Restoring the Dynamics of Clipped Audio Material by Inversion of Dynamic Range Compression

Stanislaw Gorlow, Joshua D. Reiss, Elodie Duru

Restoring the Dynamics of Clipped Audio Material by Inversion of Dynamic Range Compression

Stanislaw Gorlow
LaBRI, UMR 5800
University of Bordeaux
33400 Talence, France

Joshua D. Reiss
C4DM, Queen Mary
University of London
London E1 4NS, UK

Elodie Duru
Aquitaine Science Transfert
33600 Pessac, France
http://ast-innovations.com

***Abstract*—In this work, a novel approach for the restoration of clipped audio alias declipping is presented. It is based on the inversion of a nonlinear dynamic system varying over time. The inverse system is parametrized according to a brickwall limiter. The threshold and the makeup gain are then adjusted in such a manner that the desired effect i.e. the accentuation of transients or peaks is observed at the output. The validity of the approach is confirmed in a formal listening test, in which a performance on a par with the state of the art is achieved. The application of the approach is straight forward and the effect can be tuned to meet an objective criterion, such as a sufficiently high peak-to-average power ratio or crest factor.**

I. Introduction

One of the milestones of the Digital Revolution is the advent of the Compact Disc (CD) in the early '80s. For some 30 years now, the music industry has been resorting to this storage medium to spread new releases among music enthusiasts. An advantage of the digital format is that one can increase the perceived loudness level of a recording by the use of digital signal processing techniques, such as dynamic range compression, limiting, or clipping, to virtually the maximum peak amplitude. The gift turned out to be a curse, as often enough the listening comfort of a dynamic mix was sacrificed in favor of an enveloping sound wallpaper [1]. As the vocal is usually compressed to stand out against the instruments, the mix is overcompressed to overtrump the competition. This in the meantime falling trend is also known as the Loudness War. Nonetheless, there were millions and billions of music recordings produced during the last decade that fell victim to this war. The belief behind it was that there is a direct link between loudness and sales figures, since a more consistent mix would maintain consistent attention from the listener, and so would induce to buy [2]. Now we know better, don't we? But who is to blame? The same strategy was embarked on for decades in TV and radio broadcasting [3]. Most TV commercials, e.g., are still much louder than the TV program material. In response to permanent viewer complaints, the Commercial Advertisement Loudness Mitigation (CALM) Act in the US from 2010 made the TV sets more sophisticated but did it change the habits of the distributors? Radio stations also participate in the battle for audience attention. Beyond, there is the competition factor. The use of over- or hyper-compression to boost the apparent loudness is a popular trick of broadcasters to make their station pop out of the "blue" louder than the rest without touching the volume knob [4], [5]. The principal duty of the compressor in fact is to control the peak-to-average power ratio (PAPR) of the transfer signal, so to avoid overmodulation, and to minimize the effects of a channel with a limited dynamic range, i.e. capacity. The maximum PAPR is usually subject to legal requirements of the specific country. A strong reduction of the PAPR can also be observed for too many music recordings nowadays, although the dynamic range can still be sufficiently wide [6]. The root of the problem is fast limiting coupled with clipping, not so much the compression. Clipping causes flat-topped signal segments and adds distortion, while fast limiting removes transient punch and the dramatic impact from the music. In the radio station, the PAPR of overcompressed source material is reduced even more. As a result, on-air overcompressed music sounds contained, busy, and flat. When turned up to a higher volume, it might also sound distorted or simply bad, forcing the listener finally "to drop off the dial" [3], [7].

