MACHINE LEARNING APPROACHES TO DYNAMIC RANGE COMPRESSION IN DIGITAL MUSIC PRODUCTION

Research Review

Elias Mistler (s1675946)
School of Informatics
University of Edinburgh

ABSTRACT

Sound engineering has been subject to many changes due to digitalization of the recording process and democratization of recording facilities. While almost anyone can produce decent raw recordings with a low budget nowadays, mixing and mastering is an expensive, work- and knowledge-intensive task needed in order to combine these raw recordings into an audio file suitable for publishing. In the last few years, automated algorithms for mixing and mastering tasks have emerged, paving the way for easy and fast high quality music production in the future.

This review looks into Machine Learning based approaches to automatic dynamic range compression, a rather complex and difficult task that demands considerable knowledge and experience, as it is not only non-linear and difficult to generalize or estimate, but also tends to produce unwanted artefacts or make a mix sound lifeless if used inappropriately. Promising research has combined common best practices in dynamic range compression with Machine Learning algorithms, outperforming amateur sound engineers and even getting close to the mixing decisions of a professional.

Table of contents

1. Introduction

Machine Learning plays a big role in optimization of business processes, engineering software and countless other research areas. But what about the creative world? How can algorithmic and analytic methods be used to support creative processes, such as the digital production of music? While research has been conducted in various areas of music informatics, most efforts seem to be streamlined into melody and music generation, music recognition and classification. Machine Learning approaches, mainly artificial neural networks, are widely used in tasks like genre classification and similarity discovery, improving services for music consumers.

If we switch views from consumer to producer, from listener to musician, we will notice that many musicians could strongly benefit from similar approaches as well – for example automation of digital audio mixing and mastering through artificial intelligence, namely Machine Learning algorithms. Recording equipment has already become very affordable and easy to use, allowing musicians to record their performances with a tiny budget. However, in order to transform these recordings into a high quality audio production suitable for public distribution, expensive expert knowledge and experience is still needed. While arguably, this involves a lot of creative work and musical understanding, many tasks have a more technical nature and can often be narrowed down to best practices.

The idea of having these best practices is not only very common in sound engineering education and literature; it can also be exploited in order to automate technical and repetitive tasks, to free up a sound engineer's time for more challenging and creative tasks, or even to allow for an automated processing of a multi-track recording. This automatic mix, by its nature, cannot be particularly creative or innovative, but it may offer good value for incredibly low cost – allowing for better previews of recorded material, or even for independent artists with low budgets to publish high quality recordings.

As in other research disciplines, Machine Learning promises a way of overcoming the need to collect best practices by rather learning the necessary adjustments and effect parameters from professional records than engineering the rules and constraints by hand. In this paper, we will – after a brief introduction of underlying methods in section 2 – examine the state of research in automation of digital music production focussing on Machine Learning approaches to dynamic range compression, an important step in both mixing and mastering. Section 4 summarizes important findings and points to possible future research.

2. Methods for parameter automation

The first considerable effort to automate processes in sound engineering was made in the 1970s when voltage-controlled amplifier (VCA) fader automation allowed the way a sound engineer changed individual channel gain values during a recorded performance to be recorded and replayed. With the emerging fully digital audio workstations (DAW) in the 21st century, this mechanism became extremely popular and easy to use, leading to the overall concept of Total Recall and non-destructive mixing (compare [1], chapter "automation"). While this kind of console or effect automation is usually just referred to as "automation" and can be used to change virtually any parameter in a DAW and its plug-ins over time, we will still refer to it with the historical "fader automation" in order to clearly separate this special mechanism from more general concepts of automation.

2.1 Cross-adaptive audio effects

A more sophisticated approach uses information from the input signal in order to control the effect's behaviour. Arguably, this concept has long been in place, e.g. in dynamic range compressors, which are essentially gain fader automations based on the input signal's magnitude. However, generalizing this model allows for more complex and intelligent effect implementations, labelled Adaptive Digital Audio Effects (A-DAFx). In order to address interdependencies between multiple channels, the control signal may include several input channels, in which case we will speak of cross-adaptive digital audio effects (XA-DAFx, sometimes CA-DAFx). [2]

2.2 DIGITAL AUDIO REPRESENTATION

Audio waves in digital audio are generally represented as a series of amplitude values (samples), typically with a sample rate of 44.1kHz. While technically, the values are stored in a normalized linear range between -1 and 1, algorithmic calculations usually take place in the rectified logarithmic domain (decibel full scale, dBFS), which corresponds better to human perception:

$$x_{dBFS} = 2 * \log_{10} x_{norm}$$

EQUATION 1: TRANSFORMATION TO LOGARITHMIC DOMAIN

While samples only give information about the magnitude and phase, many effects are designed to handle the tonal aspects of a signal. In order to extract frequency information, the short-time Fourier transform (STFT) fits a number of sinus waves to a set of samples, thus closely approximating their actual frequency content.

