Parameter Automation in a Dynamic Range Compressor

DIMITRIOS GIANNOULIS, MICHAEL MASSBERG, AND JOSHUA D. REISS, *AES Member*
(Dimitrios.Giannoulis@eecs.qmul.ac.uk) (michael@massberg.org) (josh.reiss@eecs.qmul.ac.uk)

Queen Mary University of London, London UK

Dynamic range compression is a nonlinear, time dependent audio effect. As such, preferred parameter settings are difficult to achieve even when there is advance knowledge of the input signal and the desired perceptual characteristics of the output. We introduce an automated approach to dynamic range compression where the parameters are configured automatically based on real-time, side-chain feature extraction from the input signal. Parameters are all dynamically varied depending on extracted features, leaving only the threshold as a user controlled parameter to set the preferred amount of compression. We analyze a series of automation techniques, including comparison of methods based on different signal characteristics. Subjective evaluation was performed with amateur and professional sound engineers, which established preference for dynamic range compressor parameters when applied to musical signals, and allowed us to compare performance of our various approaches against manual parameter settings.

0 INTRODUCTION[1]

A superfluously used dynamic range compressor suppresses musical dynamics and may produce lifeless or even boring recordings deprived of their natural sound and character. Mastering dynamic range compression and refraining from overusing it is not an easy task even for professional engineers, due to the versatility of the effect, together with the large number of choices regarding its use [1]. Being a nonlinear effect, if used carelessly it may alter a signal in unpredictable and undesired ways [1,2]. Setting up the compressor parameters in a sensible way is nontrivial because the effects of each parameter are not obvious and there is a high degree of correlation between the different parameters.

Automating the compressor parameters and, in general, the parameters of any audio effect [3,4, 5], can provide evident advantages to the user. Although not intended to replicate artistic choices, when the compressor is used to decrease dynamic range, automation will save users the trouble of properly setting the effect to avoid sound artifacts and in many cases it will give better results. In addition to this, for a highly diverse signal there might not be a static set of parameters that would be optimal. An automated compressor with parameters that dynamically adapt to the signal's characteristics may give better results than a set of static human preferences.

Research in automating the dynamic range compressor goes back many years [6] and is still active [7]. Compressors with partly automated parameters (such as "autorelease," for instance) have already found their way to production both as analogue and digital designs [2]. In some existing designs, the automation of the time constants is performed by observing the difference between the peak and RMS levels of the signal fed in the side-chain [1,8]. In [9] an RMS measurement was used to scale the release time constant. The RMS measurement, however, is always an absolute one and dependent on overall signal level. It does not directly take into account the transient nature of the signal.

The concept of replacing a user-controlled ratio and knee width with an infinite ratio and single, user controlled knee width has been used previously in both analogue and digital compressors, albeit with a static knee width [4]. Similarly, automatic make-up gain can be found in some compressor designs [1], but only as signal-independent static compensations that do not take into account loudness, even though the main purpose of make-up gain is to achieve the same loudness between the uncompressed and compressed signals.

Related work has been concerned with reverse engineering a compressor, based on analysis of the signal before and after compression [10,11]. However, the authors are not aware of any prior work that offers full automation of compressor parameters based on the input signal characteristics. Furthermore, we are not aware of any

[1] Accompanying audio samples and source code (in Matlab and as VST plug-ins) for the compressor designs and analysis are available from c4dm.eecs.qmul.ac.uk/audioengineering/compressors/

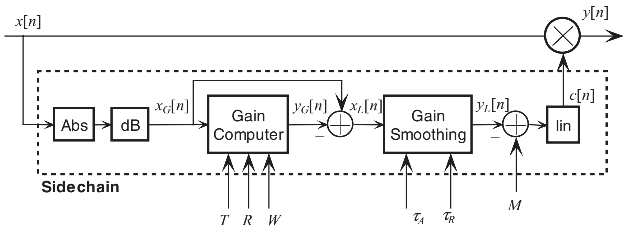

Fig. 1. Block Diagram of the Compressor Configuration.

previous listening tests or user studies regarding the effectiveness of a compressor automation approach, even for automating a single parameter within the design.

In this paper we suggest and evaluate a set of methods to automatically determine appropriate values for the standard compressor parameters in an intelligent way, depending on the input signal's properties and statistics, and the intended use of each parameter. This way the required user interaction is reduced to a minimum, ultimately to a single setting of how much compression is desired. In Section 1 we present a description of the parameters that will be automated and of the compression model that we will use. Section 2 presents a series of methods to effectively reduce the number of user-adjustable parameters. Section 3 includes a subjective evaluation of the various automation methods, with both amateur and professional sound engineers. Finally, Section 4 summarizes the results and provides suggestions for further research.

1 BACKGROUND

Our goal is to automate a dynamic range compressor, such that it can be operated or set with a single parameter. All other parameters are dependent on the input signal characteristics. Thus, we first describe the set of parameters of a typical compressor, and describe the compressor design that we will use.

1.1 Compressor Controls

Fig. 1 depicts a block diagram of the compressor configuration that was used. The compressor parameters define the behavior of the sidechain which determine the instantaneous compressor gain $c[n]$. The most commonly used compressor parameters may be defined as follows.

Threshold (denoted T) defines the level above which compression starts. Any signal overshooting the threshold will be reduced in level.

Ratio (R) controls the input/output ratio for signals overshooting the threshold level. It determines the amount of compression applied. As shown in Fig. 2, the ratio sets the slope of the static compression characteristic when the input level exceeds the threshold.

Attack and release times (τ_A and τ_R) provide a degree of control over how quickly a compressor acts. They are also known as time constants. Instantaneous compressor response is not sought because it introduces distortion on the signal. The attack time defines the time it takes the compressor to decrease the gain to the level determined by the ratio once the signal overshoots the threshold. The release time defines the time it takes to bring the gain back up to the normal level once the signal has fallen below the threshold.