A commercial solution for over-compressed audio material to regain punch and clarity is promoted by DTS [8]. Walsh *et al.* explain the technical details in [9]. The technique aims at restoring the dynamics of modern music recordings through accentuation of transient signal components which need to be tracked. A similar algorithm is presented by Zaunschirm *et al.* in [10]. The basic ideas were previously elaborated by Goodwin and Avendano [11]. On the other side, in reaction to the CALM Act, Dolby Volume was brought on the market [12]. Its claim is to provide persistent volume across different programs but it also augments the signal's dynamic range [12]. An approach inspired by image inpainting is pursued by Adler *et al.* in [13]. The proposed framework recovers clipped signal portions but only if their location is known. In addition, to achieve better results, the maximum signal amplitude must be known as well, which requires user intervention and several trials and errors. For this reason, the framework cannot be put into practice on line. Furthermore, signal portions that underwent soft limiting, and thus are bent but not missing, cannot be restored. More counterarguments are put forward by the high computational complexity of Orthogonal Matching Pursuit and by the rather high memory requirements for a

S. Gorlow is currently with the Music Technology Group (MTG) at Pompeu Fabra University, 08018 Barcelona, Spain. This work was partially funded by "Aquitaine Science Transfert" through the Ameadys project (PJ-2012-814).

large dictionary. Some other proprietary techniques can also be found in [13].

The paper is organized as follows. Section II gives a brief overview over the operation of a broadband compressor and explains how clipped audio relates to it. Section III presents our novel technology making reference to our previous work. Section IV frames the issue of image shifting that occurs in stereo sound and explains the remedy. A comparison between our technology and existing solutions is drawn in Section V. The validity of our approach is checked in a formal listening test, the outcome of which is briefly discussed. Section VI has a focus on the automation of parameter adjustment, which is to make the technology end-user friendly. Section VII finally concludes the paper and points out certain advantages of our technology over existing solutions.

II. Dynamic Range Compression

A. Feed-forward broadband compression

Dynamic Range Compression (DRC) is a sound processing technique that attenuates loud sounds and/or amplifies quiet sounds, which in consequence leads to a reduction of an audio signal's dynamic range. The latter is defined as the difference between the loudest and the quietest sound measured in decibel (dB). Throughout the paper, we mean *downward* compression when we speak in vague terms of "compression". Downward compression *attenuates* sounds above a given threshold while leaving sounds below the threshold unchanged. Fig. 1 shows such a digital compressor model. Its operation is as follows. The input signal is split and a copy is sent to the side chain. The detector calculates the level of the sidechain signal using the root mean square (RMS) or peak as a measure, while its reactivity to the current input is controlled by the attack and release times. This sidechain signal level is compared to the threshold level and, for the case it exceeds the threshold, a scale factor is calculated which corresponds to the ratio of the input level to the output level. The knee determines how quick the compression ratio is reached. At the end of the side chain, the scale factor is fed to the smoothing filter that yields the gain. The time response of the smoothing filter is controlled by another set of attack and release times. The gain control applies the gain to the input signal and adds the makeup gain to bring the output signal to its final level. For a definition of loudness and its measurement refer to [14], [15]. More about compression and other designs can be found in [16].

B. Brickwall limiting

A brickwall limiter is a special type of a compressor which is commonly found at the end of the mastering stage in music production and in broadcast applications. It makes sure that the audio never exceeds the maximum allowed level. A brickwall limiter or "clipper" is characterized by a very fast attack time, a fast release time, a very high compression ratio and it operates on the signal's full bandwidth. A clipper can hence be easily fitted into the model from Fig. 1. Our idea is to "invert" the brickwall limiter, i.e. to *declip* the broadcast signal, and in so doing to restore its dynamics.

III. Inversion of Dynamic Range Compression

Our novel and unique technology is the "decompressor". It is based on a mathematical breakthrough in the determination of how to invert a nonlinear dynamic system that varies over time [17], [18]. Knowing the parameters of the compressor, it completely and accurately inverts dynamic range compression giving back the original dynamics to the squeezed signal. The decompressor has as input an audio signal and the parameters of a compressor. It uses these to generate a signal which, if it was compressed with the given parameters, would correspond to the input signal. Hence, it can be used to completely undo compression with minimal metadata. Or, it can also be used to add dynamics to a broadcast signal, regardless of whether the signal is actually compressed in the first place. It requires a relatively low computational effort and has zero delay due to pure time-domain processing. So far, the technology was implemented and tested in C/C++. Fig. 2 shows a graphical front end that facilitates its use. It is also available as a Virtual Studio Technology[1] (VST) effects plugin.

