

A STATISTICAL APPROACH TO AUTOMATED OFFLINE DYNAMIC PROCESSING IN THE AUDIO MASTERING PROCESS

Marcel Hilsamer,

Department of Digital Signal Processing
University of Kaiserslautern
Kaiserslautern, Germany
hilsamer@eit.uni-kl.de

Stephan Herzog,

Department of Digital Signal Processing
University of Kaiserslautern
Kaiserslautern, Germany
herzog@eit.uni-kl.de

ABSTRACT

Mastering audio is a complicated yet important step in music production. It is used for many purposes, an important one is to ensure a typical loudness for a piece of music within its genre. In order to automate this step we use a statistical model of the dynamic section. To allow a statistical approach we need to introduce some modifications to the compressor's side-chain or more precisely to its ballistics. We then develop an offline framework to determine compressor parameters for the music at hand such that the signal's statistic properties meet certain target properties, namely statistical central moments, which for example can be chosen genre specific. Finally the overall system is tested with songs which are available to us as unmastered, professionally mastered, and only compressed versions.

1. INTRODUCTION

Mastering audio is a complex task which requires an experienced sound designer. There is a huge amount of literature giving experience-based tips to sound designers on how to master mastering music. One of the best-known and most cited examples would be Bob Katz' "Mastering audio" [1].

An alternative approach to set the compressor parameters is the use of presets coming with today's software mastering tools like compressors and limiters. The drawback of these presets is, that they do not take into account the properties of the piece of music at hand, therefore a satisfying result can not be guaranteed.

Recently Giannoulis et. al. [2] proposed an automation of the compressors parameters based on the input signal's temporal behavior with only a single user controlled parameter, namely the compressor's threshold. Vickers [3] presented a method to automate a compressor based on the input signal statistics. His approach still has the need of the user controlled parameter threshold, as well as two target parameters, one to define the input-output relation and one to set a make-up gain, and of course the ballistics of the compressor.

We will present a new way to determine the control settings of the compressor depending on the material at hand and statistical determined target parameters without the need of any adjustments by the user. This will be done by means of the input signal statistics and an arbitrary reference, which could be the typical statistics of its genre. The statistical properties of music have been investigated in various ways. See for example [4], [5] and [6] which can be used as a starting point for more information on genre specific statistical properties, [7] gives an overview of statistical features which can be used to describe a compressor's behaviour.

In Section 2 we describe a classical compressor and its parameters. Section 3 introduces the statistical model of the dynamic section and defines some modification concerning the ballistics and signal feature detector of the compressor. This is used in Section 4 to present an offline procedure to estimate the compressor parameters needed to match certain target central moments based on the properties of the music at hand. This procedure will be evaluated in Section 5. Finally, Section 6 will summarise the proposed framework and its result and suggest fields of further research.

2. COMPRESSOR CONTROLS

There are countless compressor topologies, but all of them are using a signal path and a control path, the so called side-chain. This side-chain is typically realised in a feedforward structure. The principle block diagram is shown in Fig. 1, with the side-chain usually working on levels and consisting of four basic blocks

- a signal feature detector, like the Root Mean Square (RMS) or Peak detector to control the compressor based on the signal feature to be altered,
- the gain calculation based on an arbitrary input-output relationship, given as a characteristic curve,
- the ballistics, a gain smoothing stage with different time constants for rising and falling edges to reduce non-linear distortion on transient signals
- and finally the so called make-up gain, an increase of the calculated gain by a constant.

We will first describe these blocks in detail before we will discuss their contribution to the statistical model of the compressor. In the following we assume normalised pieces of music as input signals.

2.1. Signal feature detector

Classic analog compressors (or digital realisations of it) use either Peak or short time RMS detectors to control the compressor's characteristic [8]. This is done because of the simple realisation, not the musical or psychoacoustical meaning of these measures. With digital signal processing we can use arbitrary measures, like for example a complex loudness model with signal level dependent spectral weighting or a simplified model like in [3]. In order to keep the presented method a general framework we will use the term C_x for the input feature used in the side-chain. Analysing the compressors output y using the same detector yields the corresponding output feature C_y .

