Observations on Guitar Music Produced by AI Reverberation and Professional Sound Engineer

CHEN, Manni; LINDBORG, Per Magnus

Published in:
International Journal of Music Science, Technology and Art

Published: 01/01/2023

Published version (DOI):
10.48293/IJMSTA-93

Publication details:
CHEN, M., & LINDBORG, P. M. (2023). Observations on Guitar Music Produced by AI Reverberation and Professional Sound Engineer. *International Journal of Music Science, Technology and Art*, *5*(1), 1-7. Advance online publication. https://doi.org/10.48293/IJMSTA-93

International Journal of Music Science, Technology and Art
(IJMSTA) 5 (1): 1-7, Music Academy "Studio Musica", 2022 ISSN: 2612-2146 (Online)

Observations on Guitar Music Produced by AI Reverberation and Professional Sound Engineers

Manni CHEN[1]✉, Per Magnus LINDBORG [2]✉

[1]School of Creative Media, City University of Hong Kong, Hong Kong SAR
[2]School of Creative Media, City University of Hong Kong, Hong Kong SAR
[1]✉ Corresponding author: manni.chen@my.cityu.edu.hk
[2]✉ Corresponding author: pm.lindborg@cityu.edu.hk

ARTICLE INFO

Received: December 30, 2022
Accepted: January 10, 2023
Published: January 16, 2023
DOI: https://doi.org/10.48293/IJMSTA-93

Keywords:
Reverberation
Listening test
Music perception

ABSTRACT

Artificial intelligence (AI) technologies have been applied in music production to create various sound effects, including reverberation. However, observation on such applications has not yet been fully explored in research studies. This paper reports results from a study comparing reverberation processing on six recordings of guitars, with musical phrasing, made by an AI software and two professional sound engineers. Audio features were extracted using the MIR Toolbox, and perceptual ratings on semantic scales were collected in two listening tests (N = 10, N = 33). Logistic regression was carried out on the two datasets in parallel. An increase in perceived Wetness or decrease in perceived Clarity was associated with a higher probability that the reverberation was made by the AI rather than a Human. For extracted audio features, lower Brightness, Rolloff, and Centroid, which are all indicators for a darker, low frequency emphasized sound, were more likely made by the AI. This study contributes to an understanding of the differences between AI- and human-generated audio effects used in music production.

1 Introduction to the paper

As AI technologies are increasingly implemented in sound effects plug-ins used in the studio music production pipeline, machines start to undertake part of creative sound work in an attempt to "simulate human skills" [1]. Contributing to a growing understanding of these technologies, our research project aims to compare sound effects made by AI with those made by human sound engineers. In this paper, we focus on reverberation. Previous studies have explored the perceived level of reverberation [2], and the amount of reverberation and early decay time [3]. Piotrowska [4] evaluated the musical output from online audio mastering platforms through objective measurements of extracted audio features. However, the perceptual effects of AI-based reverberation have not been fully explored. Since the "perceived result at the listeners' end" [5] is a crucial issue for music production to reflect multiple aspects of musical work, in this case, to deepen the understanding of AI audio effects, observations can be conducted based on musical features and perceptual listening tests.
The present study focuses on AI reverberation and integrates objective and perceptual evaluations. The objective assessment employs a range of musical features using the MIRtoolbox [6] and the subjective assessment is built upon perceptual listening tests based on ratings of stimuli using semantic scales. Therefore, this project neither focuses on how the AI was trained nor looking into the implementation of the AI reverberation, but only on the observations of the guitar music produced by AI reverberation in the pointview of music production practice by sound engineers who adopt a series of audio effects for audio mixing and mastering in daily life. Designed for mimicking the behaviour of sound engineers during work and bringing the result that is close to what human engineers produce [7], these AI-driven tools are considered as cost and time saving tools by the audio industry [8]. In this case, it is meaningful to look into how humans perceive the music output by AI-driven audio effects like reverberation to facilitate the comprehension and use of AI reverberation in actual music production work, by the

means of conducting listening tests on the music output by AI audio effects, where researching the construction of AI system is outside our scope.
In the following parts, we will first describe the preparations of musical stimuli and introduce the testing tasks, after which we illustrate the designs of objective. measurements and listening tests. Subsequently, features will be extracted from the audio signal of both AI output and human output to make an objective comparison.