IV. Stereo Sound

To avoid sudden shifts in the stereo image, it is imperative that an equal amount of gain is applied to both channels of a stereo signal, which is also referred to as "stereo linking". It is achieved, e.g., by calculating the required amount of gain reduction for each channel independently, and by applying the larger amount to both channels. This strategy is embarked on in our decompressor. Hence, we decompress both channels of the input signal separately using the same settings and so we obtain two decompressed input samples. Then, we compress the latter again and compare the recompressed samples with the original input samples. The gain of the channel with the compressed sample being equal to the original sample is our sought-after gain. The inverse gain is used to decompress the sample in the complementary channel, so that both channels are equally amplified.

V. Proof of Concept

To validate our approach, we conducted an experiment. Our proclaimed objective was to revive the dynamics of (heavily) compressed but principally clipped audio material using the decompressor. The latter was parameterized according to a brickwall limiter. The attack was set to 3.3 ms, the release to 24 ms, and the ratio to 20 : 1. The knee was set to "hard" and the detector was adjusted to peak sensing. The threshold was set to -10 dB, while the makeup gain was initialized with 10 dB, i.e. the same but unsigned value. This was to ensure that the input signal's amplitude peaked at but did not exceed the threshold, and so decompression was at the edge of becoming active. The audio material was normalized to 0 dB relative to the peak level. The makeup gain was then gradually decreased until the desired effect was observed.

The explanation is as follows. When the makeup gain has a smaller value than the threshold, the potentially clipped peaks

