0.052$ $p_{01} = 0.714$	2.077	0.4101	0.3588	0.3005
Dynamic spread	Low-Frequency	$p_{00} + p_{10} x + p_{01} y + p_{20} x^2 \rightleftharpoons + p_{11} xy + p_{02} y^2$	$p_{00} = 1.283; p_{10} = 0.043$ $p_{01} = 0.972; p_{11} = -0.003$ $p_{20} = 0.017; p_{02} = -0.159$	1.86	0.4719	0.3398	0.3049
RMS	Low-Frequency	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = 1.117; p_{10} = 0.023$ $p_{01} = 0.6.34$	1.724	0.5103	0.4677	0.2738
Percussivity	Spectral Spread	$p_{00} + p_{10} x + p_{01} y + p_{20} x^2 \rightleftharpoons + p_{11} xy + p_{02} y^2$	$p_{00} = 1.516; p_{10} = 0.3844$ $p_{01} = 0.0001$	2.196	0.3762	0.322	0.309
Percussivity	Spectral Spread	$p_{00} + p_{10} x + p_{01} \log(y)$	$p_{00} = 0.754; p_{10} = 0.3546$ $p_{01} = 0.7768$	2.206	0.3735	0.319	0.309
Percussivity	Spectral Centroid	$p_{00} + p_{10} x + p_{01} \log(y)$	$p_{00} = 1.31; p_{10} = 0.4502$ $p_{01} = 0.1546$	2.621	0.2557	0.191	0.3376
Percussivity	Brightness	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = 1.761; p_{10} = 0.4647$ $p_{01} = 0.105$	2.651	0.2472	0.1817	0.3395

greater proportion of variance. The general definition of the coefficient of determination is given by Eq. (11),

$$R^2 = 1 - \frac{SSE}{SST} \quad (11)$$

where SST is the total sum of squares proportional to the sample variance, defined as,

$$SSE = \sum_{i=0}^{n} (y_i - \bar{y}_i) \quad (12)$$

where $\bar{y}$ is the mean of y_i.

Degrees of Freedom Adjusted R^2 is generally the preferred indicator to compare two models that are nested. Like R^2, it ranges from 0 to 1, and is given by Eq. (13),

$$R^2_{adjusted} = 1 - \frac{SSE(n-1)}{SST(v)} \quad (13)$$

where $v = n\text{-}m$, v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares and m is the number of fitted coefficients estimated from the response values.

Root-Mean-Square Error (RMSE) estimates the standard error of the regression, as defined in Eq. (14),

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (14)$$

We found that the combination of percussivity and low-frequency weighting generates the best fit regardless of the modeling function evaluated. In general, the modeling functions using percussivity and low-frequency weighting yield an SSE smaller than 1.5 and RMSE smaller than 2.5, while others show a SSE larger than 2 and RMSE larger than 3. The results agree with the feature correlation coefficients obtained in Sec. 3.2.

Four models, $f(x,y) = p_{00}+p_{10}x+p_{01}y$, $f(x,y) = p_{10}x+p_{01}y+1$, $f(x,y) = p_{00}+p_{01}y+\ldots+p_{02}y^2$ and $f(x,y) = p_1 2^x + p_2 2^y$ performed a significantly better fit based on the Goodness-Of-Fit statistics.

By comparing the Goodness-Of-Fit produced by the first order polynomial $f(x,y) = p_{00}+p_{10}x+p_{01}y$ with $f(x,y) = p_{10}x+p_{01}y$, both use the percussivity and low-frequency weighing features, we see that SSE decreases by more than half and RMSE decreases by roughly 0.1. This means the accuracy of the model improves. Moreover, since the two models are nested, the adjusted R^2 increases significantly
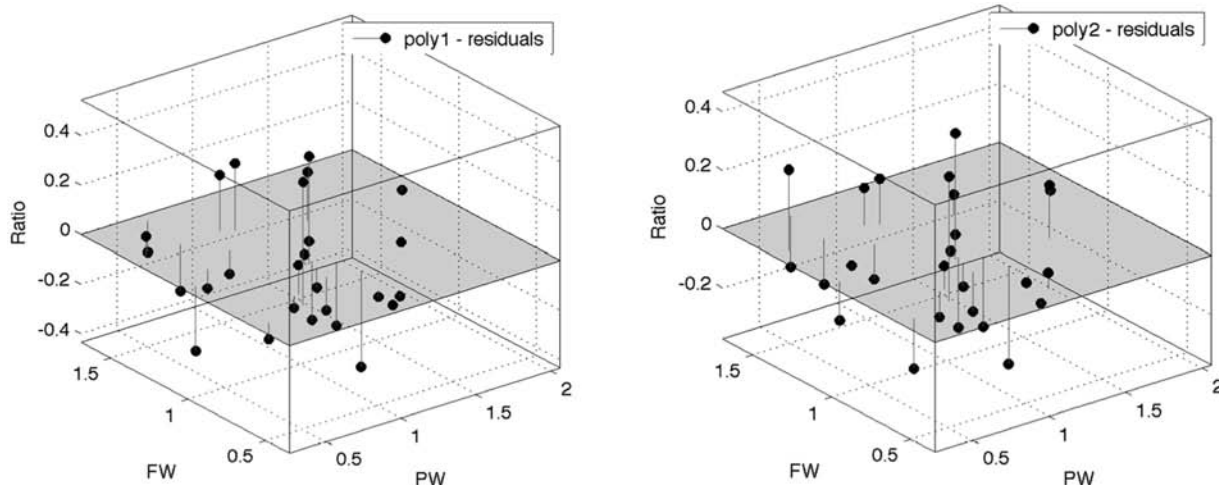
Fig. 3. Residual plots of the first and second order polynomial models, where proposed low-frequency weighting and percussivity weighting feature are denoted as *FW* and *PW* respectively.

from 0.2753 to 0.6673 when adding the additional constant term p_{00}, implying the latter performs better again. The model $f(x,y) = p_{00}+p_{10}x+p_{01}y$ also outdoes $f(x,y) = p_1 2^x + p_2 2^y$ with lower SSE and RMSE. The adjusted R^2 of the model $f(x,y) = p_{10}x+p_{01}y+1$ is larger than the one of $f(x,y) = p_{00}+p_{10}x+p_{01}y$, indicating that it excels the latter in the performance of model prediction.