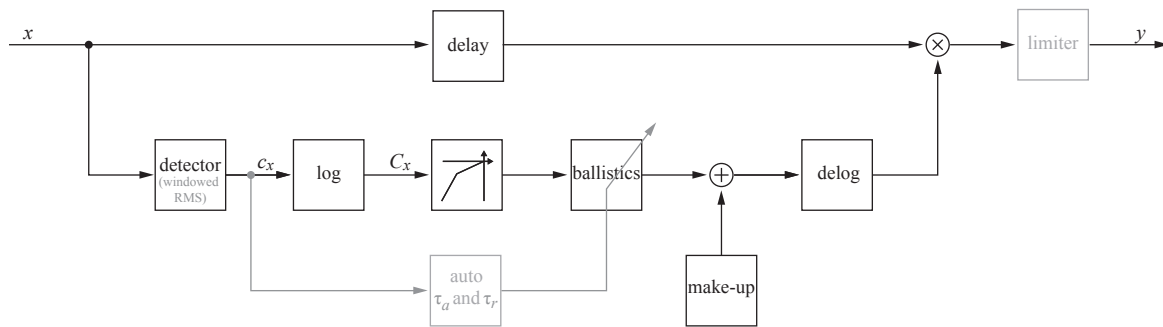


Figure 1: principle block diagram of a compressor with suggested extensions (grey).

2.2. Characteristic curve

The classical compressor characteristic is given by two parameters, namely the threshold (T) and the ratio (R). For levels below T the gain is 1, above T the slope of the characteristic is determined by the reciprocal of R . Fig. 2 illustrates the classical characteristic with $T = -30$ dB and $R = 3 : 1$ (solid) and the same characteristic with a make-up gain of $M = 20$ dB (dashed).

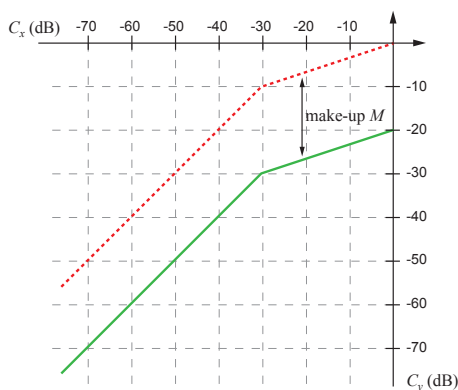


Figure 2: classical characteristic curve of a compressor with (red, dashed) and without (green, solid) make-up gain.

2.3. Attack and release times, the ballistics

The gain calculated based on the signal feature and the characteristic curve is usually smoothed to reduce non-linear distortion on transient signals and to suppress gain ripples produced by low frequencies. This is realised with different time constants for rising and falling gains, the attack time τ_a and the release time τ_r respectively. The user needs to select the attack and release times carefully which complicates the use of a compressor.

Katz [1] gives the following typical ranges: attack times ranging between 50 ms and 300 ms, common value 100 ms, release times lying between 50 ms and 500 ms, common value range between 150 ms and 250 ms. These values should be set suitable for the music's temporal behaviour, or in Katz' words: "it's counter-productive to go against the natural dynamics of music".

Due to the ballistics, the gain reduction or the actual compression depends not only on the input signal's characteristic level C_x but also on its temporal characteristics.

2.4. Make-up gain

The make-up gain is an added constant (sometimes signal dependent and time-varying as e.g. in [9]) to raise the system's output gain. We will use this block to normalise the chosen signal characteristic. This means we do not want to reduce the maximum of C_x (say the RMS value), instead we want to amplify the lower (RMS) parts of the signal. In other words the make-up gain should bring the maximum output value to the maximum input value, for example 0 dBFS. This leads to

$$M = -T \left(1 - \frac{1}{R} \right). \quad (1)$$

In Fig. 2 the dashed line shows the characteristic curve with a make-up gain according to equation (1) applied.

3. STATISTICAL MODEL OF A COMPRESSOR

The description of a compressor by means of signal statistics has recently been proposed by Shuttleworth [7]. He uses different signal features, i.e. inflection points as a peak measure, short and long term RMS as well as their partition into several frequency bands, to investigate the effects of a compressor by its Probability Density Function (PDF).