[1] http://www.steinberg.net

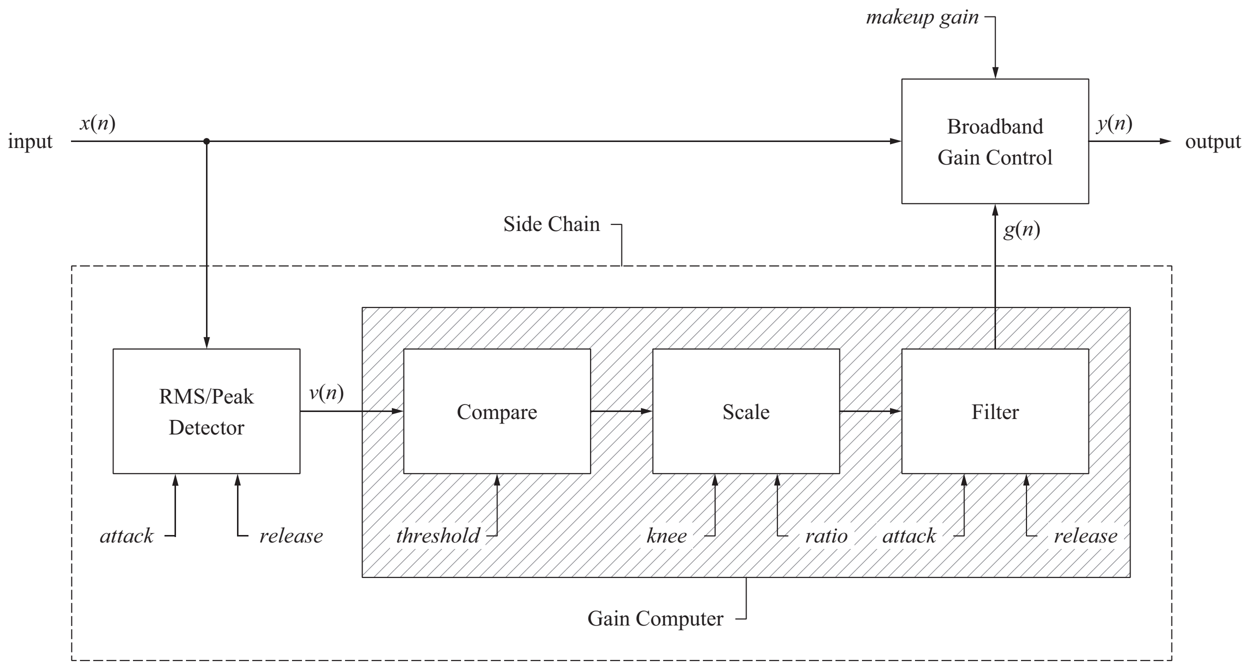

Fig. 1. Basic feed-forward broadband compressor model.

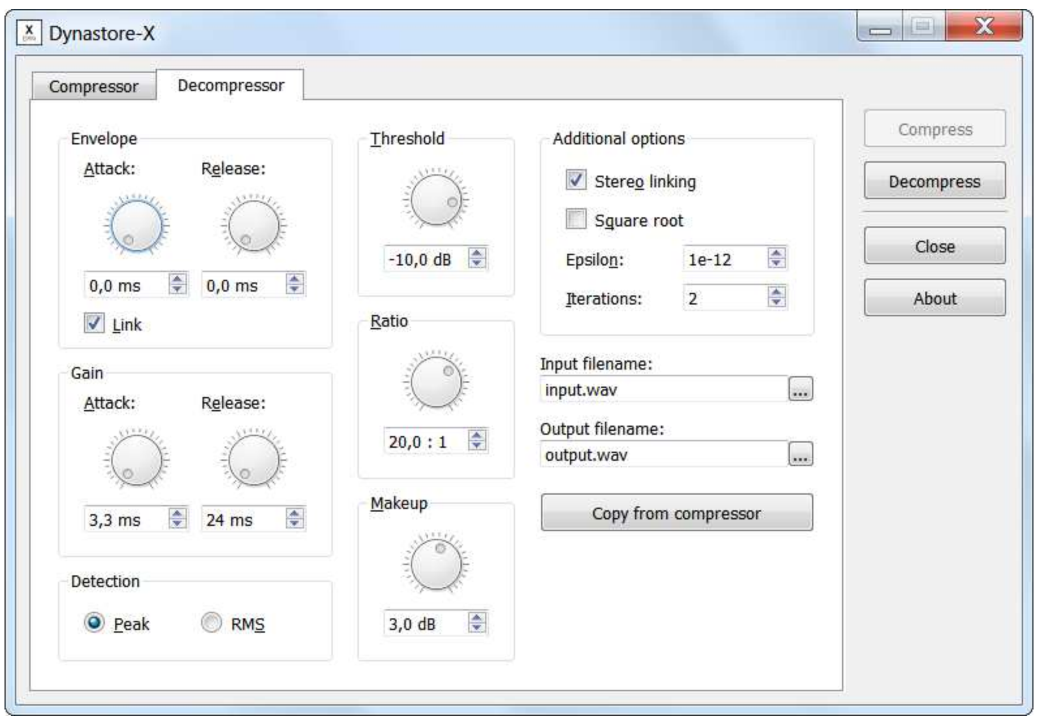

Fig. 2. Compressor/decompressor front end "Dynastore-X".

surmount the threshold and the decompressor acts on them as an *inverted* brickwall limiter or a declipper. On that account, the clipped peaks are restored. The observed effect is stronger the greater the difference between the makeup gain and the threshold.

