AN AUTONOMOUS METHOD FOR MULTI-TRACK DYNAMIC RANGE COMPRESSION

Jacob A. Maddams, Saoirse Finn, Joshua D. Reiss

Centre for Digital Music, Queen Mary University of London

London, UK

`jacob.maddams@gmail.com`

ABSTRACT

Dynamic range compression is a nonlinear audio effect that reduces the dynamic range of a signal and is frequently used as part of the process of mixing multi-track audio recordings. A system for automatically setting the parameters of multiple dynamic range compressors (one acting on each track of the multi-track mix) is described. The perceptual signal features loudness and loudness range are used to cross-adaptively control each compressor. The system is fully autonomous and includes six different modes of operation. These were compared and evaluated against a mix in which compressor settings were chosen by an expert audio mix engineer. Clear preferences were established for the different modes of operation, and it was found that the autonomous system was capable of producing audio mixes of approximately the same subjective quality as those produced by the expert engineer.

1. INTRODUCTION

Multi-track audio mixing is the production of a coherent sound mixture from multiple, individual audio sources, and is usually performed by skilled and experienced audio mix engineers. The signal processing operations routinely used in audio mixing include level balancing, spectral equalisation, spatial positioning and dynamic range compression (DRC). These and other processes are applied to individual tracks, or sub-groups of tracks, in order to produce an audio mixture in which the constituent sound sources are appropriately balanced, which sounds subjectively pleasing and which achieves a certain artistic intention.

The majority of digital audio effects (DAFX) are designed for single channel applications and process an audio input based on parameters which are specifically chosen by the user. Advances in digital signal processing have lead to the investigation of adaptive DAFX [1] in which the effect parameters are set automatically based on signal features with little or no user interaction. Yet very few of the DAFX currently available on the market are designed to automate any of the mixing process. The potential advantages of such DAFX are enormous – from allowing non-experts to produce quality mixes with little or no prior experience, to speeding up the workflow of professional mix engineers. For this reason, the development of tools designed to automate different audio mixing tasks ('intelligent multi-track DAFX') is a growing area of research.

Individual sound sources within a multi-track mix must be processed not in isolation but with respect to all other sound sources, or a subset thereof. Intelligent multi-track DAFX are, therefore, cross-adaptive inasmuch as the automatic control of one channel in the mix depends on features derived from other channels. Prior work has developed such systems for source enhancement, stereo panning, level and fader adjustment and spectral equalisation [2-6].

DRC has many applications when it comes to mixing multi-track audio. Typical examples include controlling the transient attack of percussive instruments such as drums, raising the overall loudness of a sound source by applying compression with make-up gain and providing a more consistent signal level [7]. The purpose of this paper is to develop and evaluate a system which, using high level, perceptual audio features, automatically and cross-adaptively applies DRC to individual tracks within a multi-track mix.

2. DYNAMIC RANGE COMPRESSION

DRC is a nonlinear audio effect that narrows the dynamic range (i.e. the difference between the loudest and quietest parts) of a signal. This is achieved by applying an attenuation to the signal whenever its level exceeds a given value. The set of parameters that can be used to describe a generic dynamic range compressor are threshold, ratio, knee, attack time, release time and make-up gain.

Threshold is the level above which the input signal is attenuated. Ratio is the amount of attenuation that is applied (for example, a ratio of 3:1 would result in a 1 dB increase in output signal level for every 3 dB increase in input signal level above the threshold). Knee is the dynamic range over which the ratio increases to its specified value (low values result in so-called 'hard' knees and high values result in 'soft' knees). Attack time is (approximately) the time it takes for the compressor to reach the desired attenuation ratio once the signal overshoots the threshold (or enters the knee region), and release time is (approximately) the time it takes for the compressor to return to a state of no attenuation once the signal returns back to a level below the threshold. Make-up gain is applied uniformly to the whole output signal after attenuation.

The basic goal when mixing multi-track audio is to ensure that the individual sound sources are blended together into a coherent-sounding whole, and modest amounts of DRC are particularly suitable for this task.

There are a number of different possible compressor design choices one can make when implementing DRC. In this paper, a digital implementation of a feed-forward monaural compressor with a smoothed de-coupled peak detector was used [8]. It should be noted, however, that the intelligent, multi-track dynamic range compression method described herein is independent of the compressor model.

3. LOUDNESS AND LOUDNESS RANGE

Two key signal features, loudness and loudness range, are used to control the application of DRC in this system. Although loudness is a subjective quality, objective measures have been recommended which approximate the characteristics of the human

hearing system. In this paper, the standard developed by the International Telecommunication Union, as described in ITU-R BS. 1770-2 [9], is used. Specifically, for a monaural signal, the loudness is defined as:

$$\text{Loudness} = -0.691 + 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^{N} y^2[i] \right) \qquad (1)$$

where $y[i]$ is the input signal after it is passed through a head-related transfer function filter and then a high pass filter. The unit of this loudness measurement is the LU (Loudness Unit), which is similar to the dB.

In general, short-term loudness measurements of a signal will vary over time. Loudness Range (LRA) quantifies the amount of this variation. It can be thought of as the perceptual equivalent of dynamic range, and is therefore of interest when considering intelligent DRC systems. A technical definition of LRA is given by the European Broadcasting Union (EBU) in Tech-3342 [10]. This specifies that loudness measurements are taken in sliding analysis windows of length 3 seconds with at least 66% overlap between consecutive windows. The resultant vector of short-term loudness measurements is then processed using a two-stage cascaded gating scheme. The first stage is an absolute gate set to -70 referenced relative to the maximum possible loudness of a digital signal (0 LU Full Scale). The second stage is a relative gate set to 20 LU below the integrated loudness of the absolute gated signal. 'Integrated loudness' is a measure specified in EBU Tech-3341 [11] and is intended to give a single overall loudness measurement for an entire piece of audio. LRA is then defined as the difference between the 10th and 95th percentiles of the twice-gated loudness measurements.