Equation 1 shows the short-time Fourier transform for a measurement in the frequency-time domain $X_t(k)$ for time t and frequency k, where $x(n)$ is the input at sample n and $w(n)$ is the hamming window used to smoothen overlapping time intervals.

$$X_t(k) = \sum_{n=0}^{N-1} x\left(t * \frac{N}{2} + n\right) * w(n) * e^{-\frac{j2\pi kn}{N}}$$

EQUATION 2: SHORT-TIME FOURIER TRANSFORM AS IN [3]

Digital audio effects can manipulate audio information in the frequency-time domain and then invert the Fourier transformation to retrieve a set of samples in the time domain, thus changing the frequency distribution in the resulting audio signal.

2.3 AUDIO FEATURE EXTRACTION

Machine Learning approaches are highly dependent on the representation of the data they are trained on. Useful audio features based on the Fourier transform include spectral centroid, roll-off and flux to quantify timbral texture. In order to extract rhythmic features, peaks in the audio signal are detected and combined in a tempo histogram which allows for identification of the base tempo and thus, rhythmic patterns with regards to the base tempo. [3] gives a more detailed overview of common audio features.

Low-level features specific to dynamic range, next to trivial features like current magnitude and magnitude range, also include the average loudness of a signal, defined as root mean square (RMS) of magnitude. Loudness in small time windows and loudness range over several windows are commonly used because they resemble the non-linear human perception of sound level better than the raw magnitude measures. Therefore, ratios like crest factor and dynamic spread, indicating relations between loudness and peak or current magnitude, respectively, can be useful indicators for dynamic range compression in Machine Learning approaches. [4]

2.4 AUTOENCODERS

Deep Neural Networks (DNN) have been argued to yield superior performance in many tasks such as image and sound classification if trained successfully. However, the traditional gradient-based training approaches are ineffective in DNNs due to the presence of multiple local minima. Therefore, a good fit cannot be guaranteed. Consequently, the results turn out unexpectedly poor. [5]

Alternatively, DNNs can be trained in a layer-wise, unsupervised fashion. As neural networks are inherently supervised models, the unsupervised nature has to be introduced by training with the same input and output. The layer-wise behaviour is realized by training the first layer with the input features as target and subsequently training every next layer with the weights its preceding layer, this way stepwise abstracting further. Semantically, this can be seen as the automatic construction of high-level, non-linear features – which however, are unlikely to correlate to human-perceived high-level features as described in [6].

This layer-wise, unsupervised training is referred to as autoencoder principle. It has been shown that autoencoders and Restricted Boltzmann Machines (RBM), which are based on a similar, layer-wise approach, are much more efficient in training and therefore able to outperform other classifiers like Support Vector Machines and Neural Networks trained with traditional, gradient-based methods. [7]

3. COMPRESSOR AUTOMATION IN A SOUND MIX

A good mix should always be balanced and transparent, meaning that it should be easy for a listener to hear and identify every instrument involved and where it is located in a (virtual) space relative to the listener. Thus, for the audio engineer, it is crucial to assign the "right place" to every instrument: The perception of width, depth and height[1] in a stereophonic space is mainly created by a combination of left-to-right panning and gain level, reverberation and equalizing of the single instrument tracks. [1]

3.1 RELATED RESEARCH

In the pursuit of automated mixing and mastering, a multitude of research papers have been published, mainly by researchers at the Centre for Digital Music at Queen Mary University of London. Based on the XA-DAFx architecture, automatic algorithms have been proposed, including inter-channel offset optimization to minimize unwanted comb filter effects [8–10] and automatic equalizing to optimize clarity through adjustment of relative volume [11], fitting to an idealized target curve retrieved from successful recordings [12] or by automatically identifying important frequencies of a channel and reducing masking [13] or loudness loss [14] through attenuation of the respective frequency in other input channels.