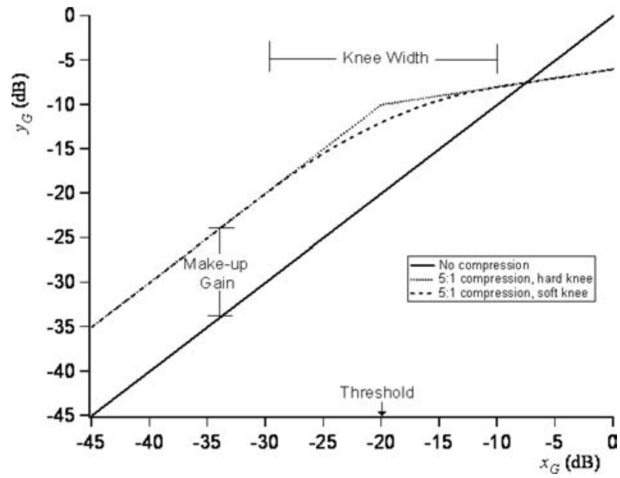

Fig. 2. Static compression characteristic with make-up gain and hard or soft knee.

A *Make-Up Gain* control is usually provided at the compressor output. The compressor reduces the level (gain) of the signal, so that feeding back a make-up gain to the signal allows for matching the input and output loudness level.

The *Knee Width* (W) option controls whether the bend in the static compression characteristic, depicted in Fig. 2, for input levels near the threshold has a sharp angle or has a rounded edge. A sharp transition is called a Hard Knee and provides a more noticeable compression. A softer transition where the ratio gradually grows from 1:1 to a set value in a transition region on both sides of the threshold is called a Soft Knee. It makes the compression effect less perceptible.

Optionally, a compressor may also have *look-ahead*, given in milliseconds. With look-ahead, the side-chain determines the amount of compression based on the current signal level, but the control is applied to a delayed version of the input signal. However, this introduces latency, and hence will not be used for our compressor design, which is intended to be applicable to live performance and broadcast. Also, our use of a peak detector (as opposed to RMS), ensures a quick response to changes in signal characteristics, hence lessening the need for look-ahead.

A compressor also has a set of additional controls that are sometimes found in modern designs. These include a Hold parameter, Side-Chain filtering, and many more.

1.2. The Compressor Model

The compressor model we employed is a feed-forward compressor with a smoothed decoupled peak-detector [12], whose output is given as the input signal times a control value determined by signal level estimation in a side-chain configuration:

$$y[n] = c[n] \cdot x[n] \tag{1}$$

where $x[n]$ denotes the input signal, $y[n]$ the output signal, $c[n]$ the control voltage.

The control voltage is calculated from a copy of the input signal that passes through the side-chain, as seen in Fig. 1. The side-chain first includes a peak detector to provide an instantaneous estimate of signal level,

$$x_G[n] = 20\log_{10}|x[n]| \tag{2}$$

The gain computer implements a static compression curve with input $x_G[n]$ and output $y_G[n]$, and is given by Eq. (3), where the sample number $[n]$ has been omitted for readability

$$y_G = \begin{cases} x_G & 2(x_G - T) < -W \\ x_G + \frac{(1/R-1)(x_G-T+W/2)^2}{T+(x_G-T)/R}\ /(2W) & 2|(x_G - T)| \leq W \\ & 2(x_G - T) > W \end{cases} \tag{3}$$

where T, R and W are the *threshold*, *ratio* and *knee width* parameters.

Fig. 2 is the graphical solution of Eq. (3), i.e. it presents both hard and soft knee and a make-up gain giving displacement from the diagonal. The (static) amount of compression is thus

$$x_L[n] = x_G[n] - y_G[n] \tag{4}$$

Smoothing is performed by a gain smoothing (also known as ballistics) stage,

$$\begin{aligned} y_1[n] &= \max(x_L[n], \alpha_R y_1[n-1] + (1-\alpha_R)x_L[n]) \\ y_L[n] &= \alpha_A y_L[n-1] + (1-\alpha_A)y_1[n] \end{aligned} \tag{5}$$

where $\alpha_A = e^{-1/(\tau_A f_s)}$ and $\alpha_R = e^{-1/(\tau_R f_s)}$are filter coefficients derived from the compressor's attack and release times, τ_A and τ_R, and f_s is the sampling frequency. The control voltage is thus found by adding the make-up gain, M, and then converting this back from decibel to linear scale,

$$c[n] = 10^{(M-y_L[n])/20} \tag{6}$$

In [12], it was shown that this compressor design yields smooth and relatively artifact-free performance for a wide variety of signals when compared with alternative designs. However, due to the influence of the attack envelope on the release trajectory in Eq. (5), the measured release time is approximately $\tau_A + \tau_R$.

2 PARAMETER AUTOMATION

Since the compressor parameters are used in different stages of the compressor design their automation methods can be independent of each other, even though they might be based on the same signal statistics.

2.1. Auto Attack and Release Times

Very short attack and release times should be avoided because they introduce a number of unpleasant artifacts such as pumping, breathing, low frequency distortion, and other artifacts [1,13]. Very long attack and release times are also rarely beneficial. The longer the attack the less responsive the compressor is to the signal. Likewise, a long release time may cause perceived dropouts after short transient sounds or reshape the decay part of notes and modify the sound of instruments.

Since most signals are time varying, dynamically varying time constants are preferred, so that they can adapt to the nature of the transient, steady state, and decay components in the signal. To minimize artifacts, a suitable auto-attack and release mechanism would choose shorter time constants when the input signal is highly transient or percussive, and longer time constants if it is a more steady state signal. In this section, we propose two methods, a time domain approach and a time-frequency processing approach [14], that give greater control and flexibility over the selection of suitable time constants.

2.1.1 Crest Factor as a Short Term Signal Measure

The crest factor is defined as the ratio of peak signal level to root mean squared (RMS) signal level over a given duration. In order to measure the short-term crest factor of a signal, without introducing any latency, we can combine a peak detector and an RMS detector, as given in Eq. (7),

$$\begin{aligned} y^2_{Peak}[n] &= \max(x^2[n], \alpha y^2_{Peak}[n-1] + (1-\alpha)|x^2[n]|) \\ y^2_{RMS}[n] &= \alpha y^2_{RMS}[n-1] + (1-\alpha)x^2[n] \\ y_C[n] &= y_{Peak}[n]/y_{RMS}[n] \end{aligned} \tag{7}$$

where forgetting factor $\alpha = e^{-1/(\tau f_s)}$ is calculated from time constant τ and sampling frequency f_s. The RMS detector, previously presented in [11,15], is a 1-pole smoothing filter applied to the square of the input signal, also known as an exponential moving average filter. The peak detector above has instantaneous attack and a smooth release trajectory. If we choose the peak detector and RMS detector time constants to be identical, we guarantee that the release envelopes of both detectors are the same, and that the peak detector's output cannot be less than the detected RMS output. The crest factor is independent of overall signal scaling and is therefore compatible with the design goal of level-independence.