LRA is a natural signal feature to consider when looking at DRC but the 3 second time scale over which the EBU definition measures loudness may not be appropriate. DRC is often used to reduce the dynamic range of a signal over much shorter time scales, for example to reduce transient attack over time scales less than 50 ms. The most appropriate time scale to use in order to approximate human perception of dynamic range is still a matter for research [12]. In this paper, LRA is calculated using loudness measured in 400 ms sliding windows at a rate of 7.5 Hz (i.e. 67% overlap of consecutive windows). Windows of length 10 ms and 3 seconds were also considered. However, it was found that when using 10 ms windows LRA is almost always high, since at this time scale signal peaks are captured, and it is more a measure of instantaneous amplitude variation than loudness range. Conversely, when using 3 second windows, LRA tends to be low since typical musical audio signals are often relatively uniform at this time scale. 400 ms was therefore considered to be a good intermediate window length, since it was usually found to result in large variations in LRA across different tracks of any given multi-track audio recording.

4. SYSTEM DESIGN

The system was designed to produce a monaural output mix by automatically applying intelligent dynamic range compression with make-up gain to each individual track of a multi-track audio recording. The overall signal flow diagram is shown in Figure 1.

4.1. Pre-Gain Automation

The primary component of the system is the signal dependent and cross-adaptive automation of DRC. However, automation of the basic gain of each track in the mix is also included since this

is perhaps the most fundamental task of any mixing process and is vital if the output is to sound at all reasonable. The aim of this automation is to ensure that all tracks have equal loudness within the mix. Pre-gain is applied to individual tracks, before they enter the compressor, so that the integrated loudness of each is equal to the maximum integrated loudness of the tracks before adding the gain. Adding a gain of G dB to a digital signal is achieved by multiplying all samples by a factor of $10^{G/20}$. Occasionally, this will result in one or more samples with an absolute value greater than 1, i.e. the signal will be 'clipped'. If such clipping is detected on any of the tracks after this gain stage, then, to avoid distortion, cross-adaptive normalisation is applied by multiplying all tracks by $1/\max\{x_{\text{clip}}[n]\}$, where $x_{\text{clip}}[n]$ is the post-gain signal with the highest clipping level. This process ensures that clipping is avoided but equal loudness between all tracks is maintained.
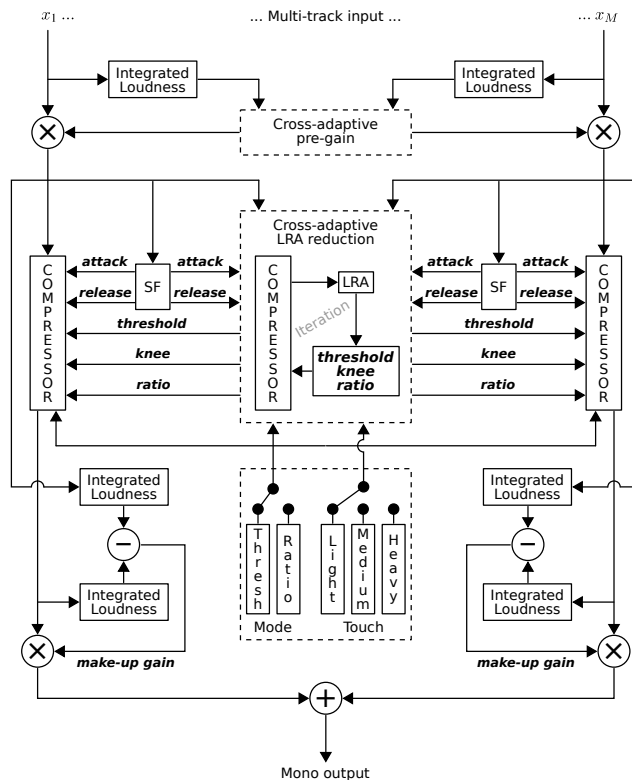

Figure 1: *Signal flow diagram of the Intelligent Multi-Track Dynamic Range Compression system.*

4.2. Compressor Automation

In order to minimize undesirable DRC artefacts (such as 'breathing', 'pumping' and 'drop outs' [13]), attack and release times are automated using the spectral flux of the signal [14-16]. This approach is described in detail by Giannoulis *et al* [17]. Two separate modes of operation are defined to automate the compressor's threshold, ratio and knee parameters.

4.2.1. 'Threshold' Mode

In 'threshold' mode, the compressor ratio is fixed at ∞:1 and the knee is set to the absolute value of the threshold. The amount of DRC is then controlled with the threshold parameter alone. As

the threshold is lowered, the compressor is triggered more frequently and the knee becomes wider (or 'softer'). See Figure 2A for an illustration of the input-output characteristics of a compressor in 'threshold' mode.