Although the second degree polynomial model has slightly larger RMSE, SSE is smaller and R^2 is larger. Since they are not nested, we cannot pick the best fit based on their adjusted R^2 coefficients. Therefore, we plot residuals and prediction bounds to assess both models graphically. The residual plots of the two models are shown in Fig. 3. Neither residual plot provides exhibits structure, suggesting that both models fit the data to an acceptable extent. The prediction bounds with 95% confidence level are presented in Fig. 4. The prediction bounds for the first degree polynomial model with 1 as constant term indicate that the model can be predicted with a small uncertainty (less than 0.8) throughout the entire data range. This is not the case for the

second degree polynomial model. It has wider prediction bounds in the area where not enough data exists, apparently because the data does not contain enough information to estimate the higher degree polynomial terms accurately. In other words, a second order polynomial model overfits the data.

With all criteria considered, $f(x,y) = p_{10}x+p_{01}y+1$ using percussivity and frequency weighting performs the best curve fit.

We perform the same analysis procedure for the model fitting of threshold. The Goodness-Of-Fit statistical results for the threshold curve fitting are presented in Table 4. Again, only modeling functions with relatively good degree of fit are listed here.

Analysis from Table 4 shows that models using a feature combination of RMS and percussivity weighting employing first and second order polynomial functions outperform others, with lowest SSE of 259.6, 246.1 and highest R^2 of 0.5565, 0.5796 respectively. Furthermore, second order polynomial models have a slightly better fit than first order
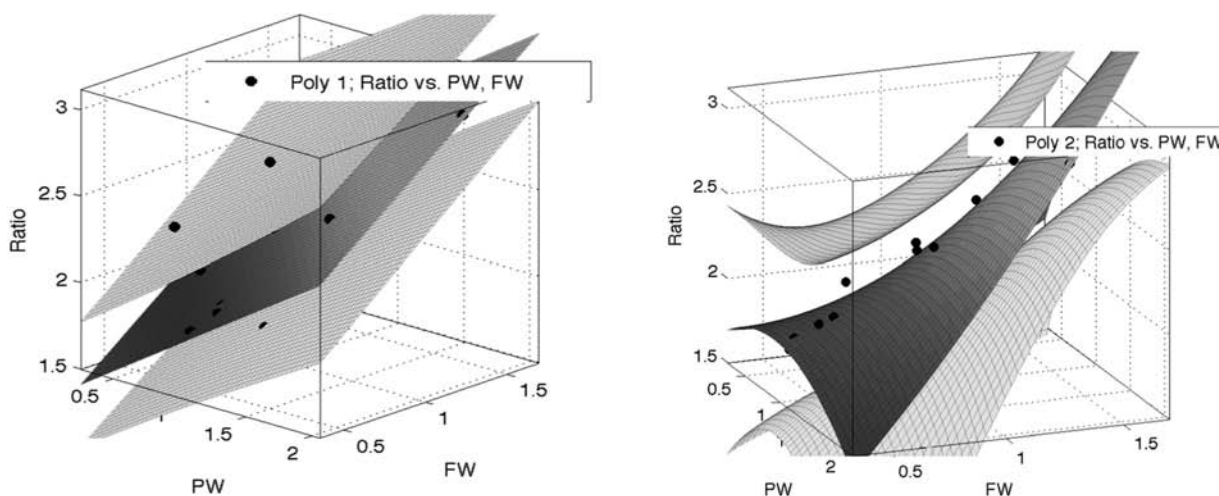

Fig. 4.   Prediction bounds with 95% confidence level of the first (top) and second (bottom) order polynomial models.

Table 4. Threshold curve fitting results with Goodness-Of-Fit

| Feature Selection | | | | Goodness-Of-Fit | | | |
X data	Y data	Modeling Functions $f(x, y) = \ldots$	Coefficients	SSE	R-square	Adjusted R-square	RMSE
RMS		$p_1 x + p_2$	$p_1 = 0.5947$; $p_2 = -12.33$	325.8	0.4434	0.4203	3.684
RMS		$p_1 x^2 + p_2 x^1 + p_3$	$p_1 = -30.29$; $p_2 = 1.341$ $p_3 = -4.584$	319.4	0.4543	0.4069	3.727
RMS		$p1 x^3 + p2 x^2 \ldots + p4$	$p_1 = 0.0007$; $p_2 = 0.06706$ $p_3 = -2.409$; $p_4 = 2.676$	319.1	0.4549	0.3806	3.808
Percussivity		$p_1 x + p_2$	$p_1 = -7.954$; $p_2 = -17.66$	373.3	0.3623	0.3358	3.944
Percussivity		$p_1 x^2 + p_2 x^1 + p_3$	$p_1 = 0.8318$; $p_2 = -9.69$ $p_3 = 16.88$	372.5	0.3637	0.3084	4.024
RMS	Percussivity	$p_{10} x + p_{01} y$	$p_{10} = 0.8659$; $p_{01} = -6.59$	441.4	0.246	0.2145	4.289
RMS	Percussivity	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = -11.03$; $p_{10} = 0.441$ $p_{01} = -4.987$	259.6	0.5565	0.5179	3.36
RMS	Percussivity	$p_{00} + p_{10} x + p_{01} y + p_{20} x^2 \rightleftharpoons + p_{11} xy + p_{02} y^2$	$p_{00} = -0.951$; $p_{10} = 1.684$ $p_{01} = 1.267$; $p_{11} = 0.032$ $p_{20} = 0.199$; $p_{02} = -0.89$	246.1	0.5796	0.4745	3.508
Dynamic Spread	Percussivity	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = -14.85$; $p_{10} = -0.3217$ $p_{01} = -8.614$	276.8	0.5271	0.486	3.469
LRA	Percussivity	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = -15.22$; $p_{10} = -0.479$ $p_{01} = -8.662$	334.9	0.4279	0.3782	3.816
RMS	Brightness	$p_{00} + p_{10} x + p_{01} y$	$p_{00} = -11.72$; $p_{10} = -0.537$ $p_{01} = -4.691$	303.3	0.4819	0.4369	3.631
RMS	Spectral Spread	$p_{00} + p_{10} x + p_{01} \log(y)$	$p_{00} = -8.551$; $p_{10} = 0.5706$ $p_{01} = -1.253$	324.8	0.4451	0.3968	3.758
RMS	Spectral Centroid	$p_{00} + p_{10} x + p_{01} \log(y)$	$p_{00} = -7.162$; $p_{10} = 0.5589$ $p_{01} = -1.815$	320.2	0.4529	0.4054	3.731