2 Materials and methods

For the preparations of musical materials, we selected six guitar samples from our own studio recordings and from free online mixing resources (electric and acoustic guitars). Each of the samples contains a musical phrase of between 15 and 20 seconds. Since we are interested in the AI reverberation that is available for all the common sound engineers working on music production to arouse their deeper discernment of reverberation powered by AI, instead of an AI reverberation that is only reachable for small groups of people who tested or used it, we selected a commercial AI reverberation product which is Sonible's smart:reverb processor [9]. According to the introduction of smart:reverb on the official website, this AI reverberation possesses an AI-based system to analyse input signals and add customised reverb based on the spectral and temporal characteristics of the input signal, which means that all parameters in reverberation can be automatically tuned by AI-driven system on the basis of features for input music signals. Fitting the original musical materials, we adopted the "Guitar" profile to calibrate the AI reverb to guitar samples. Then, the source-adaptive reverb accordingly sets all its parameters based on the computing on the attributes of guitar music. We generated three versions: one with reverberation added by Sonible's smart:reverb processor [7], one with reverberation added by two trained sound engineers; and one without any reverb, i.e., the original. The human engineers were required to manipulate the samples using only those reverberation plug-ins that they were familiar with in commonly available software, such as Pro Tools. They were not allowed to edit the samples or use other audio effects, such as equalization or compressors. Followingly, the 18 stimuli (6 samples x 3 versions) were explored in objective measurements and perceptual listening evaluations.

Table 1. List of music samples included in the two listening tests

NO.	Genre	Guitar	Description
1	Rock	Electric	Finger picking, wild
2	Classical	Acoustic	Finger picking, serious
3	Folk	Acoustic	Alternative picking, peaceful
4	Pop	Acoustic	Rhythm
5	Rock	Acoustic	Sweeping
6	Classical	Acoustic	Finger picking, tender

We extracted musical and audio features with MIRtoolbox 1.7.2 and a loudness meter plug-in to extract and compare musical features of musical excerpts. We then conducted two online listening tests. We used a comparative mean opinion score (CMOS) [8] together with a set of semantic differential rating scales described below. In this paper, we include multiple stimuli to cover various musical styles. Among six raw musical materials, two of them are obtained from professional sound engineers' recordings, and the remaining ones are from Open Multitrack testbed [9] and "Mixing Secrets" Free Download Library [10]. Table I presents raw musical materials along with corresponding genres and short descriptions.

2.1 Objective measurement

For the objective measurements, we selected features of spectral shape and dynamic envelope: *RMS*, *Brightness*, *Rolloff*, *Centroid*, and *Irregularity* [6]. *RMS* computes the general energy of the signal by taking the root average of the square of the amplitude, while the remaining features compute in terms of the spectrum. Among them *Brightness* measures the percentage of the energy in total energy above the cut-off frequency (using 1500hz by default), *Rolloff* refers to finding the frequency such that 85% of total energy in default is contained below that frequency. *Centroid* is an important description of the spectrum represented by the geometric centre (centroid) of the spectrum distribution, reflecting the frequency where the most energy is located. *Irregularity* is the degree of variation of the successive peaks of the spectrum. There are also many features listed in [6], however, five objective features containing the general descriptions of dynamic and spectral dimensions are enough as a complement to represent the computing characteristics of the musical signals. Plenty of measurement of objective features may distract the focus on human perception in listening test, especially for those statistical features that are only different in mathematics but make no perceivable difference for listeners are not meaningful enough in music production practice.

2.2 Listening tests

As for the perceptual descriptors, we conducted two listening experiments, where participants listened the 18 sound stimuli and rated them according to how they think they sounded on semantic scales. The participants were required to wear professional monitor headphones and perform the experiment online remotely through a questionnaire in front of a computer screen. Prior to the experiment, participants read a standard explanation of the task and were told to respond to each question at the first thought without hesitation to avoid repeating and choosing between options. Apart from the questions on evaluating processed materials with reference in the forms of rating, the questionnaire also included general questions on age, gender, and professional experience of sound engineering.