A. Test setup

A listening test was carried out at two sites, in France and in the UK. The test panel consisted of 10 subjects. These were professional sound engineers, music producers, and musicians with a few amateurs among them. The test material consisted of 10 music titles of roughly four different genres that can be labeled as pop, rock, metal, and R&B. The titles were chosen based on the crest factor according to [6]. Four systems were compared against each other: the commercial release ("CD"), the decompressor ("Dynastore-X"), GefenTV[2] Digital Audio Decoder ("Dolby Volume"), and Zaunschirm *et al.*'s transient modifier which is closely related to DTS Audio Restoration. The adopted protocol corresponded largely to the ITU-R BS. 1534-1 multi-stimulus test [19] but without a hidden reference or anchor. The subjects were asked to judge the four systems on a "0–100" scale ("the higher the better") according to their personal preference. They were further instructed to focus on the transients coming from percussive instruments. The audio was played back on studio monitors and headphones. Also, it should be noted that the output from GefenTV's decoder was taken "as is" and that the transient modifier was operating on six frequency bands with an adaptive threshold and optimized amplification gains.

B. Test results

The results from the listening test, shown in Fig. 3, can be summarized as follows. If the peak-to-average power ratio of the CD release is below 10 dB, i.e. "low", the decompressed signal is appreciated more than the original. Yet if the PAPR is above 10 dB, i.e. already "high", the decompressed signal is appreciated less. In that case, the perceived effect literally feels as if the drummer was thrashing your eardrum, i.e. the percussive elements are overstressed. The "transient modifier" has comparable scores, whereas Dolby Volume has the worst scores. In a questionnaire, subjects described Dolby Volume as making the mix sound "dull" or "bassy", whereas the other two systems were approved to give more "punch" and to add "clarity" and "definition" to the mix. All in all, the experiment validated our initial idea and confirmed the positive effect of decompression on heavily compressed and clipped audio. To compare the effect that each of the three systems has on the CD signal, see Fig. 4. Dolby Volume does not accentuate the transients alone, and thus the relation between the processed low- and high-PAPR audio is similar to the unprocessed CD audio: The high-PAPR audio is still preferred, irrespective of being processed.

VI. Automatic Threshold Adjustment

Before the technology can be put into a broadcast receiver or audio equipment in general, it must be adapted for the end user. So, what we need is a control mechanism which adjusts the threshold and the makeup gain autonomously, i.e. without user intervention. Then, it would be sufficient to specify how dynamic the output signal should be, for instance in terms of the PAPR, and the autonomous control mechanism would do all that is necessary to comply with one's wish. For this, one has to track the peak amplitude of the input signal, to create enough headroom for the restored peaks, to monitor e.g. the PAPR at the output and at the input, and finally to adjust the threshold in such a way that a positive effect is perceived. As it was found out in the above experiment, an accentuation of transient components that is too strong is to be avoided since it can lead to a loss of the listening comfort, especially when the PAPR is already high (see Fig. 3, shaded bars).

So far, we have implemented an initial solution that needs to be tested more and optimized. At this point that much can be said: the preliminary results look very promising and we should come up with a fully automatic threshold adjustment mechanism soon.

VII. Conclusion

Just like dynamic range compression, "declipping" is quite a subtle effect and must be handled with care. The dynamics of a music piece when expressed as a peak-to-average power ratio seem to have a sweet spot. When overstepped, the sound quality degrades. To find the sweet spot is not guaranteed, as it is highly subjective, and so it can only address the average user. But this is exactly where the end user can benefit from the decompressor. The decompressor can be custom tuned to add as much dynamics as it is wished for by the listener, for each pair of ears individually. This concept was successfully validated in the reported experiment.

The main advantage of the decompressor is that it requires no sophisticated analysis of transients. All one needs to do is to specify the parameters of the compressor that one seeks to invert and the portion of the input signal to be treated. It can be quite astonishing to see what the decompressor can bring out of clipped peaks without any prior knowledge. Naturally sounding drum sequences are one example. Another example are crescendos and decrescendos. This is beyond what can be achieved with other techniques.

Acknowledgment

The authors cordially thank Rudolf Ortner, Thomas Lund, Ben Dawson, Leigh Lawson, and everybody who took part in the listening test.