in terms of SSE and R^2. However, when comparing the $R^2_{adjusted}$ and RMSE values, first order appears to be the right choice.

Residual plots of each modeling are shown in Fig. 5. Neither residual plot provides evidence for choosing the best fit. Therefore, prediction bounds with 95% confidence level are further considered, as shown in Fig. 6. The second order polynomial model has a wider prediction bounds and tends to over-fit the data.

With all criteria considered, the first order polynomial model using RMS and percussivity weighting, performs the best data fit.

4 IMPLEMENTATION

In Cross-Adaptive Digital Audio Effects (CA-DAFX), the signal processing of any individual source is informed by the relationships between all involved sources [2, 3]. The cross-adaptive processing section of an intelligent multi-channel audio editing system exploits the interdependence of the input features in order to output the appropriate control data. This data controls the parameters in the signal processing of the multichannel content. The cross-adaptive feature processing can be implemented by a set of constrained rules that consider the interdependence between
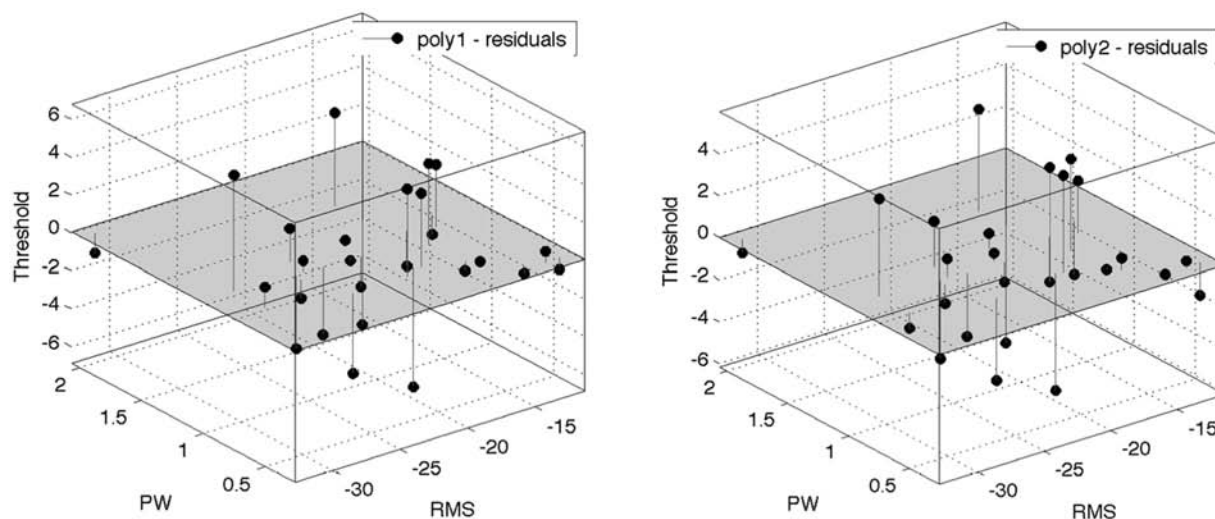

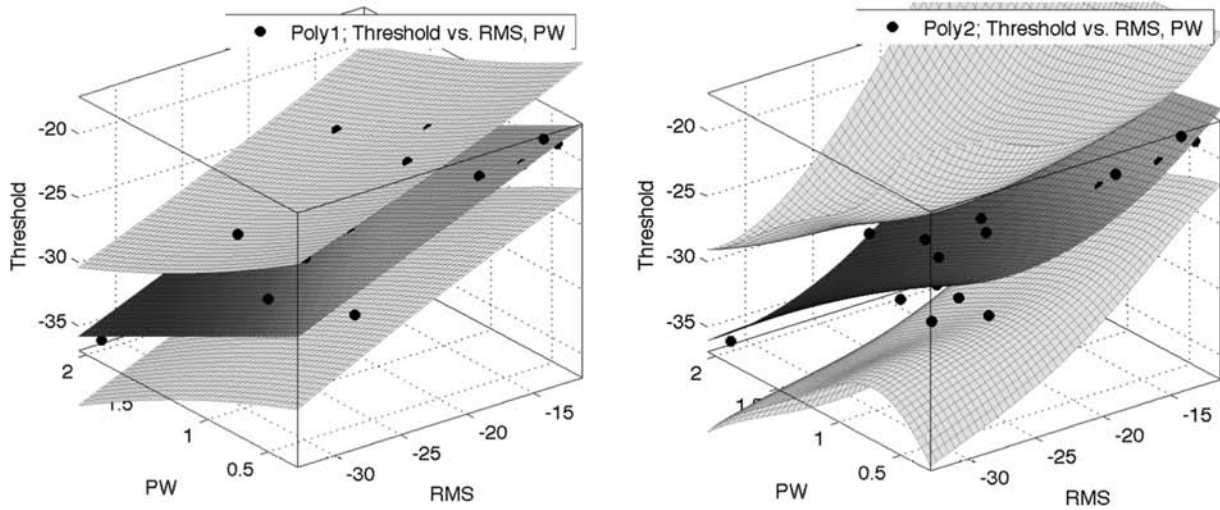Fig. 5. Residual plots for first and second polynomial models

Fig. 6.   Prediction bounds with 95% confidence level of the first (top) and second (bottom) order polynomial models.

channels. Extracted features from all channels are sent to the same feature-processing block where controls are produced.