4.2.2. 'Ratio' Mode

In 'ratio' mode, the amount of DRC is controlled via the ratio parameter alone. A fixed moderately hard knee of 3 dB was chosen for this mode. In general, a signal with a higher root mean square (RMS) level requires a higher threshold in order to have the same amount of DRC applied as a signal with a lower RMS. Therefore, in 'ratio' mode, the compressor threshold is fixed relative to the RMS of the signal. After some investigation, 12 dB below the RMS was found to be a suitable level. The input-output characteristics of a compressor in this mode are shown in Figure 2B.
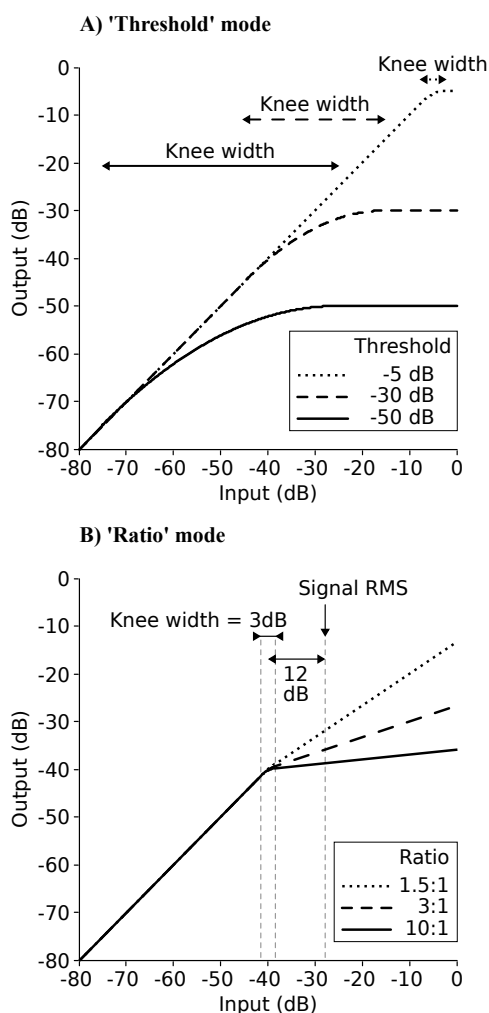

Figure 2: *Input-output characteristics of an automatic dynamic range compressor in A) 'threshold' mode and B) 'ratio' mode.*

4.3. Make-up Gain Automation

Make-up gain is included in most compressor designs to allow the overall loudness of the output and input signals to be bal-

anced. In this system, it is automated such that the integrated loudness of the compressor output is equal to the integrated loudness of the input and cross-adaptive normalisation is used to avoid clipping, as described in section 4.1.

4.4. Cross-Adaptive Control

The appropriate amount of compression for each individual track is automatically and cross-adaptively determined by analysing the LRA of all tracks in the mix. Two basic hypotheses were formulated *a priori* in order to design the cross-adaptive algorithm: 1) that DRC reduces LRA in a roughly monotonic way, and 2) that successful multi-track DRC helps to produce a coherent mix of sound sources by compressing tracks with higher LRA more than those with lower LRA, such that the difference between the highest and lowest track LRA is reduced.

Experiments were carried out to quantify the effect of DRC on LRA in each of the two compressor modes ('threshold' and 'ratio'). Seven multi-track audio recordings, covering a variety of different genres, were used. These were the same recordings used later for the subjective listening tests (see section 5.1). Each recording had up to 7 individual tracks, giving a total of 41 individual signals for testing. For each signal, LRA was calculated before and after applying DRC. In 'ratio' mode, the compressor ratio was varied from 1:1 to 10:1. In 'threshold' mode, the compressor threshold was varied between –25 dB and +25 dB relative to the signal's RMS.

The cascaded gate used when calculating LRA is designed to discount the noise floor of the signal and sections of silence. Applying make-up gain, particularly when the amount of DRC is high, can cause the noise floor to be amplified to the extent that it passes through the gate and is included in the LRA calculation. This is undesirable and can lead to anomalous LRA measurements. Therefore, the *active sections* of a signal were defined as those which contribute to the pre-DRC LRA measurement. Post-DRC LRA was then calculated based on the active sections only.

Figure 3 shows how different amounts of DRC affected the LRA measurements. The results in Figure 3A were obtained using an auto-compressor in 'threshold' mode, so that a lower threshold (moving left to right on the graph) corresponds to an increased amount of DRC. Averaged over all 41 test signals, the change in LRA was found to vary smoothly with threshold and, as expected, LRA decreased as the compressor threshold decreased. However, there was wide variation across the test signals. This comes from the fact that the maximum achievable amount of absolute LRA reduction is dependent on the LRA of the pre-DRC signal itself. It also appeared that, for many signals, there was a lower limit below which it was not possible to reduce LRA via DRC alone. Similar results were observed using an auto-compressor in 'ratio' mode (see Figure 3B).

So it was verified that, for most signals, DRC can be expected to reduce LRA in a monotonic fashion, i.e. LRA decreases as the amount of DRC increases. However, it was found that this is certainly not true in all cases and there does not seem to be an obvious way of predicting, even approximately, to what extent LRA of a given signal will change after applying DRC.

Let's define the LRA range (denoted ΔLRA) of a multi-track recording as the difference between the highest and lowest LRA of individual tracks. Using the hypothesis that DRC improves the quality of multi-track mixes by reducing ΔLRA, a cross-adaptive algorithm was developed to automate the remaining single control parameter of the compressor (ratio if in 'ratio' mode, or threshold if in 'threshold' mode). The overall amount of DRC