The statistic properties of any signal s is completely characterised by its PDF f_s . It is obvious that the compressor generates an output y with a PDF depending on the input's PDF and the compressor parameters. This holds of course for the PDF of any signal feature C_x (e.g. the RMS levels) used in the side-chain, namely f_{C_y} and f_{C_x} respectively, with

$$f_{C_y} = g[f_{C_x}, T, R, M]. \quad (2)$$

The influence of each block of the side-chain on f_{C_y} is now discussed to find an analytic expression for the function g .

The detector determines the signal feature to be altered. The static characteristic curve of the compressor is mainly responsible for the transformation of the PDF. This usually leads to a non-linear relation between input and output PDF that can be expressed analytically. The make-up gain just shifts the PDF by M and is

therefore a simple linear transformation. The ballistics however lead to a transformation which depends not only on the level of the input's signal feature C_x , but also on its temporal characteristics and thus can not be modelled appropriately in a statistical manner directly. Fig. 3 illustrates this effect in the time domain.

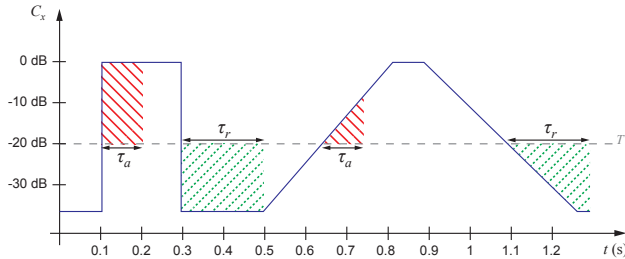


Figure 3: Illustration of the ballistics in the time-domain, red (solid lines) areas are above T but not fully compressed due to $\tau_a = 100ms$, green (dashed lines) areas are below T but partly compressed due to $\tau_r = 200ms$.

The red/green areas show the regions above/below threshold in which the gain does not follow the static characteristic due to the attack/release time. Their sizes are a direct measure for the variation from the static characteristic. By comparing the first red and green areas with the second ones it can be stated, that the faster the signal's feature cross the threshold the larger these areas are.

The ballistics behave like different low pass filters for descending and ascending gains with corner frequencies inverse proportional to the attack and release time and thus can not follow fast changes in C_x . The effect therefore can be lowered by simply adapting the attack and release time on the signal variation in time. We propose the use of automated ballistics based on recent research by Giannoulis et. al. [2]. In the following we will give a brief summary of their automated ballistics, extend their approach and show a computational efficient way to realise the proposed automation.

3.1. Auto ballistics

Giannoulis et. al. proposed an algorithm to ease the adjustment of the compressor parameters for the user. They automated the attack and release time based on the so called Spectral Flux (SF) of the input signal, which is a common tool for note onset detection. It sums the bins of the input short time spectrum in which the energy raises (positive half-wave rectifier) from one block to another, normalised by the total signal energy. This measure will be low for steady state signals and will show peaks for abrupt changes in pitch or transient level increase.

The basic concept of this approach is to use short response times for material consisting of strong transients and longer time constants for signals with constant levels over time. Giannoulis et. al. set the attack and release time inversely proportional to the SF. Listening tests with both, professionals and amateurs showed good results for this signal dependent automation of the ballistics [2].

This method seems a promising starting point to automate the ballistics in a musical sense as well as to lower the time dependency of the PDF transformation as described before. To do so, we will need to extend this approach to satisfy the following two conditions: First falling transients to automate the release time separately need to be included, whereas changes in pitch without sig-

nificant changes in signal level should be excluded in the measure. Second the measure should correlate strongly with the slope of level increase respectively decrease. Finally we will discuss the use of this measure with an arbitrary signal feature detector.

In [2] the SF alters the attack time as well as the release time which leads to good musical performance. Following this approach and taking into account that falling transients, i.e. an abrupt end of a sound, should alter the release time of a compressor, we propose to denote SF, as it is a measure for increasing energy or onset, by SF^+ and extend it by SF^- which sums falling energy bins (negative half-wave rectifier) or in other words is a measure for note offsets or falling transients.