The first listening test (N=10) was conducted with QuestionPro presented in simplified Chinese, consisted of a group of university undergraduate students in mainland China, aged 19 to 22 and majoring in sound-focused programmes and having substantial experience with sound engineering. Participants were asked to evaluate the perceived reverberation of the 18 stimuli that contain unprocessed musical materials and the stimuli processed by sound engineers and machine, presented in a randomized order in a repeated-measures design. Each was evaluated on 15 semantic rating scales [8] labelled 混响量 (hùnxiǎngliàng, "reverb amount"), 温暖感(wēnnuǎngǎn, "warmth"), 明亮感 (míngliànggǎn, "brightness"), 粗糙感(cūcāogǎn, "roughness"), 空气感 (kōngqìgǎn, "airiness"), 甜美感 (tiánměigǎn, "sweetness"), 纵深感 (zòngshēngǎn, "depth"), 尖锐感 (jiānruìgǎn, "harshness"), 喜爱度(xǐàidù, "likeability"), 清晰度 (qīngxīdù, "clarity"), 混响尾巴长度 (hùnxiǎngwěibāchángdù, "length of reverb tail"), 湿度 (shīdù, "wetness"), 距离感 (jùlígǎn, "distance"), 厚度感 (hòudùgǎn, "thickness"), and 扩散感 (kuòsàngǎn, "diffusion").

We analysed the response data with Principal Component Analysis (PCA) following the technique described by [11] in order to reduce the 15 scales above to a parsimonious lower-dimensional model. The optimal result has two components that explain 95% of the variance in the data. For the purposes of having good precision in the second listening test, we supplemented them with another pair of orthogonal components placed at a 45° angle from the first two. Thus, the plane could be robustly spanned by four bipolar constructs, which we named "DryWet", "VagueClear", "DenseDiffused", and "ColdWarm". Wet refers to the sound with reverb in our case, versus dry sound (original), which can be represented by a percentage of dry (original) or wet (with reverb) sound in the mixture e.g. 100% wetness means to only leave the sound that reverb was applied. Therefore, "DryWet" evaluates the amount of reverb by perception. "VagueClear" responds to the perception of *Clarity*, "DenseDiffused" relates to the sense of broadness of sound source, while "ColdWarm" associates with the perception of low-mid frequency. These four constructs are anchored by eight labels, as illustrated in Fig. 1. Based on Fig. 1, *Wetness* and *Clarity* as well as *Diffusion* and *Warmness* are orthogonal, which is somehow divergent to common experience in music production where when the sense of wetness and diffusion enhances apparently, the degree of *Vagueness* and Warmness increases, respectively. This is because these orthogonal results are based on the perception on the limited sound sources, where the spectrum of the instrument and added reverb are finite, for instance, in our case, the range in the amount of reverb is impossible to be extremely large so as to make the *Wetness* and *Clarity* obviously associated with each other. But when the amount of reverb is tiny, the *Clarity* is not evidently affected in perception. Therefore, the orthogonal results are different from common sense mainly since the narrower range of reverb is applied to our specific samples compared to more various situations in music production.

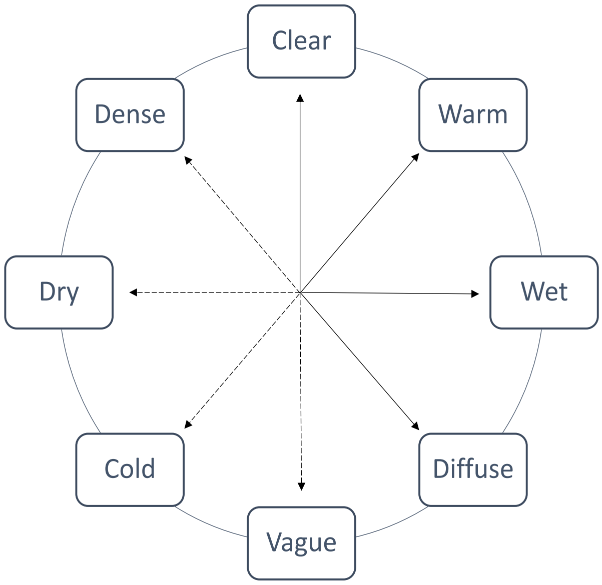

Figure 1. Result of PCA from the first listening test (N=10) for creating four bipolar semantic scales.