that is appropriate for a multi-track audio mix is largely a matter of personal taste and, therefore, three different levels of 'touch' were defined for the system: a 'light' touch results in an overall reduction in ΔLRA of 3 LU, a 'medium' touch reduces ΔLRA by 6 LU and a 'heavy' touch reduces ΔLRA by 9 LU.
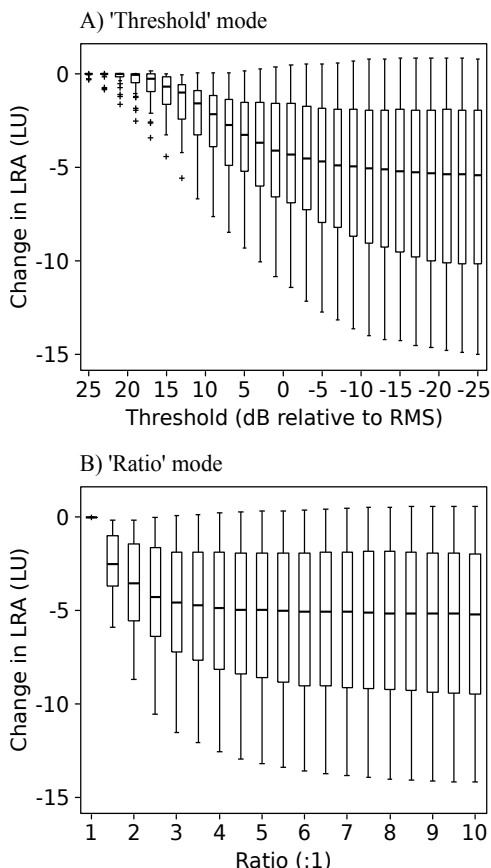
A) 'Threshold' mode

B) 'Ratio' mode

Figure 3: *Post-DRC loudness range using an auto-compressor in A) 'threshold' mode and B) 'ratio' mode. Median change in LRA with 25th and 75th percentiles (boxes), full extent (whiskers) and outliers (crosses).*

The cross-adaptive algorithm is as follows: no DRC is applied to the track with the lowest LRA; the largest reduction of LRA is sought for the track with the highest pre-DRC LRA; and the LRAs of the remaining tracks are reduced proportionally.

Specifically, for each track in the mix, a target LRA (i.e. the ideal post-DRC LRA) is defined as:

$$LRA_T(i) = LRA(i) - \Delta LRA_{\text{red}} \left(\frac{LRA(i) - LRA(i_{\min})}{LRA(i_{\max}) - LRA(i_{\min})} \right) \quad (2)$$

where i is the index number of the track, $LRA(i)$ is the pre-DRC loudness range of track i, $i_{\min} = \text{argmin}_i(LRA(i))$, $i_{\max} = \text{argmax}_i(LRA(i))$ and ΔLRA_{red} is the amount by which ΔLRA is to be reduced (dependent on the touch parameter).

Since it is not possible to know in advance precisely which compressor settings will result in the required LRA reduction for each track, these are found by iteration. Starting values are based on the empirical data presented in Figure 3 (using the average reduction in LRA for a given ratio or threshold). In 'ratio' mode, the best ratio is found to the nearest 0.1 dB. In 'threshold' mode, the best threshold is found to the nearest 0.5 dB. Figure 4 shows

an example of the target and achieved LRA reductions for tracks in a multi-track mix using Eq. 2, with different compressor modes and touches.
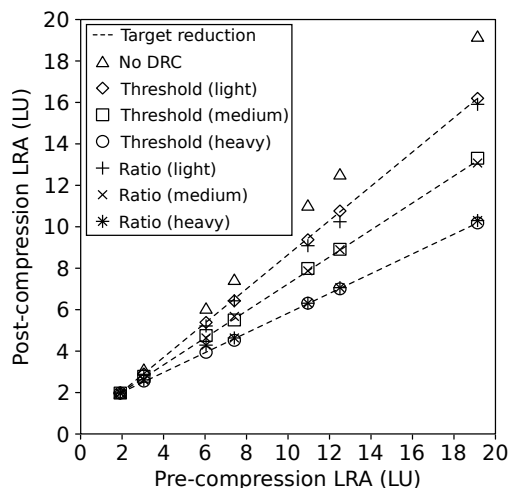
Figure 4: *Example of pre- and post-DRC LRA for a multi-track audio recording using auto-compressors in 'ratio' or 'threshold' mode with three different touches (light, medium and heavy).*

5. EVALUATION

The system described above has six different forms of operation defined by the choice of compressor mode ('threshold' or 'ratio') and touch (light, medium or heavy). Ultimately, it may be desirable to have a user-defined 'touch' parameter within the system, so that the overall amount of DRC can be broadly controlled manually. However, it may be less desirable to retain an option regarding the mode of the auto-compressor since the difference between 'threshold' and 'ratio' modes would not be at all obvious to the average user. For this reason, and to investigate the performance of the automatic system compared with a manual application of multi-track DRC, subjective evaluations were carried out. The six different forms of automatic operation were compared together with a mix in which DRC was applied manually and a mix in which no DRC was applied.

5.1. Multi-Track Test Signals

Seven multi-track audio recordings were used to test and evaluate the system. They covered a range a musical genres: four Rock/Indie songs, two instrumental Jazz pieces and one Acoustic Folk song. For each recording, a 20 second excerpt was chosen manually based on the following criteria:

1. The excerpt should be representative of the whole recording: it could be a section of a verse, a chorus or a transition, but all tracks should be active (i.e. all instruments included in the recording should be playing) for most of the excerpt.

2. It must be possible to reduce the ΔLRA of the excerpt by at least 9 LU using our approach, i.e. the pre-DRC ΔLRA must be at least 9 LU and all of the target LRA reductions must be successfully achieved in both compressor modes for all touch settings.

3. Different automatic mixes (using different compressor modes and touch settings) should sound audibly different, to the

extent that a non-expert listener would, with careful listening, be able to differentiate between them.