The proposed intelligent multitrack compressor is based on the CA-DAFX processing architecture [2, 3]. The system workflow is depicted in Fig. 7.

The ratio and threshold automation is derived from the previous curve fitting process described in Sec. 3.3.

For ratio automation we choose the model of $f(x,y) = p_{10}x + p_{01}y + 1$ with percussivity and frequency weighting features that performs the best curve fit. The final ratio automation for track m is given by Eq. (15),

$$R(m) = 0.54w_p(m) + 0.764w_f(m) + 1 \qquad (15)$$

Similarly, the threshold automation follows the first order polynomial model, $f(x,y) = p_{00} + p_{10}x + p_{01}y$ with RMS level and percussivity weighting (see Table 4). The threshold automation is defined as,

$$T(m) = -11.03 + 0.44x_{RMS}(m) - 4.987w_p(m) \qquad (16)$$

Learning from Assumption 3, we adapt the algorithms for attack and release automation in [6] using crest factor as a short term signal measure to describe the transient nature of the input signals.

To obtain the average RMS values sample by sample, we apply an Exponential Moving Average filter,

$$x_{RMS}[n] = \sqrt{(1 - \alpha)x^2[n] + \alpha x_{RMS}^2[n - 1]} \qquad (17)$$

The sample by sample average peak magnitude of the signal is calculated as

$$x_{peak}[n] = \sqrt{\max(x^2[n], (1 - \alpha)x^2[n] + \alpha x_{peak}^2[n - 1])} \qquad (18)$$

Since the peak detector's and RMS detector's smoothing constants α are set to be identical, the release envelopes of both detectors are guaranteed to be the same, and the peak

detector's output is no less than the detected RMS output. The crest factor x_{crest} of the signal is defined as

$$x_{crest}[n] = \frac{x_{peak}[n]}{x_{RMS}[n]} \qquad (19)$$

Based on [6], the attack and release time constants are calculated by

$$\tau_A[n] = \frac{2\tau_{A-\max}}{x_{crest}^2[n]}$$
$$\tau_R[n] = \frac{2\tau_{R-\max}}{x_{crest}^2[n]} \qquad (20)$$

where the maximum attack time $\tau_{A-\max}$ is set to 80 ms and the maximum release time $\tau_{R-\max}$ is set to 1000 ms [6].

According to Assumption 4, we set the knee width to half the absolute value of the automated threshold value for a soft knee configuration as

$$W(m) = \frac{|T(m)|}{2} \qquad (21)$$

that ensures that a lower threshold results in a wider knee width.

Following Assumption 5, make-up gain is set so that output loudness equals input loudness. The make-up gain is simply the loudness difference between the input and output of the DRC, measured following the EBU loudness standard [13],

$$G(m) = L_{in}(m) - L_{out}(m) \qquad (22)$$

where L_{in} and L_{out} are the input and output loudness values of individual source m, before and after the compression block. In automatic mixing, the loudness setting is usually done post-compression. This is not as simple as it may seem since an adaptive time-changing make-up gain acts as an auto-leveller and may neutralize the compression effect. Therefore, real-time implementation implies a very delicate choice of time constants for the adaptation process and a convergence strategy using cumulative values similar to the one that has been proposed for several instances of auto-mixing.
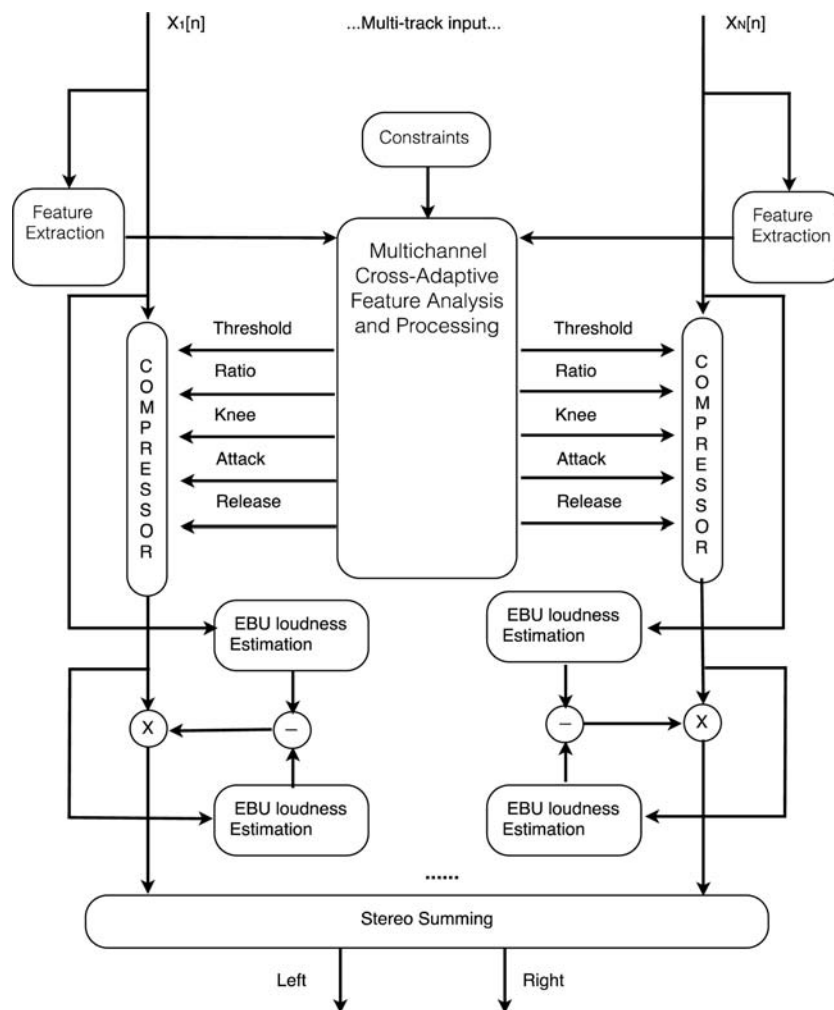
Fig. 7. System block diagram of the CA-DAFX intelligent multitrack compressor

5 SUBJECTIVE EVALUATION

5.1 Methods

Subjective evaluation of the intelligent multitrack compression algorithm was performed in the form of a multiple stimulus, MUSHRA-style [23] listening test.