The time constant τ for the two detectors determines the integration time of the crest factor measurement and was set at 200 ms based on informal testing.

Though the crest factor of a steady state signal is fairly low, it increases once the signal contains transients. Transients show high amplitude but are of short duration (typically less than 10 ms) in relation to the 200 ms integration time. Thus their contribution to the RMS value is much less than their contribution to the peak value. Thus, the crest factor can be used to locate transient parts in the signal, like note onsets.

2.1.2 Calculating Auto Attack and Release Times

The maximum attack time was set to $\tau_{A\max} = 80$ms and the maximum release time to $\tau_{R\max} = 1$s. We can find RMS detectors in some compressors with time constants on this scale to prevent low frequency distortion and other artifacts [1].

In order to avoid dropouts and pumping, the effect of a high crest factor on the release time needs to be extreme. For this reason we divide each maximum time constant by the square of the crest factor. The crest factor for a pure sine wave is $\sqrt{2}$, so we then multiply by 2 to ensure that the the maximum time constant is reached for sinusoidal input signals. Finally to compensate for the influence of the attack time on the measured release time in our compressor design [12], as discussed in Section 1, we subtract the attack time from the release time. Thus, the time varying automation of attack and release times is given by Eq. (8),

$$\begin{aligned}\tau_A[n] &= 2\tau_{A\max}/y_C^2[n]\\ \tau_R[n] &= 2\tau_{R\max}/y_C^2[n] - \tau_A[n].\end{aligned} \tag{8}$$

2.1.3 Spectral Flux as a Short-Term Signal Measure

The short time Fourier transform (STFT) of an input signal x is defined as

$$X(n,k) = \sum_{m=-N/2}^{N/2-1} x(nh+m)\omega(m)e^{-j2\pi mk/N} \tag{9}$$

where $X(n, k)$ represents the k^{th} frequency bin of the n^{th} frame, $\omega(m)$ is an N-point Hamming window and h is the hop size between adjacent windows.

The spectral flux (SF) measures how quickly the power spectrum of a signal changes and offers detection based on amplitude or energy information of the signal. It is calculated from the change in magnitude of the STFT over two successive frames, and it is restricted to count only those frequency bins where the energy is increasing. The normalized spectral flux is then defined as:

$$SF(n) = \frac{\sum_{k=-N/2}^{N/2-1} H(|X(n,k)| - |X(n-1,k)|)}{\sum_{k=-N/2}^{N/2-1} |X(n,k)|}, \tag{10}$$

where $H(x) = (x+|x|)/2$ is the half-wave rectifier function.

Though spectral flux is typically used as an onset detection function [16,17], it can easily be used for transient detection purposes [18,19]. The more transient a signal is, the higher its spectral flux value will be and the shorter the time constants that are needed to achieve proper compression.

The spectral flux is more sensitive than the crest factor and this enables it to detect more subtle changes to the signal. For the spectral flux calculation we use a window $N = 1024$ points for the Fourier transform, with a hop size between adjacent windows $h = 512$, i.e., 50% overlap between windows. These settings were chosen because they produce narrow peaks for the spectral flux function at the same time instances as the crest factor. However, as shown in Fig. 3, the spectral flux is also able to catch the change in frequency of the sine wave.

In order to overcome the problem of associating a spectral flux value to a maximum time constant, we scale the normalized spectral flux to the range of values of the crest factor. The crest factor is usually highest for the very first sample, which can be easily shown if we take equation (7)

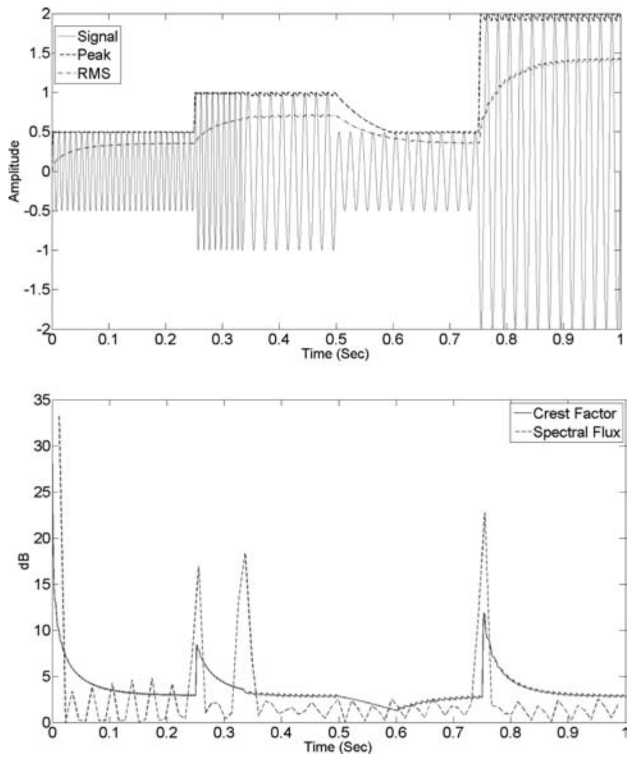

Fig. 3. (a) Sine wave with varied amplitude and frequency. (b) Crest Factor and Spectral Flux measurement on the sine wave.

and set $y_{Peak}[0] = 0$ and $y_{RMS}[0] = 0$;

$$\begin{aligned} y_C[1] &= \frac{1}{\sqrt{(1-\alpha)}}\\ &\geq \frac{|x[n]|}{\sqrt{\alpha y_{RMS}^2[n-1] + (1-\alpha)x^2[n]}}\\ &\geq y_C[n] \end{aligned} \tag{11}$$

This value is used as the high boundary for the spectral flux function and scales all other spectral flux values accordingly.