The test recordings were obtained from an online resource [18] and were downloaded as completely raw stems, i.e. no mixing, effects or post-recording processing had been applied already. The recorded instruments included: electric/acoustic guitar, electric bass guitar, double bass, piano, violin, acoustic drum kit and vocals. The drum kit tracks were sub-mixed before processing; one sub-mix for all the close drum microphones and one for overhead, room and cymbal microphones. Similarly, if a particular instrument had been 'double tracked', i.e. the same part recorded twice, then these were sub-mixed to a single track. This reduced the total number of tracks per recording to between three and seven.

The level of each track was set using an equal loudness algorithm, as described in section 4.1. However, five of the seven test recordings contained lead vocal tracks. It is common for such tracks to be slightly louder in the mix than others [19], and so each lead vocal was boosted by 3–6 dB relative to its equal loudness level. The precise amount of boost was determined manually for each recording and was the same for all mixes.

5.2. Subjective Listening Test

A subjective listening test was designed to evaluate the quality of the different mixes with intelligent compression in relation to each other, an automatic mix with no DRC (i.e. with only gain automated) and a mix in which DRC settings were chosen manually by an audio engineer with extensive experience of multi-track audio mixing in both studio and live sound environments. To allow the manual mix to be completed, a graphical user interface was developed following the conventions of a standard software Digital Audio Workstation, e.g. vertical track layout, waveform views and soloing capability (Figure 5).
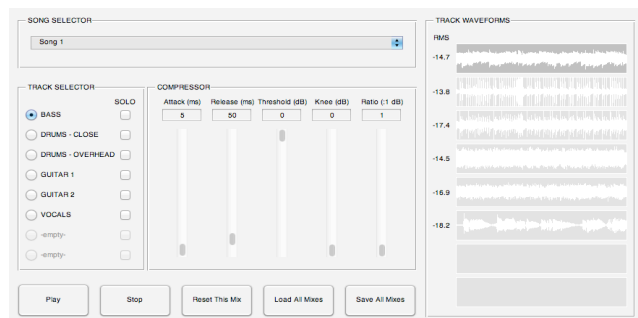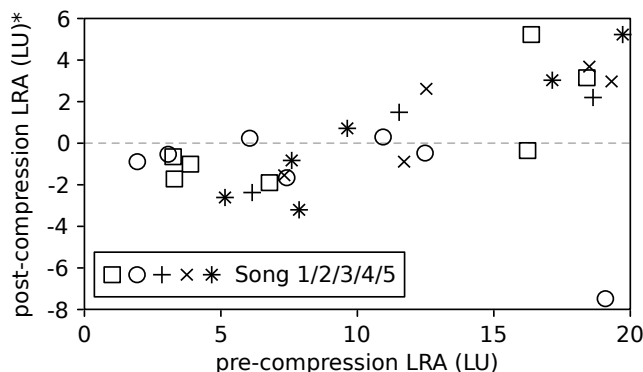
Figure 5: *Graphical User Interface for the manual application of multi-track DRC.*

The same compressor implementation [8] was used as for the automatic mixes. Attack, release, threshold, knee and ratio could be set manually for a compressor on each track, but the gain was automated in the same way as for the fully automatic mix. This will be referred to as the 'expert manual' mix.

A comparison of the LRA reductions per track achieved by the expert manual mix settings compared to the target LRA reductions of the automatic system with a medium touch is shown in Figure 6. It can be seen that the LRA reductions were, in general, quite similar. However, compared with the automatic system, the expert mixer applied more compression to the tracks with the lowest pre-compression LRA, and less to the tracks with the highest pre-compression LRA.

*Relative to target loudness range with medium touch

Figure 6: *Comparison of loudness range reductions for manual and automatic (medium touch) multi-track dynamic range compression. Different symbols represent tracks from different songs.*

All of the expert manual mixes except one resulted in a reduction in ΔLRA (see Table 1). This provides some evidence in support of the hypothesis used to design the cross-adaptive control algorithm (see section 4.4). The average reduction in ΔLRA achieved by the manual mixes was 2.3 LU, just below that which is achieved by the automatic system with a light touch (3 LU).

Table 1: *Change in ΔLRA after expert manual application of multi-track DRC.*

Song number (name)	Pre-DRC ΔLRA	Post-DRC ΔLRA	Change in ΔLRA
1 (Rock/Indie 1)	11.1 LU	5.8 LU	-5.2 LU
2 (Rock/Indie 2)	11.2 LU	11.1 LU	-0.1 LU
3 (Rock/Indie 3)	15.2 LU	14.9 LU	-0.3 LU
4 (Rock/Indie 4)	17.3 LU	7.3 LU	-9.9 LU
5 (Instrumental Jazz 1)	12.5 LU	11.1 LU	-1.4 LU
6 (Instrumental Jazz 2)	12.0 LU	10.8 LU	-1.2 LU
7 (Acoustic Folk)	14.6 LU	16.5 LU	+1.8 LU
Average	---	---	**-2.3 LU**