We then conducted the second online listening experiment (N=33) similarly to what was described above; however, this time, we used four semantic scales (清晰度 (qīngxīdù, “clarity”, 扩散感 (kuòsàngǎn, “diffusion”, 湿度 (shīdù, “wetness”, 温暖感 (wēnnuǎngǎn, “warmth”). The questionnaire was made in three language versions (number of responses in parentheses): simplified Chinese, traditional Chinese, and English. In this listening test, we recruited volunteers from both industry and academia. The average age was 30 years, ranging from 25 to 60. They were professionals currently working in the creative audio industries (N=20) or lecturers in higher education or graduate students with research on sound and/or music production (N=13). All reported normal hearing and experience in music production. In what follows, only the data from the second listening test was included.

3 Analysis

The analysis seeks to answer the primary question: what are the differences between the musical work processed by reverberation made from AI sound effects and sound engineers? The analysis started with focusing on stimuli and the check for inter-raters' reliability and listening agreements. Then, we conducted logistic regressions to explore the objective and perceptual datasets.

3.1 Data structure

The data in the second listening test consisted of 33 listeners assessing 18 stimuli (6 samples x 3 versions) on four perceptual semantic scales in a repeated-measures design. Therefore, there are 4752 points in the perceptual data and similarly, 5940 data points in the objective dataset of extracted frames of 5 audio features.

3.2 Reliability and agreement

We first conducted an intra-rater reliability analysis by correlating each participant's ratings across the four scales, six samples, and three versions. The cut-off point was set at alpha = 0.1 for Spearman's correlation, in each case. Four participants having p > 0.18 were excluded. For the 29 participants kept in the data set, we conducted an inter-rater agreement analysis. First, the matrix of grand means across all participants for the six samples and three versions was determined. Then, we took the average for each participant of their two blocks of ratings. Finally, we calculated the Spearman correlation between each participant's ratings and the grand mean. With a cut-off set at alpha = 0.1 as before, this led to another two participants being excluded, having p > 0.12. The reliability and agreement analysis reflects the consistency of the data recorded by each rater and the degree of agreement among raters, respectively. This means that, in our case, the reliability analysis measures the consistency of each participant about whether they keep consistent ratings in the same questions in terms of repeated-measures design, and then we exclude the raters who fail to retain similar judgments. The agreement analysis indicates the degree of listeners' consensus on the ratings of guitar samples. Cronbach's alpha was applied to test the degree of agreement and *Wetness* and *Clarity* indicating good and questionable agreement levels respectively (alpha = 0.81 and 0.61), remaining ones (*Warmness* and *Diffusion*) report poor agreement so they are not test for mean estimate of listening agreement among participants in Fig. 2 for further analysis. These two analyses for reliability and agreement help to establish the trustworthiness and dependability of the data we collected from the subjective listening tests.

The circumplex (Fig. 1) is a theoretical construction, which allows us to calculate values on its principal dimensions from multiple scales laid out at angles. Each 'additional rating scale' adds more precision to the two main concepts, namely *Wetness* and *Clarity*. They were calculated as follows:

$$\boxed{Wetness = \frac{DryWet + 0.71 * (DenseDiffused + ColdWarm)}{3}} \tag{1}$$

$$\boxed{Clarity = \frac{VagueClear + 0.71 * (DiffusedDense + ColdWarm)}{3}} \tag{2}$$

3.3 Wetness and Clarity in stimuli

"Wetness" and "Clarity" are the words that frequently appear in the interface of reverberation plug-ins; therefore, they were used as a semantic scale in our listening experiments, and they turned out to be the reliable constructs among the four we utilized based on the reliability and agreement analysis.

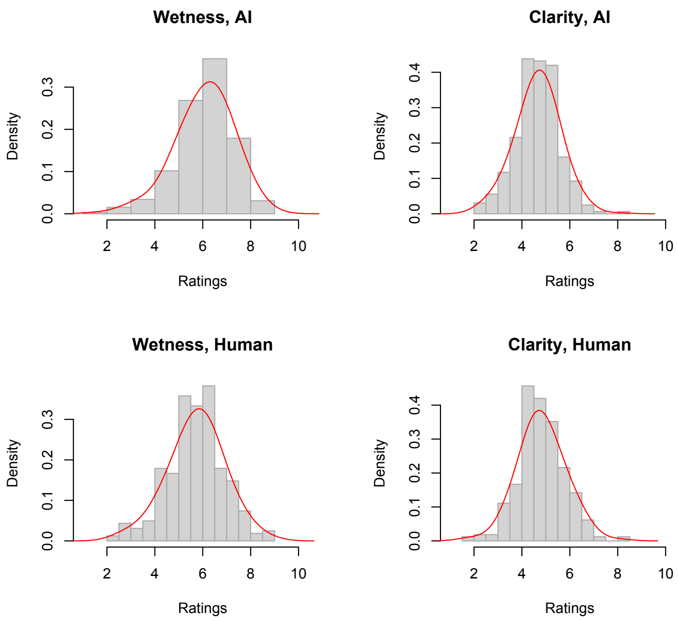

Figure 2. Histograms of ratings of *Wetness* and *Clarity* across AI and Human samples, respectively.