Five versions of six different multitrack songs (20-second segments, not used in the ratio and threshold adjustment experiment) in various genres were created and compared in the test: a no-compression mix, the proposed automatic mix, two different manual mixes produced by semi-professional mixing engineers, and another automatic mix produced by an alternative approach [10]. Analysis and evaluation of audio features of these semi-professional mixes used can be found in [24].

In all mixes, the only parameter modified was the dynamic range compression to minimize the perceptual bias caused by other audio effects as much as possible. The loudness of the final mixes were normalized manually by a group of professional mixing engineers [23], as done on the same playback system as the subjective evaluation. The order of mixing versions and songs presented to each participant was randomized by a pseudorandom number gen-

erator algorithm in Matlab. Participants were encouraged to take as much time as needed.

Sixteen participants with moderate audio engineering experience, seven of whom were from the same group of people used in the previous ratio and threshold adjustment experiment, were asked to rate the mix versions according to four specific criteria/questions on a scale of 0 (very bad) to 100 (excellent):

Q1: According to the appropriateness of the amount of dynamic range compression applied to each individual sound source in the mix.
Q2: In terms of the degree of any imperfection such as pumping, breathing artifacts, level imbalance, etc.
Q3: According to the ability to stabilize the erratic level fluctuation within the mix.
Q4: According to participants' own overall preference.

Since DRC can be relatively subtle, we chose different songs for different questions to maximize the difference. Six songs were tested in Q1 and Q4 while four songs were tested in Q2 and Q3. For Q1, a no-compression mix of each song was also presented as a "reference." However, it does
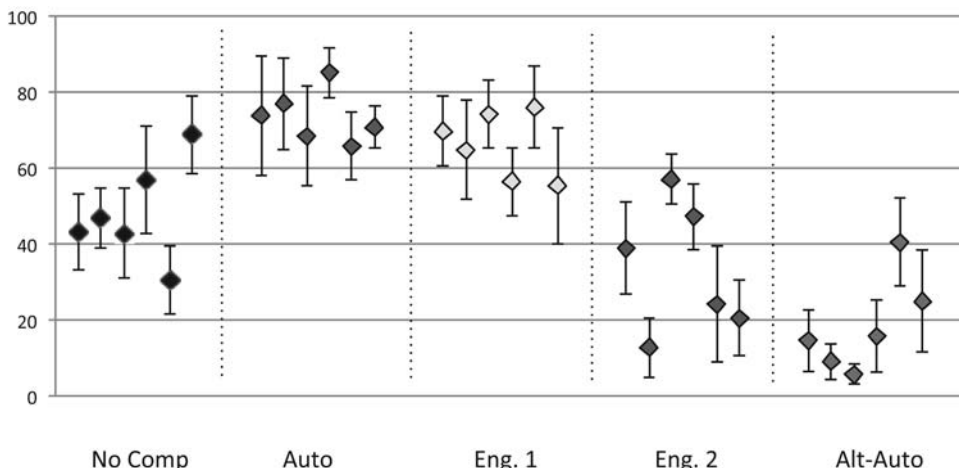
Fig. 8.   Averaged results of Q1: amount of DRC with 95% confidence intervals, grouped by mix type

not serve as objectively high quality reference or objectively low quality anchor, which was explained to the participants in advance. The order of the songs as well as the order of the versions of each individual song was randomized when presented to each participant for each question.

5.2   Evaluation Results

5.2.1   Q1: Appropriateness of the Amount of DRC

In Q1 participants were asked to rate the mixes in terms of the appropriateness of the amount of dynamic range compression applied in the mix. The results are summarized in Fig. 8, showing the mean, grouped by mix type, with error bars displaying 95% confidence intervals using the t-distribution. No compression mix, automatic mix, two semi-professional mixes, and alternative automatic mix are notated as "No Comp," "Auto," "Eng. 1," "Eng. 2," and "Alt-Auto" [10] respectively.

The "Eng. 1" and "Auto" mixes both rate consistently high throughout, with "Auto" outperforming "Eng. 1" in Song 1, 2, 4, and 6. The "Alt-Auto" mixes rate consistently low with the exception of Song 5 and 6.

5.2.2   Q2: Degree of Imperfection

Q2 investigates the degree of sound artifacts or imperfection. The results are presented in Fig. 9.

It can be seen that all "No Comp," "Auto," "Eng. 1," "Eng. 2" mixes are all rated above the middle score, with only the exception of "Eng. 2" in Song 2, which suggests these mixes do not have obvious artifacts. "Alt-Auto" rates the lowest (<20 for most cases) implying significant artifacts are produced in the mixes.

5.2.3   Q3: Ability to Stabilize Erratic Level Fluctuation

This question was designed to make the participants focus on how well the mixes can stabilize the level fluctuations. The results are shown in Fig. 10.

"Eng. 1" performs best followed by "Auto" except for Song 3. "Eng. 2" performs well in Song 2 and 3, while it is the worst in Song 4.