2.1.4 Calculating Auto-Attack and Release Times from Spectral Flux

The attack and release time constants play an important role close to the onsets of notes, since onsets will probably cross the threshold level and trigger the compression. Therefore, we correlate the time constants to the peaks of the spectral flux, which in turn are closely related to note onsets. Because spectral flux peak values are a lot higher compared to their corresponding crest factor values for a crest factor time constant of 200 ms, we do not have to use the square of these values in order to achieve short enough times after transients. Instead we use an instantaneous attack peak detector with a release time of 2 ms for calculating attack times and 9 ms for the release times to smooth the

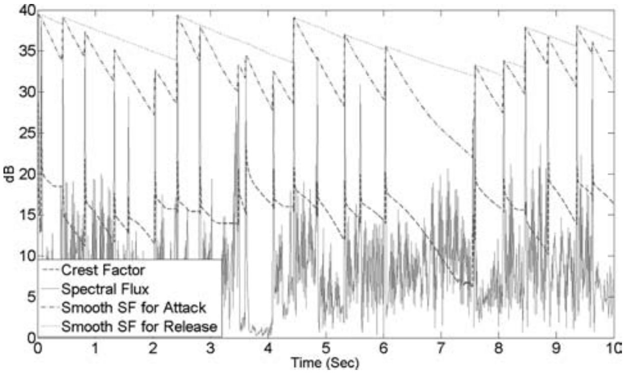

Fig. 4. Crest factor and spectral flux for calculation of attack and release times (Bass audio sample).

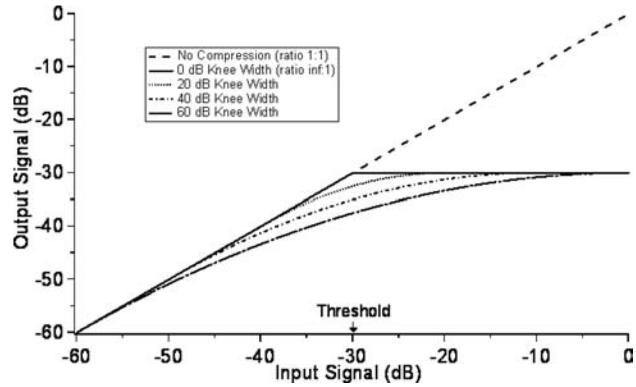

Fig. 5. Compression input/output curves with various knee widths for a set threshold at –30 dB.

spectral flux curve.

$$\begin{aligned}
\tau_A[n] &= 2\tau_{A\max}/SF_{smooth}[n] \\
\tau_R[n] &= 2\tau_{R\max}/SF^{\gamma}_{smooth}[n] - \tau_A[n] \\
SF_{smooth}[n] &= \max(x[n], \alpha SF_{smooth}[n-1] \\
&\quad + (1-\alpha)SF[n])
\end{aligned}$$

The maximum attack and release time constants are the same as those used for the crest factor method. The parameter γ was set to 0.8 to provide a less intense change of release times as opposed to attack times. Nevertheless, the difference between $\gamma = 0.8$ and $\gamma = 1$ is small and for the sake of simplicity it can be dropped in a simple implementation of the method.

Fig. 4 is an example of the crest factor and the smoothed spectral flux methods used to obtain the attack and release times.

2.2 Threshold and Ratio in the Auto Compressor

Both threshold and ratio parameters relate to the static compression characteristics. In an automated compressor we want the user to only have to adjust one parameter that will define the desired compression amount they want to apply to the audio signal.

As in [4], we let the threshold be manually chosen, set the ratio parameter to infinity, and use a soft knee with an automated knee width that will vary with time depending on the compression of the signal. This method is based on the idea that a very soft knee can also be seen as an automatic ratio.

2.3 Auto Knee

By using a soft knee in the gain computer stage and setting the ratio to ∞:1, the slope of the static compression curve of Eq. (3) becomes

$$\frac{dy_G}{dx} = \begin{cases} 1 & 2(x_G - T) < -W \\ 1 - (x_G - T + W/2)/W & 2|(x_G - T)| \leq W \\ 0 & 2(x_G - T) > W \end{cases} \tag{13}$$

Hence, the signal will be perfectly limited once it exceeds T+W/2 . Below that point the slope will gradually increase, reaching ½ (equivalent to a ratio of 2:1) exactly at T, and it will keep decreasing until T-W/2, where it will become 0 (no compression at all, equivalent to a ratio of 1). So by setting the ratio to infinity and varying the knee width one can access the whole range of compression ratios. Fig. 5 presents the static compression behavior for various knee widths and a set threshold.

2.3.1 Automating the Knee Width

If the compression applied is for short periods of time, so only a few peaks are trimmed and the average gain reduction is small, then one might want the compressor to act as a hard limiter. On the other hand, if the signal is heavily compressed and the average gain reduction is high, one might want a smoother and less obvious compression effect.

For the automatic knee mechanism we propose an adaptive method based on the average gain reduction of the compressor, given by the following equation;

$$\begin{aligned}
c_{Dev}[n] &= \alpha c_{Dev}[n-1] + (1-\alpha)(c[n] - c_{Est}) \\
W[n] &= 2.5(c_{Dev}[n] + c_{Est})
\end{aligned} \tag{14}$$

c_{Dev} provides a smooth estimate of how much the control deviates from an estimated value based on the parameter settings of the compressor. A control voltage estimate, c_{Est}, is used to bias the averaging filter, by subtracting the estimate before the filtering and adding it back in afterwards. A reasonable setting for the estimated value is given from the Threshold and Ratio settings, $c_{Est} = T(1\text{-}1/R)/2$. This initializes the average gain reduction at a value close to its intended values and allows the control voltage estimate to quickly adapt to changes in parameter settings during real-time operation.

The averaging time constant of the filter, found from $\tau = -1/(f_s \ln \alpha)$, needs to be carefully chosen. A time constant that is too short will follow the compressor's gain reduction too quickly and a long time constant will be too slow in following the gain reduction curve and reaching the intended values. $\tau = 2$s was chosen, so that use of average gain reduction for make-up gain in Section 2.4 does not interfere with the release envelope.