Similarly, there have been approaches to automate stereo panning by moving highly correlated signals to opposite directions and thus reducing masking [15–19]. A first Machine Learning approach in frequency equalizing relies on a multiple linear regression model fitted to a variety of target mixes and is able to generalize to new input data [20,21]. While for reverberation, little automation research has been published, there exist some adaptive reverb algorithms exist which automatically change their behaviour with regards to the input[2]. Moreover, promising approaches to dynamic range compression have been made which we will focus on in the rest of this paper.

While generally speaking, most of the research has been based on best practices and expert knowledge, we believe that there is a high potential for Machine Learning approaches for automation of sound engineering tasks which we can support by looking at the little research specific to dynamic range control automation that has already been published.

3.2 DYNAMIC RANGE COMPRESSION

There are various reasons to control – mainly compress or limit – the dynamic range of a signal. For one thing, some instruments have a high dynamic range which often cannot be reproduced for the listener due to technical limitations (e.g. radio transmission, compressed audio, loudspeakers) or a big difference between individual channel loudness (e.g. rather quiet notes in a piano melody becoming inaudible if an electric guitar is playing). Another thing is that over time, compression also became an important style element for certain music genres – for example, a modern rock recording will always use a high amount of compression. [1]

A dynamic range compressor (DRC) is an adaptive audio effect which operates non-linearly based on the input signal. However, it has several control parameters which make its use cumbersome and subject to individual experience of the producer, usually at least attack and release time, compression ratio and threshold. Many

[1] Owsinki [1]: "Tall, Deep, and Wide"
[2] E.g. http://www.zynaptiq.com/adaptiverb/ - Zynaptiq Adaptiverb

compressors also implement a knee function and offer make-up and pre-gains. Figure 1 shows the static interaction of compressor parameters: Samples exceeding the threshold are squashed by the compression ratio. Additionally, a knee of a given width is used to smoothen the transition. In figure 1, horizontal dashed lines indicate threshold and knee. [22]

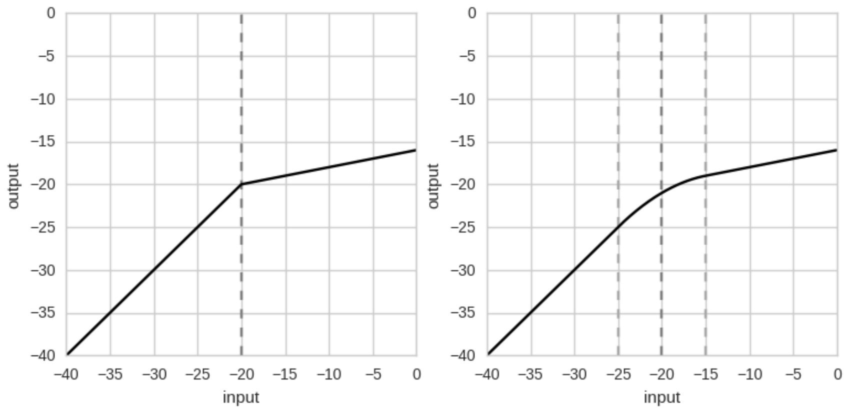

FIGURE 1: COMPRESSOR WITH AND WITHOUT KNEE (THRESHOLD = -20, RATIO = 5:1, KNEE WIDTH= 10)

The attack and release time constants are used to smoothen a copy of the input signal which is then used as a control signal to compress the input. An accurate and immediate compression would strongly distort the signal. Due to the resulting non-linear behaviour, as well as the non-linearity in the gain calculation (Figure 1), it is a difficult task to set up a compressor or estimate the applied settings from a compressed signal.

3.3 Compressor parameter automation with best practices

In [23], the complexity of a compressor is reduced to a single control by introducing a more fine-grained adapting mechanism with regards to the input, as well as interdependencies between the parameters: Attack and release times are dynamically calculated by detecting and adapting to transients in the input signal. The ratio is effectively controlled by the threshold setting and the input signal, as well as a feedback mechanism using the effective average gain reduction. The make-up gain is simply implemented by raising the output gain until it reaches the original loudness. The only manual setting left is the threshold. Some commercially available compressors already implement similar concepts[3] successfully, simplifying the use of a compressor noticeably.