5.2.4   Q4: Overall Preference

Q4 was designed to study participants' overall preference for the DRC. The results are shown in Fig. 11.
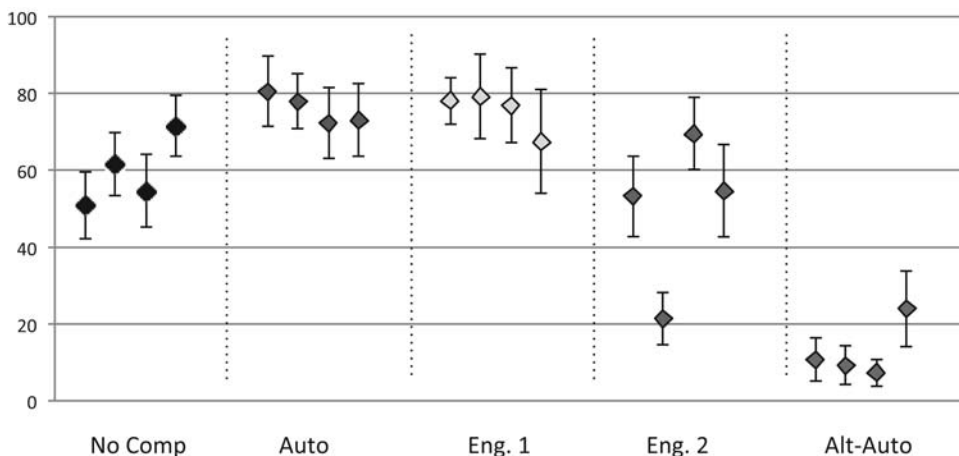

Fig. 9.   Averaged results of Q2: degree of imperfection with 95% confidence intervals, grouped by mix type
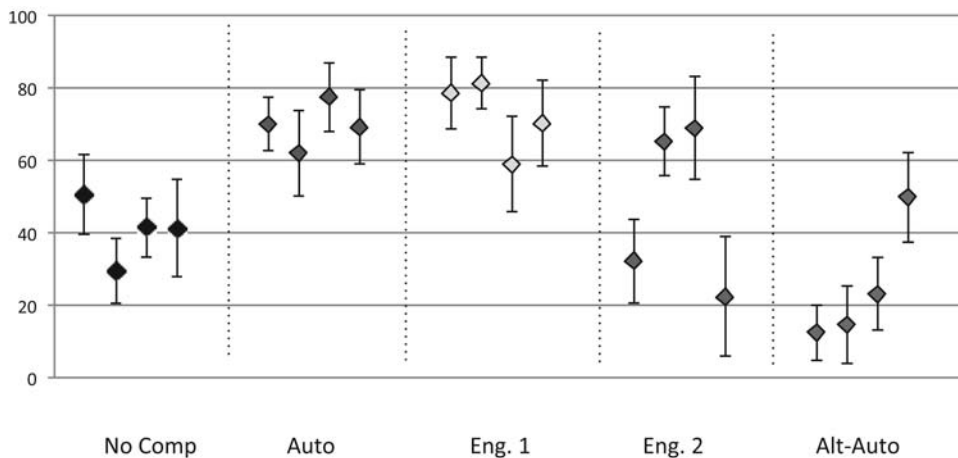
Fig. 10. Averaged results of Q3: level stabilizing with 95% confidence intervals, grouped by mix type

Participants generally prefer "Auto" and "Eng. 1" mixes throughout all the songs. "Eng. 2" has a strongly varying rating depending on the songs.

5.3 Summary

To give a clearer depiction of the overall performance of each mix type, the averaged mean results with 95% confidence intervals across all participants and songs are displayed in Fig. 12.

Fig. 12 shows that the proposed "Auto" performs best in Q1, the appropriate amount of DRC. "Auto" also performs quite well in terms of stabilizing the erratic level fluctuation in Q3. More importantly, the proposed automatic mixes are the participants' favorite in Q4, the overall preference. The no compression version is preferred to "Eng. 2" in Q1, Q2, and Q4, suggesting that people sometimes dislike the use of compression. The "Alt-Auto" is clearly the worst overall performer. The mean results in Q2 show that the Alt-Auto causes obvious sound artifacts or unpleasant effects. "Alt-Auto" uses pre-processing to equalize the loudness (measured by EBU loudness standard) of multitracks before compression, which is likely to make the percussive instrument much louder than other instruments, resulting in an unpleasant listening experience. The same issue was

addressed in [25]. This could be the reason that the "Alt-Auto" performs poorly. Notice that although Auto and Eng. 1 perform similarly in the subjective evaluation, results also show participants' preference is song dependent. For example in Fig. 10, Song 3 is rated the highest for Auto, while the same song is rated the lowest for Eng. 1.

Overall, the results show that the proposed automatic compression has very good performance based on various criteria.

6 CONCLUSION

In this paper we have proposed a novel intelligent multitrack dynamic range compression algorithm. The algorithm utilizes the CA-DAFX processing architecture [2, 3], exploits the interdependence of the input audio features, and incorporates best practices as well as subjective evaluation results to produce the optimal amount of dynamic range compression for multitracks.

To the best of the authors' knowledge, this presents the first fully automated multitrack dynamic range compressor where all classic parameters of a typical compressor (ratio, threshold, knee, attack, and release) are dynamically adjusted depending on extracted features and control rules.
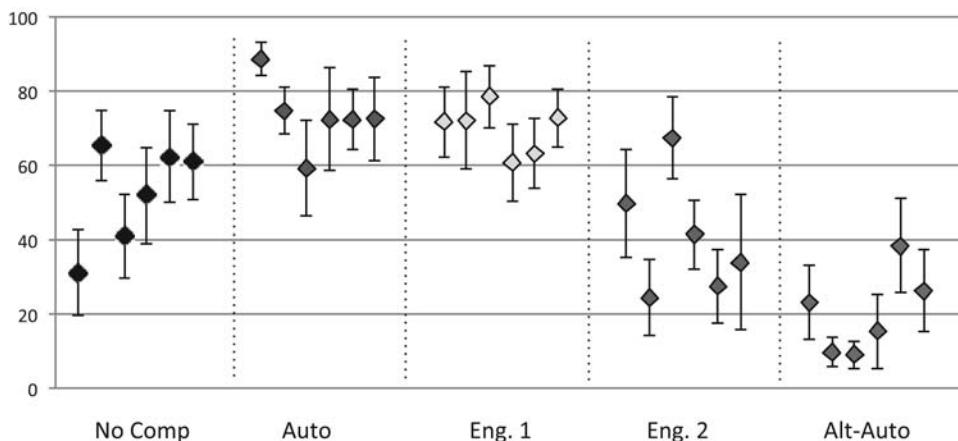