References

[1] D. Viney, "The obsession with compression," Master's thesis, London College of Music, 2008.
[2] E. Vickers, "The loudness war: Do louder, hypercompressed recordings sell better?" *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 346–351, May 2011.
[3] F. Foti and R. Orban, "What happens to my recording when it's played on the radio?" in *AES Conv. 111*, Nov. 2001.
[4] B. C. J. Moore, B. R. Glasberg, and M. A. Stone, "Why are commercials so loud? Perception and modeling of the amplitude-compressed speech," *J. Audio Eng. Soc.*, vol. 51, no. 12, pp. 1123–1132, Dec. 2003.
[5] M. A. Stone, B. C. J. Moore, C. Füllgrabe, and A. C. Hinton, "Multi-channel fast-acting dynamic range compression hinders performance by young, normal-hearing listeners in a two-talker separation task," *J. Audio Eng. Soc.*, vol. 57, no. 7/8, pp. 532–546, Jul. 2009.
[6] R. M. Ortner, "Je lauter desto bumm! — The Evolution of Loud," Master's thesis, Donau Universität Krems, 2012.
[7] N. B. H. Croghan, K. H. Arehart, and J. M. Kates, "Quality and loudness judgments for music subjected to compression limiting," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1177–1188, Aug. 2012.
[8] *DTS Audio Restoration: Compressed Audio Regains Punch and Clarity*, DTS, 2010.
[9] M. Walsh, E. Stein, and J.-M. Jot, "Adaptive dynamics enhancement," in *AES Conv. 130*, May 2011.

[2]http://www.gefen.com/gefentv/

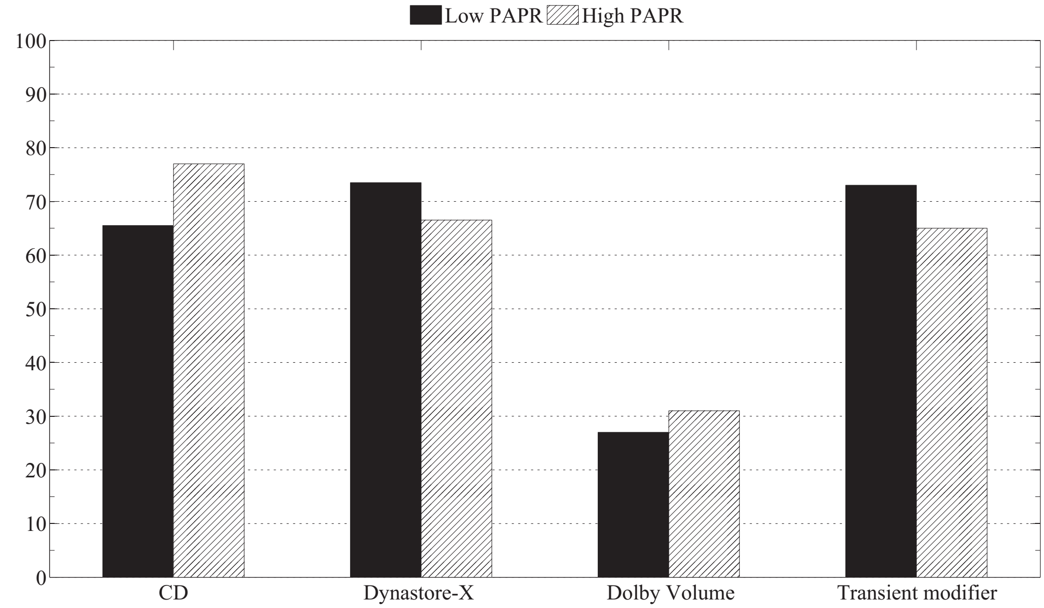

Fig. 3. Listening test results (median opinion scores).