Fig. 2 presents the histograms of *Wetness* and *Clarity* in AI-generated and human-processed versions. Looking further into the data, we observed that *Wetness* was evaluated higher after adding reverberation. The original Guitar 2 and raw Guitar 5 were assessed as having the two highest *Wetness* scores, while the sample Guitar 1 was considered the driest. However, the highest and lowest wet musical materials were not assessed the same after processing them. In the AI-generated version, although the processed Guitar 1 is still the lowest wet and Guitar 5 is the highest, Guitar 6 expressed significantly wetter than Guitar 2, which is one of the two highest raw materials. In respect of the human-processed version, the Guitar 1 which is the lowest one in the original version and the AI-generated version, became the wettest among all human-made musical stimuli. The Guitar 5 which is the wettest in raw material and the AI-generated version, transferred to the driest sample after the sound engineer adding the reverberation.

Overall speaking, AI-generated versions were considered wetter, since apart from the Guitar1, the other five samples were evaluated wetter in the AI-generated version than human-processed version. Additionally, the order of *Wetness* in AI-generated versions correspond more to the raw musical materials, in contrast, the change in the human-processed version is more dramatic. This transformation between processed and unprocessed music work delivered a certain degree of differences in the pattern of adding reverberation between machine and human sound engineers.

3.4 Logistic regression

To further observe the relations between semantic scales and investigate the interactions between scales and two processed versions, logistic regression was applied to see whether the scales allow a prediction of progeny (AI-generated versus human-processed version). We transformed two different processed versions into a dichotomous factor; the AI version was coded as 0 while the Human version was coded as 1. The predictors were *Wetness* and *Clarity*, and progeny was the predictand (AI or Human). Table 2 reports the significant semantic scales and objective features of AI-generated versions versus human-processed versions, meaning that only the variables that are significant ($p < 0.01$) in the reduced model are shown, which is the reason why the *RMS* and *Irregularity* that the corresponding p-values are insignificant are excluded from the table. In Table 2, the p-value and beta coefficients in regression analysis work together to indicate which relationships in our model are statistically significant and how they relate to each other. To be specific, the beta coefficient describes the relationships between the independent variables, including AI-generated and human-processed versions, and the dependent variables, including *Brightess*, *Rollof*, *Centroid*, *Wetness* and *Clarity*. The negative coefficient means as the independent variable increases, the dependent variable tends to decrease, while the positive coefficient suggests as the independent variable increases, the dependent variable tends to increase. The coefficient value signifies the degree of the mean of the dependent variable changes given a one-unit change in the independent value when holding other variables constant in the model. Therefore, since the p-value displays whether these relationships are statistically significant, the semantic scales and objective features listed in the table are all associated with human perception of the reverbed guitar samples as AI- or human- generated versions.

Table 2. Report of logistic regression models in terms of significant semantic scales and objective features of AI-generated versions versus human-processed versions

FEATURE	Estimate	ß coef.	odds ratio	Z	p
Brightness	-4.02	-0.64	0.53	-4.48	0.000007 ***
Rolloff	0.00	-1.33	0.26	-2.70	0.0069 **
Centroid	0.00	1.94	6.97	3.61	0.00031 ***
Wetness	-0.53	-0.53	0.48	-8.44	< 2e-16 ***
Clarity	0.21	0.21	1.23	2.67	0.0076 **
Brightness	-4.02	-0.64	0.53	-4.48	0.000007 ***

* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$

Moreover, an increase in perceived *Wetness* by one standard deviation is associated with an increased probability that the reverberation was made by AI. The beta coefficient is -0.53, indicating a 52% increased odds ratio for AI over Human, which is the same case for *Brightness*, *Rollof* and *Centroid*. However, an increase in perceived Clarity by one standard deviation is associated with an increased probability that the reverberation was made by Human. The beta coefficient is 0.21, indicating a 23% increased odds ratio for Human over AI.