By adding SF^+ and SF^- to SF^Σ the change of a note without a significant change in loudness will no longer be detected. A positive value will show an abrupt raise in signal level which can be used to scale the attack time of the compressor. A negative value indicates an abrupt fall in signal level and can therefore be used to scale the release time.

Following [2], for our simulations¹ we used a $N = 1024$ sample Fast Fourier Transformation (FFT) with a hann window and an overlap of $N/2$ to produce a value every 512th input sample (hop size $h = 512$) or every block k . Windowing the data prior to the FFT is important in order to reduce so called end-effects and therefore smooth the spectrum and thus the SF.

Fig. 4 (b)² illustrates SF^+ , SF^- and SF^Σ for a sine-wave with abrupt changing level and frequency. It is clearly visible, that SF^Σ is a good measure to detect rising and falling transients in signal level.

In order to reduce the computational effort we will realise a similar measure in the time domain. As the introduced SF^Σ basically indicates the differences in signal energy between two adjacent blocks we can also use the differentiation of the length N RMS value of the windowed³ input signal \hat{x} calculated with a moving averager, namely

$$\Delta_{RMS}[k] = \sqrt{\sum_{i=0}^{N-1} \hat{x}^2[kh+i]} - \sqrt{\sum_{i=0}^{N-1} \hat{x}^2[(k-1)h+i]} . \quad (3)$$

Fig. 4 (c) illustrates this measure. It can be seen that the proposed measure is similar to SF^Σ . It even correlates more with the amount of signal energy increase respectively decrease. This is not surprising since SF was introduced to detect note onsets and not their levels, hence due to the normalisation by the block energy every change in level will be detected almost equally.

Δ_{RMS} satisfies the required modifications, namely positive values are proportional to the slope of rising transients, negative values are proportional to the slope of falling transients and changes in frequency are ignored. Using this measure to scale the ballistics of the compressor guarantees a signal dependent attack and release time which will smooth the gain accordingly to the temporal behaviour of the input signal.

The use of any signal feature detector other than RMS leads to the need of a different measure as the characteristic and of course

¹With a sampling rate of $f_s = 48kHz$.

²The chosen test signal is equivalent to that in [2] for a better comparability of the results.

³Windowing in time domain calculations seems unusual, but the reduction of end-effects will help to suppress the RMS error due to the difference between integration time and the unknown signal's period. This leads to smoother, less oscillating results. A correction factor must be applied. For an overview of different windows and their correction factors see [10].

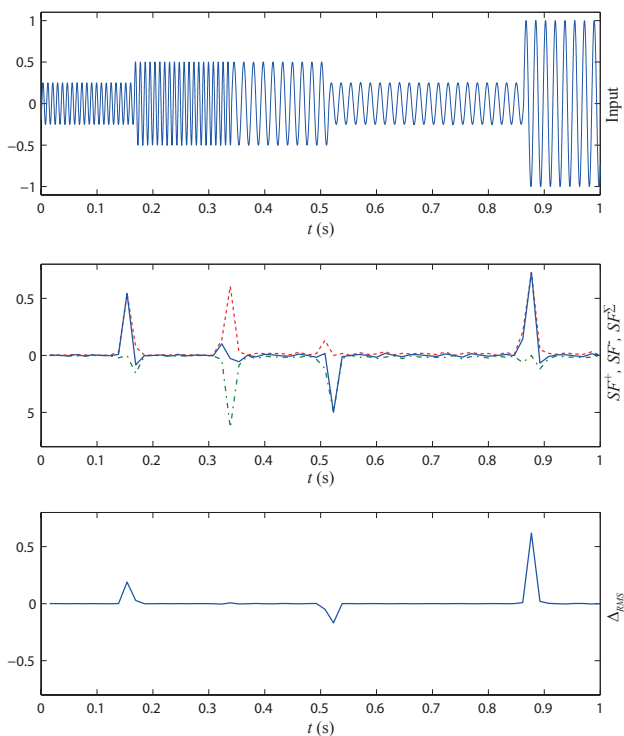


Figure 4: (a) input signal, sine-wave with abrupt changing amplitude and frequency. (b) SF^+ (red, dashed), SF^- (green, dash-dotted) and their sum (blue, solid) with an FFT length of $N = 1024$ using a hann window and 50% overlap. (c) Δ_{RMS} with block length $N = 1024$ using a hann window and 50% overlap.