Based on these algorithms for automatic parameter control, [24] presents a fully automated dynamic range controller. In a first step, all parameters are, similarly to the previous approach, reduced to one parameter, either threshold or gain ratio. The remaining parameter is then automated in a cross-adaptive manner, minimizing the difference of loudness ranges between input channels, therefore mainly compressing inputs with higher loudness ranges. However, the loudness range alone is an insufficient approximation of human perception and preference.

3.4 Intelligent multitrack compression with linear regression

Ma et al. [4] introduce Machine Learning to their existing approaches to automatic dynamic range compression. By training a multiple linear regression model, crucial features for compressor parameter automation should be identified. Mainly, common features such as dynamic spread, spectral centroids and spectral spread, as well as loudness and loudness range are used. On top of that, two new cross-adaptive features representing the relative share of transients and the relative share of low frequency energy in a track compared to the sum of tracks are introduced in order to allow the regression model to optimize for a list of assumptions thought to be crucial for the settings of a compressor:

[3] E.g. https://hofa-plugins.de/plugins/iq-limiter/ - HOFA IQ limiter

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

LISTING 1: ASSUMPTIONS ABOUT DYNAMICAL RANGE COMPRESSION IN [4]

Based on these features, a multiple linear regression model is trained on a set of manual mixing decisions previously obtained, thus learning the interdependency between audio features and the compression to be applied. The two features with the highest correlation to ratio and threshold, respectively, are selected for separate linear models which estimate ratio and threshold. In result, the ratio will be determined by percussivity and frequency weighting while thresholds are determined by RMS level and percussivity weighting. Other parameters are derived from ratio and threshold similarly to the aforementioned best-practices approaches.

While a small-scale subjective evaluation shows overall performance even better than two semi-professional mixes, there is no further information about the responsible sound engineers' skill level and the results are not compared to the work of an experienced professional. This leaves the overall result open for interpretation. Also, it is not made entirely clear why the linear model is restricted to two input features. We can only assume the calculations are kept to a minimum in order to allow for real-time deployment and intuitive human understanding. However, it is the belief of the author that the use of more features may lead to even better results. Ma et al. suggest the combination of equalization and compression into a single approach in further research – a rather novel approach that has already been made commercially available for manual use[4] and is also utilized in the proposal presented in the next section.

3.5 COMPRESSION FOR MASTERING WITH NEURAL NETWORKS

Mimilakis et al. [25] make use of a Deep Neural Network (DNN) to estimate compressor parameters. In order to do so, the inputs are first transformed to the frequency-time domain through use of the STFT. The linear frequency information is then mapped to a non-linear scale tailored to human perception and fed into the DNN.

The DNN utilizes seven layers and uses the rectified linear unit (ReLU, see Eq. 3) as an activation function for the first six layers. The output layer consists of a set of exponents to be applied to the inputs in the human-perception based time-frequency-domain. In order to retrieve gain factors from the resulting output, both input and output are transferred back to the linear frequency domain and then used to calculate the effective, frequency dependent compression ratios. After applying these compression ratios to the inputs through multiplication, the resulting frequency distribution is converted back to the time-domain using an inverted STFT to retrieve the compressed audio file.

$$ReLU(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

EQUATION 3: RECTIFIED LINEAR UNIT

The 7-layer neural network is constructed from two autoencoders, one 4-layer autoencoder to represent the input features and one 3-layer autoencoder trained on the target ratios which were retrieved from a mastering sample bank. The resulting model was proven to produce better results than a chosen reference system, but still worse than a professional audio engineer. [25]

Contrary to Ma et al., Mimilakis et al. focus on compression of a single track of audio recording which is fully analysed prior to applying any effect and then edited similar to reference compressed signals, making the approach suitable for mastering purposes, but much less so for mixing or even live applications. However, the

[4] https://hofa-plugins.de/plugins/iq-comp/ - HOFA IQ Compressor

full analysis allows for stable compressor settings which is likely to better preserve the musical dynamics of a song. Also, due to the frequency-dependent nature of the compression used, the approach effectively combines equalization and compression, an effect much looked for in mastering applications.

It came to the authors attention that the FFT utilized in the bottom layer of the DNN, due to its nature of summing up over time frames, behaves like a convolutional operation that is able to extract higher-level features (frequency information) from low-level features (magnitudes), but it is only used in the bottom layer. This suggests that the general use of Convolutional Neural Networks may be promising in future research approaches. High-level features with good correlation to human perception may be found in the process, thus enabling better estimation of the need of dynamic range compression and other audio effects.