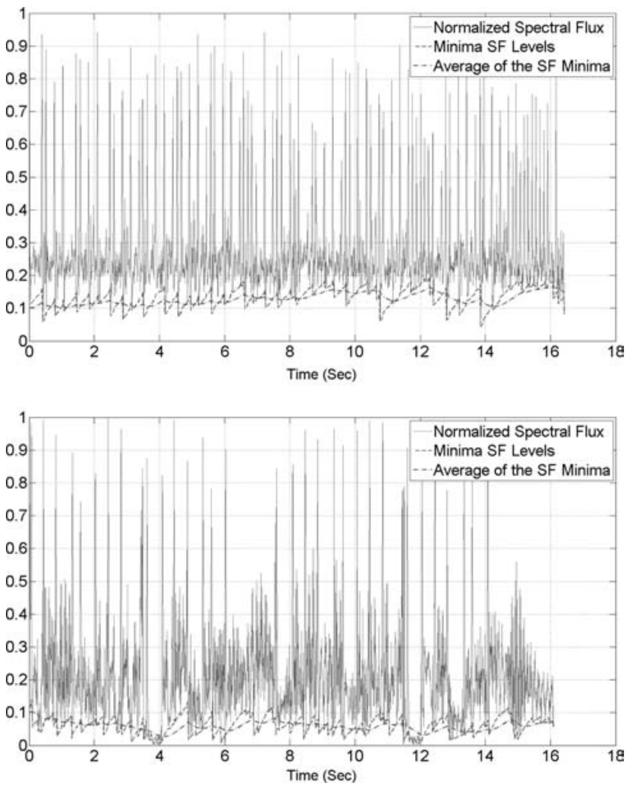

Fig. 6. Spectral flux of the Drums sample (top) and the Bass sample (bottom).

The scale factor 2.5 was derived empirically from informal listening tests. The result is a knee width that is slowly and smoothly varied with time.

The main weakness of the proposed model is that the knee width is exclusively related to the average gain reduction and not directly related to any characteristics and information of that signal.

2.3.2 Optimizing the Knee Width Automation with Information on the Input Signal

The following method suggests a way to optimize the knee width using information on the input signal extracted with the normalized spectral flux. Signals with extensive transient content will have their spectral flux values above a certain level, considerably higher compared to that of a signal with fewer transients, since in every frame step there will be significant transient content captured by the SF. For example, in Fig. 6, the spectral flux values of the drums sample are constantly above 0.1 and around 0.2 while the spectral flux values of the bass sample reach a minimum level of 0.05.

We first calculate the minimum levels of the spectral flux using a modified version of an instantaneous attack decoupled peak detector and then we use a low-pass filter to find the moving average of these values. The method can be summarized as:

$$\begin{aligned} SF_{\min}[n] &= \min(|SF[n]|, \alpha SF_{\min}[n-1] + (1-\alpha)SF[n]) \\ SF_{\min,avg}[n] &= (1-\alpha_2)SF_{\min}[n] + \alpha_2 SF_{\min,avg}[n-1] \end{aligned} \tag{15}$$

where the coefficients α, α_2 were based on time constants $\tau = 2$ ms and $\tau_2 = 1$ ms respectively, so as to guarantee the desired performance.

The evaluation results for the knee width automation (see 3.2) showed that the relationship between the average gain reduction and the preferred knee width is nonlinear and instrument independent. Therefore, a polynomial of order k was used to describe this relation.

$$W[n] = 2.5c^{k}_{Avg}[n] \tag{16}$$

A steady state signal should result in a roughly constant knee width to prevent unnecessary modulation of compressor parameters. On the other hand, signals that are very transient in nature should produce a compressor whose knee width varies more with gain reduction, so that it can both act like a limiter for high amplitude signals and provide a smooth transition to no compression on low amplitude signals.

As shown in Fig. 6, a highly transient signal, such as a percussive drums sample, will never present very low values for minima since there will always be transient activity captured by spectral flux. A steady-state signal will have spectral flux minima reaching lower values since initial transients of the attack part of the notes will quickly fade while the steady-state part will remain longer. Based on the average of the spectral flux minima we set k to values that, as seen in Section 3, produce the desired behavior for the knee width.

$$k = \begin{cases} 0.6 & SF_{\min,avg} > 0.1 \\ 0.05 & SF_{\min,avg} \leq 0.1 \end{cases} \tag{17}$$

2.4 Auto Make-up Gain

The aim of the make-up gain function is to achieve equal loudness between the compressor input and output signals (though it may also be used to maximize loudness of compressed recordings, contributing to the "loudness war" [20]). Our first approach is to estimate the make-up gain based on the average amount of applied compression. From Eq. (14),

$$c_{make-up}[n] = -(c_{Dev}[n] + c_{Est}) \tag{18}$$

Our second approach uses a loudness function to compare perceived loudness before and after compression. In [21], the EBU standard for loudness, based on a thresholded implementation of the ITU 1170 standard [22,23], was used to compare the loudness of tracks in multitrack audio, in order to automate time varying fader controls. Here, the EBU standard is used to measure loudness of the uncompressed and the compressed signal. This enables us to extract the loudness difference between the two signals and use it to calculate the make-up gain needed for the compressed signal. Even with the application of loudness-based make-up gain, the compressor is still able to significantly reduce the loudness range of the signal [24].

3 EVALUATION

Subjective evaluation was performed with two groups of subjects: nine expert mixing engineers (Professional

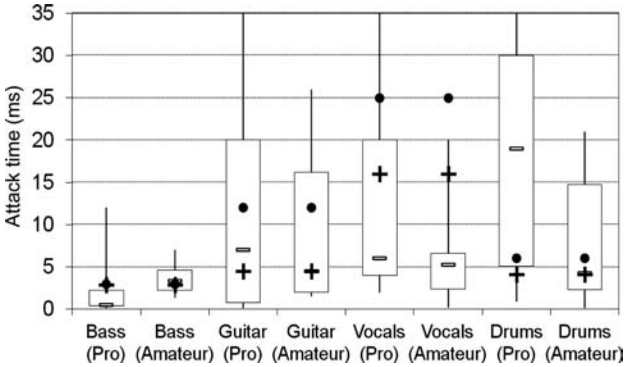

Fig. 7. Box plots for the attack time evaluation. Results in ms with median value (dash), crest factor automation (dot), and spectral flux automation (cross).