the ballistics will not work on the RMS values anymore. Therefore we will introduce the general measure

$$\Delta_{C_x}[k] = \hat{C}_x[kh] - \hat{C}_x[(k-1)h] \quad (4)$$

with again windowed input data \hat{x} .

In order to achieve approximately equal areas shown in Fig. 3, the actual attack and release times should be calculated block-wise by

$$\begin{aligned} \tau_a[k] &= \tau_{a_{max}} (1 - 2 \max\{0, \Delta_{C_x}[k]\}) \\ \tau_r[k] &= \tau_{r_{max}} (1 + 2 \min\{0, \Delta_{C_x}[k]\}) . \end{aligned}$$

The maximum attack and release times should be set accordingly to typical values as given in Section 2.3, for example $\tau_{a_{max}} = 100\text{ms}$ and $\tau_{r_{max}} = 200\text{ms}$.

Using this automation, we can simplify the temporal characteristic-dependent, non-linear characteristic of the compressor to a temporal characteristic-independent system with kindly smoothed non-linearity and therefore a statistical analysis becomes acceptable. The smoothing is because of the now signal dependent ballistics, or in other words signal independent error areas in Fig. 3 and is ignored in the following parameter estimation process as it would not modify the needed parameters significantly.

4. AUTOMATED COMPRESSOR USING THE STATISTIC MODEL

We assume similar statistic properties for pieces of mastered music within a specific genre. This assumption is based on genre-specific instrumentation and arrangements as well as a similar overall sound and loudness as sound designers usually let themselves be inspired by currently popular productions within the same genre. This leads to genre-specific PDFs for the RMS levels [5].

As a compressor can alter the PDF of an input signal, if the transformation of the input PDF f_{C_x} to the output f_{C_y} is known analytically, one could invert this transformation and thus the proper parameters needed to meet a genre specific target PDF could be determined explicitly.

The input signal feature distribution is given by f_{C_x} and can be approximated by the histogram of the levels of the detector's output. This distribution will be altered by the compressor, more precisely its characteristic. We will now formulate analytic expressions for this alteration with the use of a PDF transformation.

4.1. Transformation of the PDF due to a compressor and its inversion

A compressor with a characteristic curve based on threshold and ratio alters the detected signal characteristic (without taking the ballistic into account) to the output distribution

$$f_{C_y} = \begin{cases} R f_{C_x}[C_x = RC_y], & C_x \geq T, C_y \geq \frac{T}{R} \\ f_{C_x}[C_x = C_y - M], & C_x < T, C_y < \frac{T}{R} \end{cases} \quad (5)$$

We now know how the compressor modifies the input PDF. With a genre typical target PDF, which could be the mean of a statistical relevant number of analysed mastered songs within a genre we can try to minimise the difference between this target PDF and f_{C_y} by adjusting R and T . An exact match in general is not possible as equation (5) does not arbitrarily modifies the input PDF. However it is possible to match certain (central or standardised) moments of the target PDF for example its mean and variance. Hence we will now analyse the transformation of these moments due to the compressor.

The i -th central moment $\mu_{i,y}$ is transformed to

$$\begin{aligned} \mu_{i,y} &= \int_{-\infty}^T (M + C_x - \mu_{1,y})^i f_{C_x} dC_x \\ &+ \int_T^{\infty} \left(\frac{C_x}{R} - \mu_{1,y}\right)^i f_{C_x} dC_x . \end{aligned} \quad (6)$$

With the help of equation (6) it is possible to determine the central moments of the output of the compressor based on its input and parameters R and T . As the parameter T is part of the integral limits it is not possible to directly invert equation (6) to determine R and T to match certain target moments $\mu_{i,t}$. In addition the input signal PDF f_{C_x} is not known as an analytic expression and needs to be approximated by a histogram.