A design similar to the 'Multiple Stimuli with Hidden Reference and Anchor' (MUSHRA) test was used [20]. Participants were asked to rate each mix on a scale of 0 (very bad) to 100 (excellent), based on specific criteria. There were three tests in total, each with a different criterion for evaluation. The MUSHRAM Matlab interface [21] was used to administer the tests (see Figure 7) but was altered to allow playback of each mix to be interrupted. The MUSHRA test design was originally developed primarily for the evaluation of audio codec quality and specifies the inclusion of an objectively high quality 'reference' and an objectively low quality 'anchor'. However, since there is no obvious way of defining an objectively bad application of DRC within a multi-track mix, no anchor was used in these listening tests. Tests 1 and 2 used a reference sample, but test 3 did not. All mixes (both automatic and manual) were normalised to overall equal integrated loudness in order to avoid bias related to subjective preference for louder or quieter signals. For each test, four multi-track recordings were used from a range of genres and, for each of these, the order of presentation of the mixes was randomised. 15 participants were recruited (13 male, 2 female) and pre-screened to ensure that they all had no hearing impairments, were familiar

with the concept and typical sound of DRC and had experience of analysing and listening critically to audio. The tests were administered under controlled conditions in a good listening environment. Post-screening of participants was conducted by analysing the correlation between each participant's scores and the median of the scores from all participants. Pearson's correlation and Spearman's rank correlation were calculated to identify potential outliers. Then, after manual inspection of each participant's ability to award consistent grades, one participant was excluded from the test 1 results and two were excluded from the test 2 results. No participants were excluded from the results of test 3.
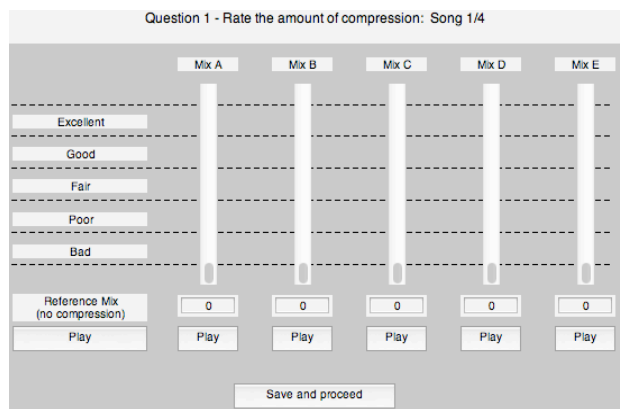

Figure 7: *Graphical User Interface for the subjective listening tests*.

Test 1 was designed to evaluate subjective preference for the overall amount of DRC that was applied. Different levels of the touch parameter were used with the compressor mode fixed to 'threshold'. Two Rock/Indie recordings, one Jazz and one Acoustic Folk recording were used as the test material. The 'no DRC' mix was used as the (hidden) reference – i.e. it was labelled as the 'reference' but also included without a label amongst the other test mixes for evaluation. The 'expert manual' mix was also included. Participants were instructed to "*rate the following according to the appropriateness of the relative amounts of dynamic range compression applied to each individual sound source in the mix*" and were required to score at least one mix in the 'bad' category.

The results of this test are shown in Figure 8. Relative scores for the different mixes were fairly consistent across all four recordings. It was only the threshold–light automatic mix that scored higher overall than the no DRC mix, indicating that participants did not think that very much DRC was required. The threshold–heavy mix scored significantly lower than all other mixes and was the only one to be rated bad overall. The fact that the no DRC mix scored relatively highly is also perhaps indicative of the difficulty in choosing appropriate DRC settings, and the extent to which badly chosen parameters can seriously degrade the quality of a mix.

Test 2 was designed to evaluate subjective preference for the two different compressor modes. The compressor mode was therefore varied but, since the amount of DRC was not being evaluated, the touch parameter was fixed; a heavy touch was chosen to provide the most audible differences between mixes. A different set of songs, but from the same genres as test 1, were used as the test material. Once again, the 'no DRC' (hidden) reference mix and the 'expert manual' mix were also included. Participants were asked to "*rate the following according to the*

sound quality of the dynamic range compression applied to each individual sound source in the mix".

The results of this test are shown in Figure 9. Overall, the expert manual mix scored the highest and both automatic mixes were rated poor. The sound quality of the ratio–heavy mix was preferred only slightly to the threshold–heavy mix. It is clear from tests 1 and 2 that the heavy touch automatic mixes were not favoured by the participants, most likely due to the audio quality being severely compromised by excessive amounts of DRC.

Finally, test 3 was designed to be an overall general evaluation of all six different automatic mixes against each other, a mix with no added DRC and the 'expert manual' mix. The test material consisted of four recordings from the Rock/Indie genre. The 'no DRC' mix was included, but not as an explicit reference. Participants were asked to "*rate the following according to the overall quality of the mix*."

The results of this test are shown in Figure 10 and are fairly consistent with those from the previous two tests. The heavy touch automatic mixes had the lowest scores and, overall, the two light touch automatic mixes and the expert manual mix scored highest. The expert manual and threshold–light mixes were virtually tied with the top scores, but the ratio–light mix scored only slightly higher than the mix with no DRC.

In this test, as in test 2 in which ratings were based on the "sound quality" of DRC (Figure 9), the expert manual mix of the 'Rock/Indie 4' song scored particularly highly compared with the automatic mixes. An examination of the manual mix compressor settings showed that they were largely similar to the automatic settings on all tracks, except one of the vocals. The expert had heavily compressed this track and reduced its LRA by 13.5 LU (by comparison, the heavy touch automatic mixes reduced its LRA by just 9 LU). This was clearly a successful strategy and is an example of the difficulties of designing an automatic system that is capable of consistently out-performing or matching the expertise and listening experience of a human engineer.
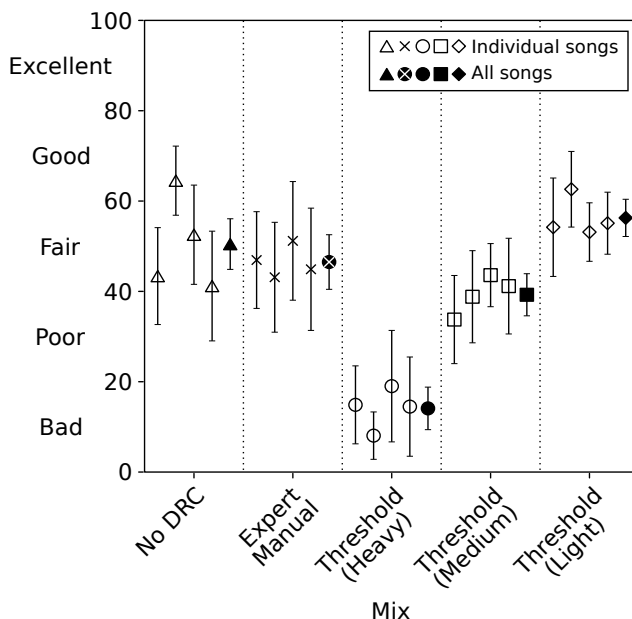