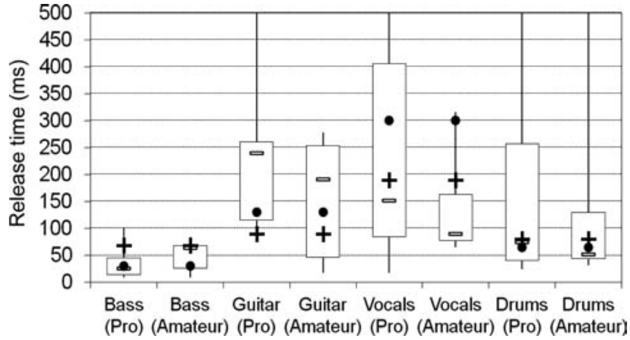

Fig. 8. Box plots for the release time evaluation. Results in ms with median value (dash), crest factor automation (dot), and spectral flux automation (cross).

group) and seven amateurs who had experience with dynamic range compression (Amateur group). A "method of adjustment" style test [25] was performed to obtain quantitative data on how humans set up and use a dynamic range compressor in their environment with their own equipment. Each test subject was provided with a VST plug-in of the compressor, test instructions, and four short audio tracks of drums, bass played in "slap" style, soft vocals, and acoustic guitar. The instructions included a series of listening tests in which the users had to tune individual parameters to their preferred setting while keeping all other parameters fixed at predefined settings. The predefined values were usually such that they would generate obvious amounts of compression and make any compression artifacts easily spotted by the listener.

The results were compared with what the automation method had chosen as preferred automated parameter settings. While the preferred human choices for each setting were single, static values, the compressor's automation method is an adaptive process, producing dynamically varied values. Therefore, some form of grouping or averaging of the dynamic values had to be performed.

3.1 Evaluation of the Auto-Attack and Release Times

For the evaluation of attack and release times, other parameters were predefined as follows: threshold at –30 dB, ratio at ∞:1 and knee width at 0 dB (hard knee). For auto-attack time we calculated the average out of all attack time values that fall in a time period equal to the maximum attack time after every possible onset (peak of the spectral flux) of the signal. For the auto-release time, since we cannot predict exactly when release times will be used we simply found the mean out of all the values.

Figs. 7 and 8 present box plots for the preferred attack and release time respectively. The box in each column indicates the interquartile range. The bottom of the box corresponds to the lower quartile (25th percentile) and the top of the box to the upper quartile (75th percentile). The dash within the box shows the median value of the data set, and the vertical black line shows the sample range from the minimum to the maximum sample value.

The small interquartile ranges for bass sample show that most of the testers agree that it requires a very fast attack time in order to prevent the initial transient of each note from slipping through. To avoid a drop-out after the very hard initial transient, most also set the release time to a very fast value (median of 26 ms for the professionals).

Test subjects preferred a longer attack and release time for the guitar compared to the bass. This resulted in the spectral flux providing good results for the auto-attack and the crest factor providing good results for the auto-release. For vocals, the automatic release is quite slow, especially for the crest factor approach. The median for both amateurs and professionals suggests a much faster release time constant (of 100 to 150 ms).

For the drums sample, which is similar to the bass in terms of richness in transients, the choices of the professionals are highly diverse. If we concentrate at the median time constant, it is much longer this time (19 ms), which indicates that it might be desirable to preserve the initial transient of each drum hit.

Due to the lack of transients in the soft vocals sample, the automatic compressor chooses a slow attack time in order to prevent it from being distorted. Although the interquartile range for the professionals is high, the median for both professionals and amateurs suggests an attack time of only approximately 6 ms.

Generally, the spectral flux automation method performs better than the crest factor method. This is mainly due to the fact that for the crest factor method we were depending on the long time constant to produce a smooth result while for spectral flux we used a subsequent smoothing filter to achieve this.

3.2 Evaluation of the Auto Knee

The parameters for knee width evaluation were ratio at ∞:1, attack time at 0.5 ms, and release time at 100 ms. Test subjects were asked to choose their preferred knee width for threshold values –18 dB, –25 dB, and –40 dB. For comparison of automation with preferred user knee width setting we calculated average gain reduction for each threshold value and from that found the corresponding average knee width that the automation methods used. We concentrated

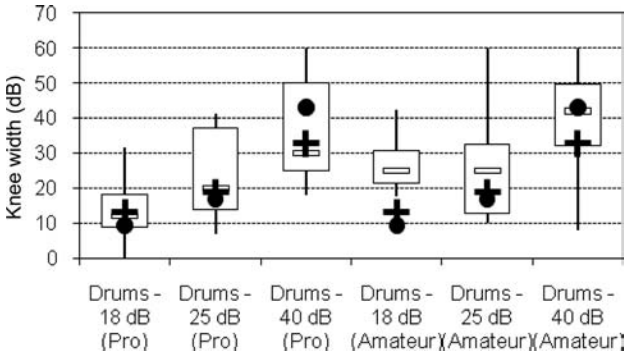

Fig. 9. Box plots for the knee width drums evaluation. Results in dB with median value (dash), Gain reduction dependent automation (dot), and automation with side-information from the spectral flux (cross).

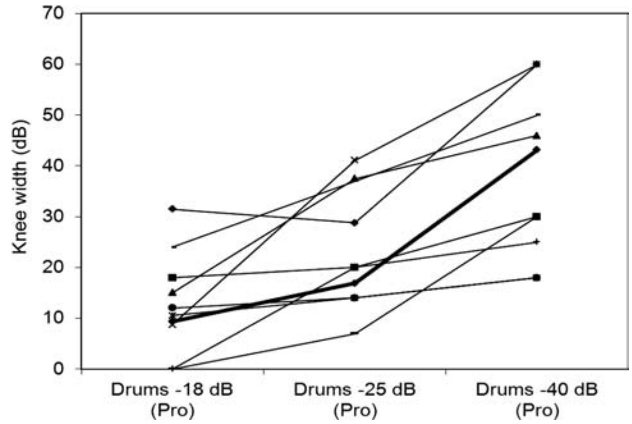

Fig. 11. Individual choices (in dB) for the drums sample knee-width experiment for professionals. Thick black line represents the median.