We define the cumulative sums over the product of C_x^i and the histogram \hat{f}_{C_x} for all possible thresholds T as

$$m_i[T] = \sum_{l=C_{x_{min}}}^T l^i \hat{f}_{C_x}[l] . \quad (7)$$

This vector is an approximation for the i -th non central moment of the compressor input $\int_{-\infty}^T C_x^i f_{C_x} dC_x$ and can be used to evaluate the output central moments using simple vector additions. We will need two target moments to determine the two parameters R and T , in the following we will use the first two, namely the mean and variance.

The mean value is then approximated by

$$\mu_{1,y} = Mm_0[T] + m_1[T] + \frac{1}{R} (m_1[0] - m_1[T]) \quad (8)$$

and the variance to

$$\begin{aligned} \mu_{2,y} = & M^2 m_0[T] + 2Mm_1[T] + m_2[T] \\ & + \frac{m_2[0] - m_2[T]}{R^2} - \mu_{1,y}^2, \end{aligned} \quad (9)$$

with $m_i[0]$ representing the i -th moment with $T = 0$ dB, or in other words of the whole histogram of C_x .

By rearranging equation (8) to solve for R we obtain the vector

$$R_{\mu_1}[T] = \frac{Tm_0[T] - m_1[T] + \mu_{1,x}}{Tm_0[T] - m_1[T] + \mu_{1,t}} \quad (10)$$

containing the values for R which are needed to meet the target mean value.

By evaluating equation (9) at the points determined in (10) we obtain a vector containing all output variances $\mu_{2,y}[R_{\mu_1}[T]]$ for each parameter combination. Then the minimum of the squared difference of the variance $\mu_{2,y}$ of the output and its target $\mu_{2,t}$ determines the desired parameters T_{est} and R_{est} . In cases where an exact match in both moments is not possible with the use of just a compressor, this approach will lead to an exact match in mean $\mu_{1,y}$ and a minimum squared error in variance $\mu_{2,y}$. To ensure an exact match in variance and a minimum error in mean for all cases, equation (9) should be rearranged to solve for $R_{\mu_2}[T]$, then the minimum squared difference between $\mu_{1,y}$ and the target mean $\mu_{1,t}$ determines the desired parameters. The described procedure of course can easily be expanded to higher central moments or even standardised moments like the skewness or the kurtosis.

R_{μ_1} can contain negative values, as for higher thresholds, depending on the input signal and the target mean value, it will become impossible to match it with the make-up gain not letting C_y exceed 0 dBFS. Then a negative ratio R and therefore a negative slope of the characteristic above T will cause C_y to exceed 0 dBFS around $C_x = T$ and still meet 0 dBFS for $C_x = 0$ dBFS. This of course is not what we desire, so we will only use the range with positive values for R_{μ_1} to evaluate $\mu_{2,y}$.

4.2. Limiter to suppress overshoots due to attack time

Even though the make-up gain M is defined to meet 0dB at the output for a 0dB input, the non-zero attack time produces peaks of very short duration exceeding 0dB. In order to prevent the output signal from clipping these peaks have to be eliminated. This can be done with what is typically called a brick-wall limiter.

The use of such a limiter in the mastering process is a typical procedure as it can be used to make the piece of music at hand louder without changing its sound significantly. As distortion of short duration are nearly inaudible even gain reductions of several dB are possible for short peaks as long as the limiter's attack and release time are short enough [1]. We therefore propose the use

of a limiter with very short attack and release time and a so called look ahead, an infinite ratio and a threshold set near to 0 dB. Our simulations show good results using $\tau_{l,a} = 0.5$ ms, $\tau_{l,r} = 50$ ms and $T_l = -0.1$ dB in terms of no clipping and no significant deterioration in the matching of the target mean and variance⁴. The limiter is shown as the last block of the framework in Fig. 1.