Furthermore, the used mastering sample bank is a rather limited source (< 300 songs) of training data compared to the amount of training data usually used in Machine Learning. Also, it has to be acknowledged that the songs in the database do not have the status of famous top records. A bigger amount of training data, possibly top-quality work of renowned studios and artists may yield better results.

4. CONCLUSION AND FURTHER WORK

Based on previous compressor automation, two research proposals utilizing Machine Learning have been shown, one for multi-track compression improving transparency and individual audibility of single tracks in a mix by using a multiple linear regression model; the other one for automatic mastering of a mix using a 7-layer neural network. Both approaches show high potential, but also need further improvement.

For audio production, it turns out to be particularly difficult to get sufficient training data for effectively training Machine Learning models. Most sound engineers obviously never publish unmastered or even raw multi-track recordings. Small test banks[5] have been established, but seem fairly small for use in Machine Learning and do not fully match the quality of a top recording studio. However, the Machine Learning approaches themselves look very promising and seem to generalize well, even from the small training sets used.

4.1 COMMERCIAL STATUS

While automatic mixing algorithms such as multi-track dynamic range compression have in big parts been subject to a single group of researchers based at the Centre for Digital Music at Queen Mary University of London, automatic mastering approaches have already been commercialized.[6] However, details about the underlying algorithms are kept tightly under wraps. The presumably most sophisticated and popular automatic mastering service LANDR – which has emerged from the same research group – only stated that their toolset "uses A.I. and machine learning […], to replicate the processes human engineers make when mastering a track." [26], or, a little more specific "machine and deep learning", but not in which way it is used. [27] User experience of the tool suggests the use of genre recognition, based on which the actual mastering decision are taken, possibly on best practices or also on Machine Learning. Aside from the mastering service, there are hints of the company planning to create a novel "music mixing platform" [28] in the near future.

4.2 FUTURE WORK

Future work should investigate the possibilities, as well as the legal and ethical implications of working with a professional sound engineering studio that has access to thousands of multi-track records, as well as high-quality mixed and mastered results. Alternatively to transcribing the settings, reverse engineering approaches such as described in [19] and [29] could be used in combination with instrument and genre classification algorithms. The use of sound source separation techniques as suggested in [30] also seems promising in order to generate training data for Machine Learning algorithms or even separate sound sources and re-mix them as proposed in [31].

Future research may further focus on adapting other techniques from music classification and genre recognition. Also, methods from image classification and object recognition may be transferred to audio processing. Based on the behavioural similarity to the Fourier transformation, we especially suggest investigating the use of convolutional neural networks for automatic dynamic range compression in the future.

[5] E.g. http://www.cambridge-mt.com/ms-mtk.htm - 276 songs (as of 25th Dec 2016), http://multitrack.eecs.qmul.ac.uk/ - 593 songs (as of 25th Dec 2016), http://jazzomat.hfm-weimar.de/dbformat/dbcontent.html - 299 songs (as of 15th Jan 2017)

[6] landr.com, masteringbox.com, wavemod.com, emastered.com, curioza.com …

REFERENCES

[1] B. Owsinski, The Mixing engineer's handbook - third edition, Thomson Learning, Boston, 2014.

[2] V. Verfaille, U. Zolzer, D. Arfib, Adaptive digital audio effects (a-DAFx): A new class of sound transformations, IEEE Transactions on Audio, Speech and Language Processing 14 (5) (2006) 1817–1831.

[3] G. Tzanetakis, P. Cook, Musical Genre Classification of Audio Signals, IEEE Transactions on Audio, Speech and Language Processing 10 (5) (2002) 293–302.

[4] Z. Ma, B. de Man, P. Pestana, D. Black, J.D. Reiss, Intelligent Multitrack Dynamic Range Compression, Journal of the Audio Engineering Society 63 (6) (2015) 412–426.

[5] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, Journal of Machine Learning Research 11 (2010) 3371–3408.

[6] T. Wilmering, G. Fazekas, M.B. Sandler, High-Level Semantic Metadata for the Control of Multitrack Adaptive Digital Audio Effects, Convention of the Audio Engineering Society 133, San Francisco (2012).

[7] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, Y. Bengio, An empirical evaluation of deep architectures on problems with many factors of variation, International conference on Machine learning 24 (2007) 473–480.

[8] E. Perez-Gonzalez, J.D. Reiss, Determination and correction of individual channel time offsets for signals involved in an audio mixture, AES Convention 125, San Francisco (2008).