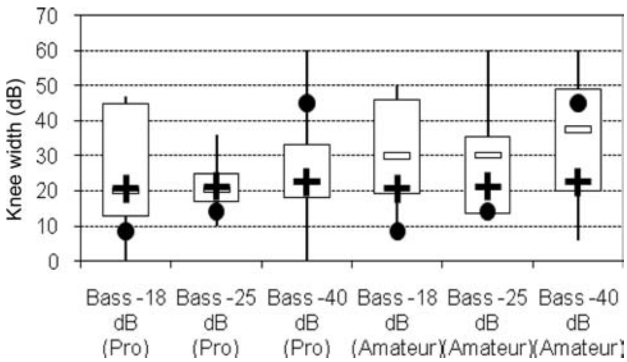

Fig. 10. Box plots for the Bass sample knee width evaluation. Results in dB with median value (dash), Gain reduction dependent automation (dot), and automation with side-information from the spectral flux (cross).

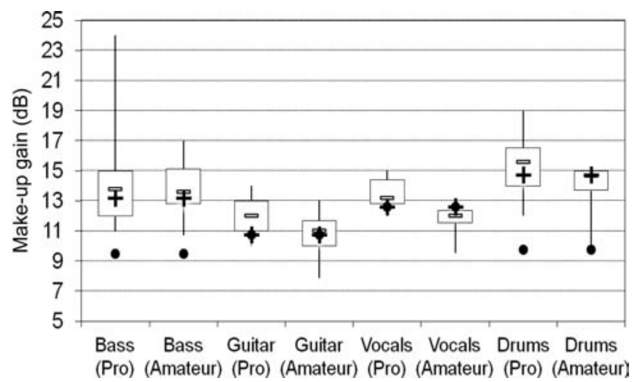

Fig. 12. Box plots for the Make-up gain evaluation. Results in dB with median value (dash), average control voltage automation (dot), and loudness-based automation (cross).

on the two more percussive signals, drums and slap bass, since the influence of the knee is is more pronounced on such content. Figs. 9 and 10 present the results from this test.

The results on the drums sample confirm that testers prefer a softer knee for heavier compression, i.e., lower threshold. The trend is more clearly seen in the professional results than in the amateur ones. It seems that using average gain reduction as a means of adjusting knee width was successful.

For the slap bass sample, the median for the professionals suggests a fairly constant knee width regardless of threshold, and amateurs prefer a softer knee at a threshold of –18 dB than at –25 dB. Therefore, using the same automation as that for the drums sample gives poor results. The spectral flux modification corrects this shortcoming and the results are more consistent with what users would use.

Fig. 11 shows the choices of the professionals for the drums sample as a function of threshold, with the choices of the gain-reduction dependent automation method indicated by a thick line. Almost all results show an upward trend (broader knee width for lower threshold), although the trajectories themselves differ regarding scaling and offset. This justifies our choice of using the information from the average gain reduction to adjust the width of the knee. Furthermore, including information from the spectral flux helped fine-tune the method to better fit the preferred user choices.

3.3. Evaluation of the Auto Make-up gain

The parameters for the make-up gain evaluation test were threshold at –30 dB, ratio at ∞:1, attack time at 0.5 ms and release time at 100 ms. These settings (low threshold and very short time constants) were chosen to guarantee that all four test signals would be heavily compressed. Test subjects were asked to manually vary the make-up gain, until the compressed signal has the same loudness as the uncompressed signal. Results are presented in Fig. 12.

The full range of results for this experiment varied significantly (e.g., one professional tester applied 24 dB of make-up gain to the bass sample, which is more than 10 dB above the median value). A few testers reported that they found it difficult to judge whether the two signals appear equally loud when their dynamic range is so different and came to different results, whether they concentrated on the attack (transients) or the sustain (steady-state) part of the sound. However, the interquartile range is quite small, within only 3 dB for all of the make-up gain experiments. This means that most testers agreed on a make-up gain for a given sample.

Comparing the results to the automation, the average control-voltage dependent make-up gain is quite accurate for the guitar and the vocal samples. In both cases most professionals would apply slightly more and most amateurs would apply slightly less make-up gain. However, it failed to provide the desired gain for the slap bass and drums audio samples. Both signals were characterized by short-lived high peaks with high transient content located mainly in their loud onsets. These loud transients contained in the original signals are quite significant for perception of the signal's overall loudness, so when those transients are suppressed by the compressor, we require more make-up gain than the actual average gain reduction in order to achieve equivalent loudness.

The loudness-based make-up gain comes a lot closer to the median value of the users' experiment and can be characterized as accurate for all cases apart from the drums where the make-up gain is about 3 dB more than what the testers believe it should be. That can be explained due to the transient nature of drums, which makes loudness measurement difficult.

4 CONCLUSION

In this paper we have proposed a series of methods to automate most of the parameters of a digital dynamic range compressor. These methods are independent of one another for each parameter and can be used together or separately in different compressor models. We studied the performance of these methods and compared them against the choices of human operators.

To the best of the authors' knowledge, this presents the first subjective evaluation of preference for dynamic range compressor parameters when applied to musical signals. However, the purpose of the evaluation was to understand how the proposed automation methods perform in comparison to user preference. Because test subjects used their preferred listening environment, no knowledge was obtained about how human preferences might change with controlled conditions. Furthermore, the number of test subjects and test signals was limited. The test was performed on individual tracks that were not part of a general mix. As a result what was tested was not compression of tracks in order to nicely fit into a mix but rather individual track compression. Finally, the preferred human choices for each setting had to be compared against an adaptive automation method. It is clear that more evaluation is needed for further work.

Both methods proposed for time constants, crest factor, and spectral flux produced good results. Using a knee width dependent only on the amount of compression did not provide the best results. However, the modification we introduced to include information from the spectral flux in the auto knee width calculation, led to a successful improvement with very good results. We managed to follow the choices of the human operators very closely. Using the average gain-reduction as make-up gain worked well and was described as helpful by some testers. It saves the user from adjusting the gain whenever they made a significant change to any of the other parameters (especially the threshold). The introduction of a loudness measure for the make-up gain was significant in improving the performance of the method.

The thresholds used in Eq. (16) and (17) were derived empirically and attempted to optimize knee width automation based on the results from the human users. A more extensive evaluation would help in testing a proper hypothesis for preferred knee width, but this goes beyond the scope of the paper.

A different approach to the automation of the static compression characteristic would be to automate the threshold to follow the RMS of the signal and let the user adjust the ratio based on what they prefer. This would avoid keeping the ratio fixed at infinity, which confines the compressor's operation to be close to a limiter.