Fig. 11. Averaged results of Q4: overall preference with 95% confidence intervals, grouped by mix type
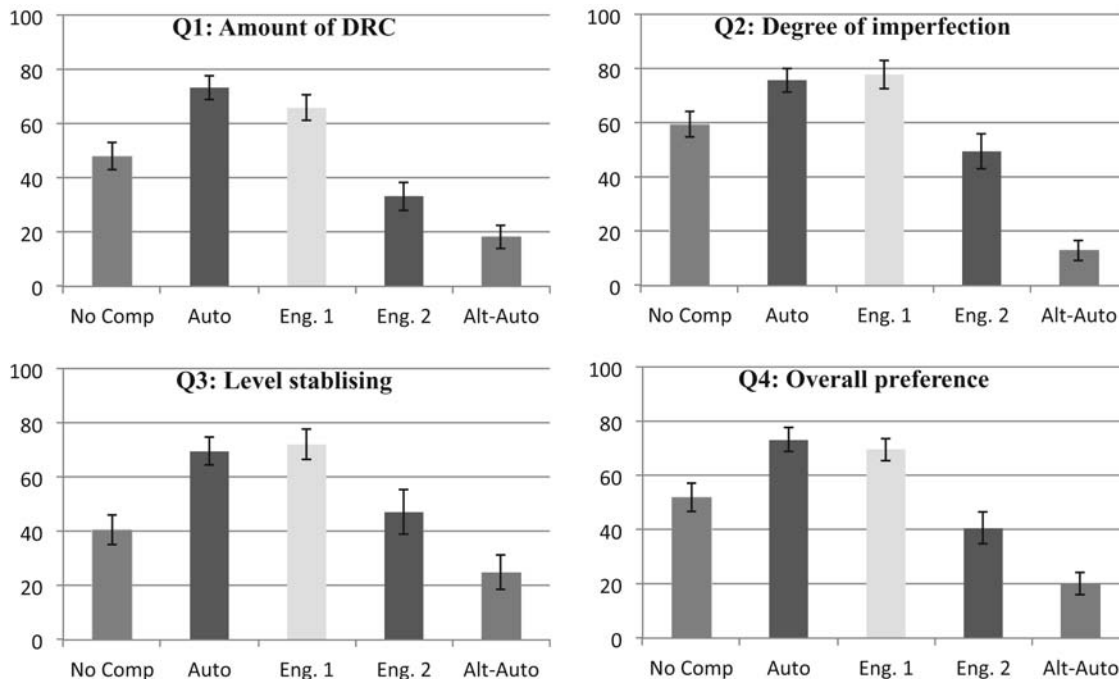
Fig. 12.   Overall mean results with 95% confidence intervals for Q1-Q4 grouped by mix type

In the pursuit of intelligent algorithms, two new audio features, namely percussivity weighting and low-frequency weighting, were proposed to describe the transient nature and spectral content of the signal. A method of adjustment experiment was conducted to investigate the relationship between human preference for ratio and threshold and the proposed audio features in multitrack content. We applied multiple linear regression models to the subjective results to formulate the ratio and threshold automations that follow the choices of the human operators.

The output mix produced by the proposed algorithm has an outstanding performance in the final subjective evaluation when compared against a raw mix, two semi-professional mixes, and a previous automatic compression approach. The results showed that the algorithm is able to compete with or outperform the semi-professional mixes in terms of four different perceptual criteria: the appropriateness of the amount of DRC applied, the degree of imperfection, ability to stabilize the erratic level fluctuations, and overall preference. However, evaluation with a larger number of subjects of wider range of background is needed to more formally assess the performance of the proposed algorithm.

Subjective evaluation results have shown that spectral content plays an important role in the pursuit of an intelligent solution to dynamic processing. The authors plan to expand the algorithm to a more general intelligent frequency and dynamic processing system, by providing a marriage between dynamic processing and equalization processing to eliminate the shortcomings of static EQ and conventional dynamic processors.

All songs used in this work can be accessed from the Open Multitrack Testbed at   multitrack.eecs.qmul.ac.uk [26].

7 ACKNOWLEDGMENT

8 REFERENCES

[1] R. Izhaki, *Mixing Audio: Concepts, Practices and Tools* (Focal Press, 2013).

[2] U. Zolzer, ed., "Adaptive Digital Audio Effects," in *DAFX: Digital Audio Effects*, 2nd Ed., p. 321 (Wiley, 2011).

[3] J. D. Reiss, "Intelligent Systems for Mixing Multi-channel Audio," *17th International Conference on Digital Signal Processing (DSP)*, pp. 1–6 (2011).

[4] G. W. McNally, "Dynamic Range Control of Digital Audio Signals," *J. Audio Eng. Soc.*, vol. 32, pp. 316–327 (1984 May).

[5] A. Schneider and J. V. Hanson, "An Adaptive Dynamic Range Controller for Digital Audio," *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 339–342, Victoria, Canada (1991).

[6] D. Giannoulis, M. Massberg, and J. D. Reiss, "Parameter Automation in a Dynamic Range Compressor," *J. Audio Eng. Soc.*, vol. 61, pp. 716–726 (2012 Oct.).

[7] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital Dynamic Range Compressor Design—A Tutorial and Analysis," *Jo. Audio Eng. Soc.*, vol. 60, pp. 399–408 (2012 Jun.).

[8] J. Parker and V. Valimaki, "Linear Dynamic Range Reduction of Musical Audio Using an Allpass Filter Chain," *Signal Processing Letters, IEEE*, vol. 20, pp. 669–672 (2013).

[9] T. Wilmering, G. Fazekas, and M. Sandler, "High Level Semantic Metadata for the Control of Multitrack Adaptive Audio Effects," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8766.