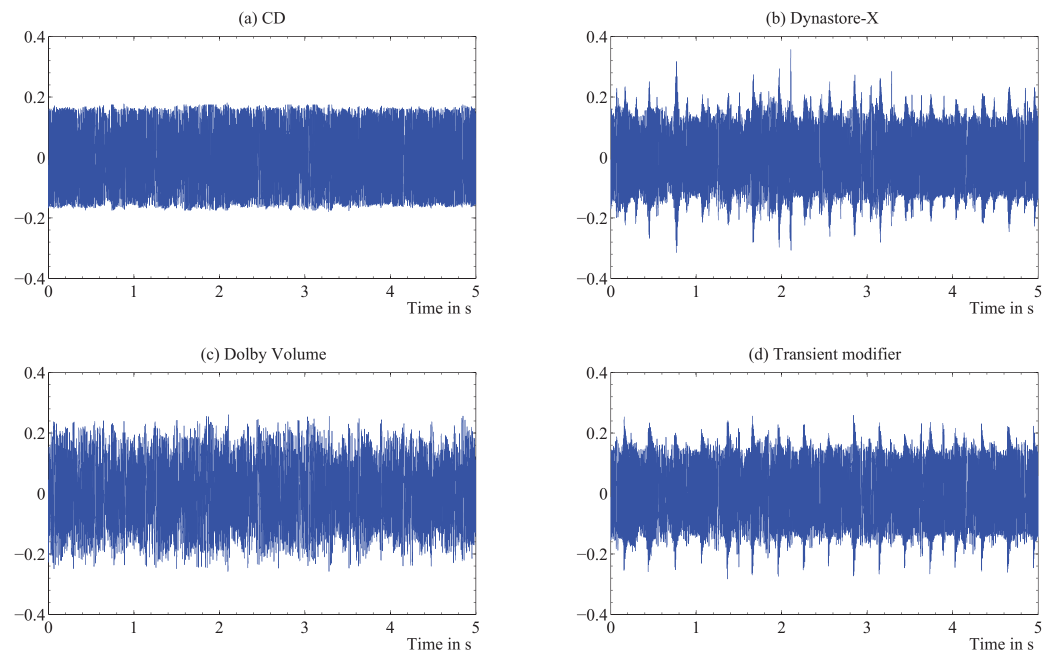

Fig. 4. Metallica's "My Apocalypse" from the infamous "Death Magnetic" album (excerpt).

[10] M. Zaunschirm, J. D. Reiss, and A. Klapuri, "A sub-band approach to modification of musical transients," *Comput. Music J.*, vol. 36, no. 2, pp. 23–36, 2012.

[11] M. M. Goodwin and C. Avendano, "Frequency-domain algorithms for audio signal enhancement based on transient modification," *J. Audio Eng. Soc.*, vol. 54, no. 9, pp. 827–840, Sep. 2006.

[12] *Dolby Volume: An Innovative Solution to Inconsistent Volume Issues*, Dolby Laboratories, 2010.

[13] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Audio, Speech, Language Process.*, vol. 20, no. 3, pp. 922–932, Mar. 2012.

[14] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.*, vol. 5, no. 2, pp. 82–108, Oct. 1933.

[15] *Algorithms to measure audio programme loudness and true-peak audio level*, ITU-R, Aug. 2012, rec. ITU-R BS.1770-3.

[16] D. Giannoulis, M. Massberg, and J. D. Reiss, "Digital dynamic range compressor design—a tutorial and analysis," *J. Audio Eng. Soc.*, vol. 60, no. 6, pp. 399–408, Jun. 2012.

[17] S. Gorlow and J. D. Reiss, "Model-based inversion of dynamic range compression," *IEEE Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1434–1444, Jul. 2013.

[18] ——, "A method for inverting dynamic range compression of a digital audio signal," International Patent Application 13 050 461, 03 04, 2013.

[19] *Method for the subjective assessment of intermediate quality level of coding systems*, ITU-R, Jan. 2003, rec. ITU-R BS.1534-1.