For the attack and release times we proposed the use of Spectral flux as a method for transient/onset detection. Other methods exist [16,17] that could provide similar performance.

When using the crest factor to obtain the time constants one can achieve smoother operation by increasing the time constant used for the peak and the RMS detector in the crest factor calculation. A similar approach could also be followed for the spectral flux. The use of a detector with long attack and release times could smooth the spectral flux curves. This is an alternative to the approach we used to calculate the attack and release times with the spectral flux.

Our evaluation of the automatic make-up gain suggests that the use of EBUR-128 and ITU-R BS.1770-2 [22,23] is very effective and shows general agreement with user preference. However, there is a slight underestimation of required make-up gain with the more percussive samples, as noted in previous work [24]. This loudness standard was intended for broadcast content, not isolated sources. Further research may suggest a modification of the loudness measure, which could more effectively take into account percussive signals, with high amplitude, broadband transients.

In general, creating an automatic compressor should become an easier task if we knew to what type of signal it will be applied. An autocompressor that only has to work on drums for instance can make many more assumptions about its input signal than a compressor that is expected to sound well on a arbitrary tracks. The autorelease mechanism, for instance, could potentially benefit from some form of tempo-dependence, at least for very rhythmic signals.

Finally, an interesting idea suggested by some of the professional testers would be to let the user control how the automation behaves by being able to set the meta parameters controlling the release time. The compressor behavior would still adapt to the signal but allow the user to maintain control over compression characteristics.

5 REFERENCES

[1] R. Izhaki, *Mixing Audio: Concepts, Practices and Tools* (Focal Press, 2008).

[2] U. Zolzer, *Digital Audio Signal Processing*, 2nd ed. (John Wiley and Sons, Ltd., 2008).

[3] Z. Ma, J. D. Reiss and D. Black, "Implementation of an Intelligent Equalization Tool Using Yule-Walker for Music Mixing and Mastering," presented at the *134th Convention* of the Audio Engineering Society (2013 May), convention paper 8892.

[4] A. T. Sabin, R. Zafar and B. Pardo, "Weighted-Function-Based Rapid Mapping of Descriptors to Audio Processing Parameters," *J. Audio Eng. Soc.*, vol. 59, pp. 419–430 (2011 June).

[5] J. D. Reiss, "Intelligent Systems for Mixing Multichannel Audio," *presented at the 17th International Conference on Digital Signal Processing (DSP2011)*, Corfu, Greece (2011).

[6] L. Tyler, "An Above Threshold Compressor with One Control," *presented at the 63rd Convention* of the Audio Engineering Society (1979 May), convention paper 1505.

[7] R. J. Cassidy, "Level Detection Tunings and Techniques for the Dynamic Range Compression of Audio Signals," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6235.

[8] A. Schneider and J. V. Hanson, "An Adaptive Dynamic Range Controller for Digital Audio," *IEEE Pacific Rim Conf. Comm., Comp. and Sig. Proce.*, p 339–42, Victoria, Canada (1991).

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

[10] S. Gorlow and J. D. Reiss, "Model-Based Inversion of Dynamic Range Compression," *IEEE Trans. Audio Speech Lang. Proc.*, vol. 21 (2013).

[11] D. Barchiesi and J. D. Reiss, "Reverse Engineering the Mix," *J. Audio Eng. Soc.*, v. 58, pp. 563–576 (2010 Jul./Aug.).

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

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

[14] F. Rumsey, "Time-Frequency Processing," *J. Audio Eng. Soc.*, vol. 60, pp. 464–469 (2012 June).

[15] F. Floru, "Attack and Release Time Constants in RMS-Based Feedback Compressors," *J. Audio Eng. Soc.*, vol. 47, pp. 788–804 (1999 Oct.).

[16] R. Zhou and J. D. Reiss, "Music Onset Detection," in *Machine Audition: Principles, Algorithms and Systems*, W. Wang, Ed., ed. (IGI Global, 2010).

[17] J. P. Bello,L. Daudet,S. Abdallah, et al., "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 1035–1047 (September 2005).

[18] I. Samaali, G. Mahé and M. Turki, "Watermark-Aided Pre-Echo Reduction in Low Bit-Rate Audio Coding," *J. Audio Eng. Soc.*, vol. 60, pp. 431–443 (2012 June).

[19] M. Zaunschirm, J. D. Reiss and A. Klapuri, "A Sub-Band Approach to Musical Transient Modification," *Computer Music J.*, vol. 36, no. 2, pp. 23–36 (Summer 2012).

[20] E. Vickers, "The Loudness War: Do Louder, Hypercompressed Recordings Sell Better?" *J Audio Eng. Soc.*, vol. 59, pp. 346–51 (2011 May).

[21] 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.

[22] ITU, "Rec. ITU-R BS.1770-2, Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," *International Telecommunication Union* (2011).

[23] EBU, "Tech Doc. 3341, Loudness Metering: 'EBU Mode' Metering to Supplement Loudness Normalisation in Accordance with EBU R-128," *European Broadcasting Union* (2010).

[24] J. Maddams, S. Finn and J. D. Reiss, "An Autonomous Method for Multi-Track Dynamic Range Compression," *presented at the Digital Audio Effects (DAFx)*, York, UK (2012).

[25] N. Zacharov and S. Bech, *The Perceptual Audio Evaluation: Theory, Method and Application* (Wiley-Blackwell, 2006).

THE AUTHORS

Dimitrios Giannoulis

Michael Massberg

Joshua D. Reiss

Dimitrios Giannoulis received the B.Sc. degree in physics from the National University of Athens, Greece, where he specialized, among others, on signal processing and acoustics. He received the M.Sc. degree in digital music processing in 2010 from Queen Mary University of London. He is currently pursuing a PhD degree in the Centre for Digital Music (C4DM) in Queen Mary University of London. His research interests are machine learning and signal processing.

●

Michael Massberg studied media technology at FH Oldenburg / Ostfriesland / Wilhelmshaven in Emden, Germany, and digital music processing at Queen Mary University of London, UK, where he received a Master's degree in 2009. He has been with German pro-audio company Brainworx since 2008, working as as an R&D engineer and specializing in digital modeling of analog recording studio hardware.

●

Dr. Josh Reiss is a Senior Lecturer with the Centre for Digital Music at Queen Mary University of London. He received his PhD 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.