Figure 8: *Mean scores and 90% confidence intervals for "appropriateness of the relative amounts of dynamic range compression applied to each individual sound source in the mix" listening test.*
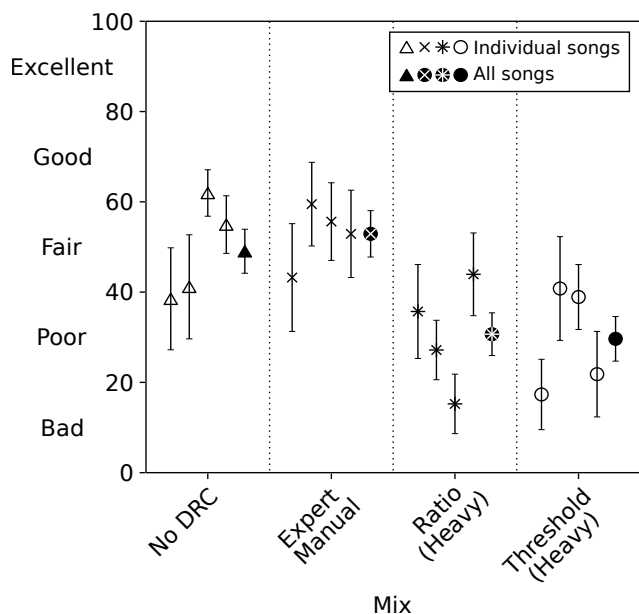
Figure 9: *Mean scores and 90% confidence intervals for "sound quality of the dynamic range compression applied to each individual sound source in the mix" listening test.*
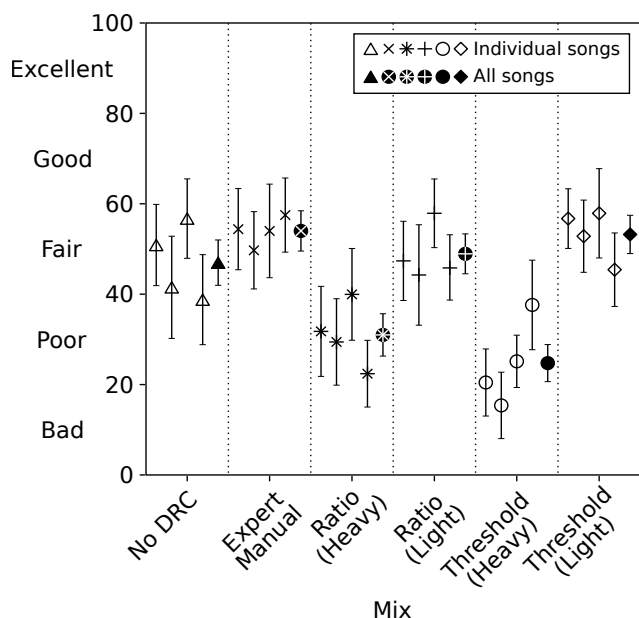

Figure 10: *Mean scores and 90% confidence intervals for "overall quality of the mix" listening test.*

6. CONCLUSIONS AND FUTURE WORK

The system described in this paper uses the perceptual audio features loudness and loudness range to automatically and cross-adaptively control multi-track DRC. The control strategy is based on the *a priori* hypothesis that the fundamental role of DRC in multi-track audio mixes is to reduce the difference between the highest and lowest individual track LRA, and that sound sources with higher LRAs require greater amounts of DRC. This hypothesis was substantiated empirically by examining the post-DRC changes in LRA achieved when an experienced mix engineer chose the compressor settings manually.

A number of different modes of automatic operation were designed and evaluated using a subjective listening test. Automatic mixes using a compressor in 'ratio' mode were not consistently rated higher than the 'threshold' mode mixes, but when using a heavy touch there was some evidence that 'ratio' mode was preferred. A light touch was found to be the most subjectively appropriate; both the ratio–light and threshold–light mixes performed very well when compared to the expert manual mixes. Indeed, the average reduction in ΔLRA achieved by the manual mixes was very similar to that achieved by the light touch automatic mixes. The best performing automatic mode of operation was threshold–light, and there was good evidence to suggest that, in this mode, the system is capable of automatically applying multi-track DRC to at least the same subjective standard as would be achieved if settings were chosen manually by an experienced mix engineer.

However, there is a need for further listening tests to establish more concretely which settings are preferred. Two of the most striking aspects of the listening test results were the generally low scores (no mix was rated above 'fair' on average) and the consistency with which the 'no DRC' mix was rated as one of the best. There are a number of possible explanations for this. For example, the differences between the mixes may not have been clearly audible and the overall sound quality of the recordings themselves may not have been considered high. Other choices of multi-track recordings may, therefore, have yielded more conclusive results, or higher ratings overall.