5. SIMULATION RESULTS

To evaluate the proposed automated compressor we first tested the method to determine the parameters R and T described in Section 4.1. Then we used the proposed automatic compressor with songs which are available to us in unmastered and professional mastered versions⁵ in two scenarios: first to match the mean and variance of unmastered songs to those of the same songs processed with a classic compressor and second to match the mean and variance of these songs to those of the corresponding mastered version. Finally the results were validated in an informal listening test. For all simulations we solved the mean value for R to choose R and T to achieve the minimum squared error in variance.

By using a compressor with the proposed auto ballistics it should be possible to exactly reconstruct the chosen parameters R and T from the input and output PDF with the described parameter estimation method from section 4.1. In order to test this method we processed songs with a compressor employing the described windowed RMS detector and automatic ballistics using randomly chosen parameters R_{ref} and T_{ref} . The resulting versions were used to calculate the target moments for the parameter estimation process. The estimated parameters R_{est} and T_{est} were identical (with small variations due to the chosen histogram density) to the ones employed for generating the reference signal, as expected.

The complete automatic compressor was tested in the first scenario by using unmastered songs processed with a classic compressor, more precisely the one coming with Apple's Logic Pro, using typical settings according to [1] and with the build-in limiter activated, as the reference signal. Then the raw unmastered versions were processed with the framework and the moments of its output were compared to those of the reference signal. The results are shown in Table 1.

	song 1	song 2	song 3	song 4	song 5
$\mu_{1,t}$	-10.57	-14.47	-9.25	-18.47	-21.89
$\mu_{1,x}$	-12.55	-16.95	-12.36	-19.55	-26.76
$\mu_{1,y}$	-10.60	-14.50	-9.29	-18.47	-21.90
μ_{1,y_l}	-10.87	-15.01	-9.80	-19.15	-23.13
$\mu_{2,t}$	14.60	35.22	20.28	48.46	90.71
$\mu_{2,x}$	21.20	47.05	31.27	51.78	91.29
$\mu_{2,y}$	17.18	37.93	25.24	49.01	91.12
μ_{2,y_l}	17.12	37.90	24.96	49.01	91.10

Table 1: Mean and variance values of input signals, their targets calculated from the compressed signals, the output moments without ($\mu_{1,y}$, $\mu_{2,y}$) and with limiter (μ_{1,y_l} , μ_{2,y_l}).

It can clearly be seen, that the overall system shows a good matching of the mean values $\mu_{1,y}$ for each song. The variance of the output $\mu_{2,y}$ is kindly higher than its target, which is due to

⁴A small decrease in both, mean and variance, will occur (see Table 1).

⁵Song 1 - 3 from [11] and song 4 and 5 from the album 'Mind Meets Matter' by Claude Pauly.

the ballistics, leading to higher levels during attack and lower levels during release than modelled. This effect is of short duration and was therefore not audible in our tests. Furthermore this effect could be reduced by introducing an additive correction term to the target's variance, depending on $\tau_{a_{max}}$ and $\tau_{r_{max}}$. The discussed kind reduction of the moments due to the usage of the limiter is also clearly visible. This good overall matching is achieved although the parameters R and T of the automated compressor differ from those used to create the reference, which is not surprising as Logic's compressor and the automatic compressor employ different signal feature detectors.

In the second scenario we matched the unmastered versions to the mastered versions of the same songs. As mentioned before an exact match of the PDFs is not possible. This is mainly due to the possibly excessive use of a limiter (see Fig. 5 around -4 dB), and in this case, in addition, the possible use of an equalizer and/or a multi-band compressor during the professional mastering process, as these effects can not be modelled by just a compressor. However the statistical moments of the output meet the targets in all our simulations very well (comparable to the results in Table 1). As an example, Fig. 5 shows the PDF of one of the songs for its unmastered and mastered version as well as the output of the automatic compressor. The target moments were $\mu_{1,t} = -9.89$ dB, $\mu_{2,t} = 16.80$, the resulting compressor parameters $T = -16.5$ dB and $R = 1.281$ and the resulting moments at the output were $\mu_{1,y} = -9.92$ dB and $\mu_{2,y} = 19.20$ before and $\mu_{1,y_l} = -10.28$ dB and $\mu_{2,y_l} = 18.96$ after limiting.