4 Discussions and conclusions

This paper evaluated the musical work processed by AI reverberation compared to the reverberation added by human sound engineers. Through two listening tests where participants rated four semantic scales based on 18 stimuli in a repeated-measures design and the objective measurement of 5 audio features using MIRtoolbox, the musical materials which were only processed via reverberation by machine and human can be analyzed to compare the characteristic of reverberation added from different entities. Regarding the data of the perceptual experiment, we reduced 15 semantic scales from the first listening test to merely four semantic scales, only further reduced to two dimensions.

We used descriptive words that have been applied in the interface of reverberation, the result suggests that *Wetness* reached a good agreement and reliability among our participants, all having substantial sound engineering experience. Then we looked further into *Wetness* and *Clarity* to observe the differences between three version groups. We found the AI-generated versions tend to have higher mean scores of *Wetness* than human-processed versions and also tend to remain in the order of original degree of *Wetness* as the original versions, while the changes of *Wetness* in human-processed work seem to be more unpredicted, which is probably because of the human understanding of appreciation of music. This, to some extent, displays the differences of adding reverberation between human and machine. When it comes to the logistic regression, the results indicate that the odds of evaluation on human-processed reverberation increases with the assessment of Clarity while the evaluation on AI-made reverberation increases with *Wetness*, which verifies the findings of previous stimuli analysis regarding *Wetness*. *Brightness*, *Rollof* and *Centroid* that are related to lower frequency tend to indicate AI-made reverberation. This work has the potential of being a valuable contribution as it can shed light into how humans perceive AI-generated audio effects such as reverberation differently to human-generated ones.

In future work, we evaluate the musical output from AI audio effect generators in comparison to that of experienced sound engineers regarding other audio effects like compression or equalization. In general, we believe that more perceptual data and audio features should be included in the evaluations to facilitate the reliable data analysis and in order to draw valid and comprehensive conclusions on the music processed by intelligent audio effects.

References

[1] Birtchnell, T. and Elliott, A. (2018). Automating the black art: Creative places for artificial intelligence in audio mastering. *Geoforum* **volume** (96): pp. 77–86

[2] Cauchi, B., Javed, H., Gerkmann, T., Doclo, S., Goetze, S., & Naylor, P. (2016). Perceptual and instrumental evaluation of the perceived level of reverberation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*

[3] De Man, B., McNally, K., & Reiss, J. D. (2017). Perceptual evaluation and analysis of reverberation in multitrack music production. *Journal of the Audio Engineering Society.*

[4] Piotrowska, M., Piotrowski, S., & Kostek, B. (2017, May). A Study on Audio Signal Processed by" Instant Mastering" Services. In *Audio Engineering Society Convention 142. Audio Engineering Society*.

[5] Rumsey, F., & Berg, J. (2001). Verification and correlation of attributes used for describing the spatial quality of reproduced sound. In *AES International Conference: Surround Sound*-Techniques*, Technology, and Perception* 21/06/2001-24/06/2001. Audio Engineering Society, Inc..

[6] L. Olivier, "MIRtoolbox 1.7.2 User's Manual," University of Oslo, Norway, 2019. https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7-2.pdf

[7] Rumsey, F. (2013). Mixing and artificial intelligentce. Journal of the Audio Engineering Society, 61(10), 806-809

[8] Birchnell, T (2018). Listening without ears: Artificial intelligence in audio mastering. Big Data $ Society, 5(2), 2053951718808553

[9] Smart: reverb manual. Sonible official website. https://www.sonible.com/wp-content/uploads/2020/08/manual-smartreverb.pdf

[10] Pulkki, V., & Karjalainen, M. (2015). *Communication acoustics: an introduction to speech, audio and psychoacoustics*. (John Wiley & Sons)

[11] De Man, B., Mora-Mcginity, M., Fazekas, G., & Reiss, J. D. (2014, October). The open multitrack testbed. In *Audio Engineering Society Convention 137*. Audio Engineering Society.

[12] M. Senior, Mixing Secrets for the Small Studio (Focal Press, 2011). https://www.cambridge-mt.com/ms/mtk/

[13] Kabacoff, R. I. (2015). *R in action: data analysis and graphics with R*, 2nd ed. (Shelter Island: Simon and Schuster, NY)