In addition, some of the strategies for automating individual compressor parameters (section 4.2) were based on prior research [17] which focussed on the automation of a single compressor with the goal being to improve the ease with which DRC can be applied by reducing the required user input to a minimum. This fact may have also had an effect on the listening test results and the success of this system. Further examination of these strategies is required to determine whether they could be improved and optimised for multi-track DRC applications. A more sophisticated knee automation strategy (for example, as described in [17]) could also be employed.

Future areas of research could concentrate on gaining a greater understanding of perceptual dynamic range. In this paper, the EBU definition of LRA was adapted somewhat arbitrarily to measure the variation in 400 ms (rather than 3 second) short-term loudness. The way in which dynamic range is perceived, and how this varies with different instruments, genres, etc., is not currently well understood, but this knowledge would greatly benefit the kind of intelligent system described here.

The fundamental assumptions used in this system were that a reduction in ΔLRA would be preferred by listeners, and that the best way to achieve this is by applying DRC to each track in proportion to its pre-DRC LRA. The target LRA reductions are then defined by equation (2). However, there are many other possible approaches one could use. For example, equation (2) could be modified so that DRC is applied to all tracks in the mix, rather than all but one. Or the target LRA reduction could depend on additional signal features, or equal LRA across all tracks could be sought (i.e. ΔLRA = 0). There are many such alternative strategies which are certainly worthy of further investigation.

7. REFERENCES

[1] V. Verfaille, U. Zölzer and D. Arfib, "Adaptive Digital Audio Effects (A-DAFx): A New Class of Sound Transformations," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1817–31, 2006.

[2] E. Perez Gonzalez and J.D. Reiss, "Automatic equalization of multi-channel audio using cross-adaptive methods," in *Proceedings of the 127th Audio Engineering Society Convention*, New York, NY, USA, October 2009, paper 7830.

[3] E. Perez Gonzalez and J.D. Reiss, "Automatic gain and fader control for live mixing," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, October 18–21, 2009.

[4] E. Perez Gonzalez and J.D. Reiss, "A real-time semi-autonomous audio panning system for music mixing," *EURASIP Journal on Advances in Signal Processing*, 2010.

[5] E. Perez Gonzalez and J.D. Reiss, "Automatic Mixing," in *DAFX – Digital Audio Effects* by U. Zölzer (ed.), John Wiley, Chichester, UK, second edition, 2011.

[6] S.P. Mansbridge and J.D. Reiss, "Implementation and evaluation of autonomous multitrack fader controls for automatic mixing," in *Proceedings of the 132nd Audio Engineering Society Convention*, Budapest, Hungary, April 2012.

[7] M. Senior, "Mixing Secrets for the Small Studio," Focal Press, Oxford, UK, 2011.

[8] D. Giannoulis, M. Massberg and J.D. Reiss, "Digital dynamic range compressor design – A tutorial and analysis," to appear in the *Journal of the Audio Engineering Society*, June 2012.

[9] International Telecommunication Union, "Recommendation ITU-R BS.1770-2 – Algorithms to measure audio programme loudness and true-peak audio level," Available at http://webs.uvigo.es/servicios/biblioteca/uit/rec/BS/R-REC-BS.1770-1-200709-I!!PDF-E.pdf, Accessed February 12, 2012.

[10] European Broadcasting Union, "TECH 3342 – Loudness Range: A descriptor to supplement loudness normalisation in accordance with EBU R128", Available at http://tech.ebu.ch/publications, Accessed February 12, 2012.

[11] European Broadcasting Union, "TECH 3341 – Loudness metering: 'EBU Mode' metering to supplement loudness normalisation in accordance with EBU R 128", Available at http://tech.ebu.ch/publications, Accessed February 12, 2012.

[12] J. Boley, M. Lester and C. Danner, "Measuring Dynamics: Comparing and contrasting algorithms for the computation of dynamic range," in *Proceedings of the 129th Convention of the Audio Engineering Society*, San Francisco, CA, USA, November 4–7, 2010.

[13] B. Katz, "Mastering Audio: The Art and the Science," Focal Press, Boston, MA, USA, 2007.

[14] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M.B. Sandler, "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 1035-47, 2005.

[15] M. Goodwin and C. Avendano, "Enhancement of Audio Signals Using Transient Detection and Modification," in *Proceedings of the 117th Audio Engineering Society Convention*, San Francisco, CA, USA, October 2004.

[16] R. Zhou and J.D. Reiss, "Music Onset Detection," in *Machine Audition: Principles, Algorithms and Systems* by W. Wang (ed.), IGI Global, 2009.

[17] D. Giannoulis, M. Massberg and J.D. Reiss, "Parameter automation in a dynamic range compressor," Submitted to the *Journal of the Audio Engineering Society*, Available at http://www.elec.qmul.ac.uk/digitalmusic/audioengineering/compressors/, Accessed February 12, 2012.

[18] M. Senior, "The 'Mixing Secrets' Free Multitrack Download Library," Available at http://www.cambridge-mt.com/ms-mtk.htm, Accessed February 7, 2012.

[19] R. Izhaki, "Mixing Audio: Concepts, Practices and Tools," Focal Press, Oxford, UK, second edition, 2012.

[20] International Telecommunication Union, "Recommendation ITU-R BS.1534-1 – Method for the subjective assessment of intermediate quality level of coding systems," Available at http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-1-200301-I!!PDF-E.pdf, Accessed February 12, 2012.

[21] E. Vincent, "MUSHRAM – A Matlab Interface for MUSHRA listening tests," Available at http://www.elec.qmul.ac.uk/digitalmusic/downloads/#mushram, Accessed February 9, 2012.