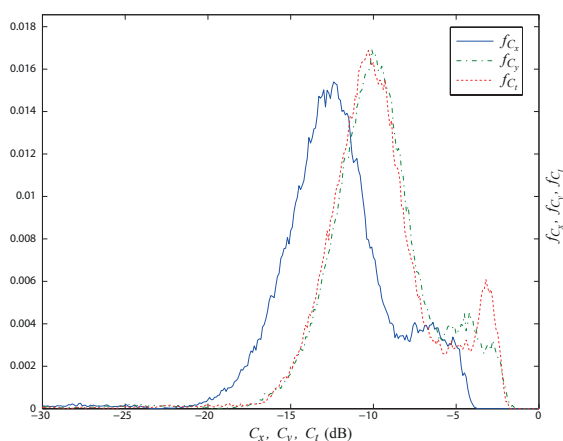


Figure 5: PDF of the RMS values of the input (blue, solid), the output (green, dash-dotted) and the target (red, dashed).

Finally, in order to evaluate the overall performance of the system we validated the results in an informal listening test, comparing the target wave file with the automatic compressor's output. In all examples there was no significant change in the overall loudness between both versions, which is not surprisingly due to the use of an RMS detector in combination with the relative exact match in its mean. In addition, no pumping or any other audible distortions were present, which indicates a good working automation of the ballistics. Finally the overall sound of the songs processed with the automatic compressor, besides some spectral differences compared to the mastered versions due to the use of an equalizer or multi-band compressor, was similar to that of their targets.

6. CONCLUSION

We presented a method to determine the parameters R and T for a compressor to ensure its output, based on the piece of music at hand, to meet certain target moments. The only modifications to a classic compressor needed for this statistical approach in mastering audio was automating the ballistics in the compressor and the use of a simple brick-wall limiter to eliminate overshoots due to the ballistics. In total this leads to a framework to automate the audio mastering process. With the use of target moments which differ significantly between different genres the proposed framework is able to match these moments and therefore a genre specific loudness and sound automatically.

Further research should focus on the identification of the most significant statistical moments to use as targets. In addition the framework can easily be extended to an automated multiband compressor by using several paths with different bandpass filters and an instance of the proposed compressor in parallel, which will help to take a genre typical spectral distribution during automated mastering into account. A useful starting point for getting proper multiband target moments or even PDFs can be [4] or [7].

7. REFERENCES

- [1] Bob Katz, *Mastering audio: the art and the science*, Taylor & Francis, 2007.
- [2] Dimitrios Giannoulis, Michael Massberg, and Joshua D. Reiss, "Parameter automation in a dynamic range compressor," *J. Audio Eng. Soc.*, vol. 61, no. 10, pp. 716–726, 2013.
- [3] Earl Vickers, "Automatic long-term loudness and dynamics matching," in *Audio Engineering Society Convention 111*, Nov 2001.
- [4] Goran Ferenc, "Dynamic properties of musical signals in genre ethnic electronica," in *INFOTEH JAHORINA Vol. 11*, 2012.
- [5] Mimir Mijic, Drasko Masovic, Dragana Sumarac Pavlovic, and Milan Petrovic, "Statistical properties of music signals," in *Audio Engineering Society Convention 126*, May 2009.
- [6] Patrick Agostini, "Investigation of statistical properties of music signals," Bachelor's thesis, Dept. of Digital Signal Processing, University of Kaiserslautern, Germany, 2014.
- [7] Tim Shuttleworth, "Evaluation of dynamics processors effects using signal statistics," in *Audio Engineering Society Convention 135*, Oct 2013.
- [8] Guy W. McNally, "Dynamic range control of digital audio signals," *J. Audio Eng. Soc.*, vol. 32, no. 5, pp. 316–327, 1984.
- [9] Jacob A. Maddams, Saoirse Finn, and Joshua D. Reiss, "An autonomous method for multi-track dynamic range compression," in *Proceedings of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, 2012, vol. 15.
- [10] F.J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan 1978.
- [11] Analog dimension - mastering studio, "Audio samples," <http://www.analogdimension.com/audiosamples.html>.