[10] J. A. Maddams, S. Finn, and J. D. Reiss, "An Autonomous Method for Multi-track Dynamic Range Compression," in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)* (2012).

[11] P. D. Pestana, "Automatic Mixing Systems Using Adaptive Audio Effects," Ph.D. thesis, Universidade Catolica Portuguesa (2013).

[12] P. Pestana and J. Reiss, "Intelligent Audio Production Strategies Informed by Best Practices," presented at the *AES 53rd International Conference: Semantic Audio* (2014 Jan.), conference paper S2-2.

[13] ITU-R, "*Rec. ITU-R BS.1770-1, Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level*," International Telecommunications Union (2007).

[14] E. Vickers, "Automatic Long-Term Loudness and Dynamics Matching," presented at the *111th Convention of the Audio Engineering Society* (2001 Nov.), convention paper 5495.

[15] J. Boley, C. Danner, and M. Lester, "Measuring Dynamics: Comparing and Contrasting Algorithms for the Computation of Dynamic Range," presented at the *129th Convention of the Audio Engineering Society* (2010 Nov.), convention paper 8178.

[16] P. D. Pestana, J. D. Reiss, and A. Barbosa, "Loudness Measurement of Multitrack Audio Content Using Modifications of ITU-R BS. 1770," presented at the *134th Convention of the Audio Engineering Society* (2013 May), convention paper 8813.

[17] P. H. Kraght, "Aliasing in Digital Clippers and Compressors," *J. Audio Eng. Soc.*, vol. 48, pp. 1060–1065 (2000 Nov.).

[18] D. M. Huber and R. E. Runstein, *Modern Recording Techniques* (CRC Press, 2013).

[19] N. Thiele, "Some Thoughts on the Dynamiics of Reproduced Sound," *J. Audio Eng. Soc.*, vol. 53, pp. 130–132 (2005 Jan./Feb.).

[20] N. Foudi, "A Preliminary Framework for Automatic Cross-Adaptive Compression," Masters thesis, University of Massachusetts Lowell (2012).

[21] O. Lartillot, P. Toiviainen, and T. Eerola, "A Matlab Toolbox for Music Information Retrieval," in *Data Analysis, Machine Learning and Applications*, pp. 261–268 (Springer, 2008).

[22] J. M. Lattin, J. D. Carroll, and P. E. Green, *Analyzing Multivariate Data* (Thomson Brooks/Cole Pacific Grove, CA, USA, 2003).

[23] ITUR BS Recommendation, "1534-1, Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems (MUSHRA)," International Telecommunication Union (2003).

[24] B. De Man, B. Leonard, R. King, and J. D. Reiss, "An Analysis and Evaluation of Audio Features for Multitrack Music Mixtures," *15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (2014 Oct.).

[25] S. Mansbridge, S. Finn, and J. D. Reiss, "Implementation and Evaluation of Autonomous Multi-Track Fader Control," presented at the *132nd Convention of the Audio Engineering Society* (2012 Apr.), convention paper 8588.

[26] B. De Man, M. Mora-Mcginity, G. Fazekas, and J. D. Reiss, "The Open Multitrack Testbed," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), eBrief 165.

THE AUTHORS

| Zheng Ma | Joshua Reiss | Brecht De Man | Pedro Duarte Pestana | Dawn Black |

Zheng Ma is currently a Ph.D. student at the Centre for Digital Music at Queen Mary University of London (QMUL), specializing in intelligent mixing tools for multitrack frequency and dynamic processing. He received a Bachelor degree with First Class Honours in telecommunication engineering with management from Beijing University of Posts and Telecommunications and QMUL. He had been an active member of the Audio Engineering Society and was the secretary for the London Student Section in 2014. He has been awarded a scholarship under the Chinese State Scholarship Fund for four years to pursue his research in the United Kingdom.

●

Dr. Josh Reiss is a Senior Lecturer with the Centre for Digital Music at Queen Mary University of London. He received his Ph.D. in physics from Georgia Tech, specializing in analysis of nonlinear systems. Dr. Reiss has published over 150 scientific papers and serves on several steering and technical committees and is co-founder of the company MixGenius. He has investigated music retrieval systems, time scaling and pitch shifting techniques, polyphonic music transcription, loudspeaker design, automatic mixing, and digital audio effects, among others. His primary focus of research, which ties together many of the above topics, is on the use of state-of-the-art signal processing techniques for professional sound engineering.

●

Brecht De Man is currently a Ph.D. student with Josh Reiss at the Centre for Digital Music at Queen Mary University of London. He received a Master degree in electronic engineering from the University of Ghent, Belgium, in 2012. Since the start of his Ph.D. he has published and presented research on the perception of record-

ing and mixing engineering, intelligent audio effects, and the analysis of music production practices. Brecht is also an experienced musician and recording engineer. An active member of the Audio Engineering Society, he is Chair of the AES Student Delegate Assembly for Europe and International Regions, Chair of the London UK Student Section, committee member of the British Section of the AES, and as such involved in the organization of various past and future AES events. In 2013, and again in 2014, he received the HARMAN Scholarship from the Educational Foundation of the Audio Engineering Society.

●

Pedro Duarte Pestana is currently a professor at the School of Arts at the Catholic University of Portugal (UCP) and a researcher at the Center of Science and Technology of the Arts (CITAR) of the same institution. He also works as lead audio research at Mixgenius, a company that specializes in intelligent audio production software. He received his Ph.D. in computer music at UCP, specializing in adaptive digital audio effects for automatic mixing. He has investigated topics pertaining to perception and cognition in audio, machine learning systems, acoustics, interactive sound design, and digital audio effects. He has been an active member of the Audio Engineering Society for over a decade and won the best paper award at the 134th AES Convention in Rome.

●

Dawn Black is currently employed as audio algorithm engineer in a new music instrument company called ROLI, after lecturing at the Centre for Digital Music at Queen Mary University of London, teaching and traveling between London and Beijing.