[9] A. Clifford, J.D. Reiss, Reducing comb filtering on different musical instruments using time delay estimation, Journal of the Art of Record Production (5) (2011).

[10] N. Jillings, A. Clifford, J.D. Reiss, Performance optimization of GCC-PHAT for delay and polarity correction under real world conditions, AES Convention 134, Rome (2013).

[11] E. Perez-Gonzalez, J.D. Reiss, Automatic equalization of multichannel audio using cross-adaptive methods, AES Convention 127, New York (2009).

[12] Z. Ma, J.D. Reiss, D. Black, Implementation of an Intelligent Equalization Tool Using Yule-Walker for Music Mixing and Mastering, AES Convention 134, Rome (2013).

[13] S. Hafezi, J.D. Reiss, Autonomous Multitrack Equalization Based on Masking Reduction, Journal of the Audio Engineering Society 63 (5) (2015) 312–323.

[14] G. Wichern, H. Robertson, A. Wishnick, Quantitative Analysis of Masking in Multitrack Mixes Using Loudness Loss, AES Convention 141, Los Angeles (2016).

[15] E. Perez-Gonzalez, J.D. Reiss, Automatic Gain and Fader Control for Live Mixing, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics New York (2009).

[16] S. Mansbridge, S. Finn, J.D. Reiss, Implementation and Evaluation of Autonomous Multi-Track Fader Control, AES Convention 132, Budapest (2012).

[17] S. Mansbridge, S. Finn, J.D. Reiss, An Autonomous System for Multitrack Stereo Pan Positioning, AES Convention 133, San Francisco (2012).

[18] E. Perez-Gonzalez, J.D. Reiss, A Real-Time Semiautonomous Audio Panning System for Music Mixing, EURASIP Journal on Advances in Signal Processing (Special issue on digital audio effects) (2010) article 436895.

[19] E. Perez-Gonzalez, J.D. Reiss, Automatic mixing: live downmixing stereo panner, International Conference on Digital Audio Effects 10 (DAFx-07), Bordeaux (2007).

[20] J.J. Scott, Y.E. Kim, Analysis of Acoustic Features for Automated Multi-Track Mixing, International Society for Music Information Retrieval (ISMIR) Conference 12 (2011).

[21] J.J. Scott, M. Prockup, E. Schmidt, Y.E. Kim, Automatic Multitrack Mixing Using Linear Dynamical Systems, Sound and Music Computing Conference 8, Padova (2011).

[22] D. Giannoulis, M. Massberg, J.D. Reiss, Digital Dynamic Range Compressor Design: A Tutorial and Analysis, Journal of the Audio Engineering Society 60 (6) (2012).

[23] D. Giannoulis, M. Massberg, J.D. Reiss, Parameter Automation in a Dynamic Range Compressor, Journal of the Audio Engineering Society 61 (10) (2013) 716–726.

[24] J.A. Maddams, S. Finn, J.D. Reiss, An autonomous method for multi-track dynamic range compression, International Conference on Digital Audio Effects 15 (DAFx-12), York (2012).
[25] S.I. Mimilakis, K. Drossos, T. Virtanen, G. Schuller, Deep Neural Networks for Dynamic Range Compression in Mastering Applications, AES Convention 140, Paris (2016).
[26] S. Mansbridge, LANDR.com: Professional Audio Mastering, available at https://www.landr.com/en/about (accessed on December 27, 2016).
[27] Angel.co, MixGenius: Intelligent Music Production SaaS, available at https://angel.co/mixgenius (accessed on December 27, 2016).
[28] K. Strauss, Making Money On The Music Mix: Meet MixGenius, available at http://www.forbes.com/sites/karstenstrauss/2014/06/16/making-money-on-the-music-mix-meet-mixgenius/ (accessed on December 27, 2016).
[29] D. Barchiesi, J.D. Reiss, Reverse Engineering of a Mix, Journal of the Audio Engineering Society 58 (7/8) (2010).
[30] A.J. Simpson, G. Roma, M.D. Plumbley, Deep Karaoke: Extracting Vocals from Musical Mixtures Using a Convolutional Deep Neural Network, MSc, Cornell University, New York (2015).
[31] S.I. Mimilakis, E. Cano, J. Abeßer, G. Schuller, New Sonorities for Jazz Recordings: Separation and Mixing using Deep Neural Networks, AES Workshop on Intelligent Music Production 2, London (2016).