EVALUATING THE PERFORMANCE OF OBJECTIVE AUDIO QUALITY METRICS IN RESPONSE TO COMMON AUDIO DEGRADATIONS

Xie He, Duncan Williams and Bruno Fazenda

Acoustics Research Center
Univesity of Salford
Manchester, UK
h.xie@edu.salford.ac.uk

ABSTRACT

This study evaluates the performance of five objective audio quality metrics—PEAQ Basic, PEAQ Advanced, PEMO-Q, ViSQOL, and HAAQI —in the context of digital music production. Unlike previous comparisons, we focus on their suitability for production environments, an area currently underexplored in existing research. Twelve audio examples were tested using two evaluation types: an effectiveness test under progressively increasing degradations (hum, hiss, clipping, glitches) and a robustness test under fixed-level, randomly fluctuating degradations.

In the effectiveness test, HAAQI, PEMO-Q, and PEAQ Basic effectively tracked degradation changes, while PEAQ Advanced failed consistently and ViSQOL showed low sensitivity to hum and glitches. In the robustness test, ViSQOL and HAAQI demonstrated the highest consistency, with average standard deviations of 0.004 and 0.007, respectively, followed by PEMO-Q (0.021), PEAQ Basic (0.057), and PEAQ Advanced (0.065). However, ViSQOL also showed low variability across audio examples, suggesting limited genre sensitivity.

These findings highlight the strengths and limitations of each metric for music production, specifically quality measurement with compressed audio. The source code and dataset will be made publicly available upon publication.

1. INTRODUCTION

Audio quality is a crucial factor in ensuring that created content is effectively communicated throughout the production process. As [1] observes, "Even for the layperson, sonic quality does matter." Similarly, [2] highlights that audio quality has a significant impact on the overall listening experience. Pop or rock genres music thrives on Home-studio recordings and proliferation of the digital audio technology. However, they are more susceptible to accumulate various distortions throughout the signal chain. Degradation can stem from a variety of sources—including, but not limited to, cable hiss or grounding noise [3], clipping artefacts introduced by effects plugins [4], glitches within digital audio workstations (DAWs) [5], and compression artefacts from lossy CODECs. These degradations can disrupt both released recordings, broadcasts and live sound. Despite their impact, the combined effects of these multiple distortions on perceived audio quality remain underexplored.

Audio quality is traditionally assessed using subjective and objective methods. While subjective evaluation is considered more accurate, it is often constrained by its time-intensive and costly nature. In contrast, objective metrics, such as PEAQ [6], PEMO-Q [7], ViSQOL [8], and HAAQI [9], are commonly utilised. These intrusive objective models typically follow a standardised process: (1) employing an auditory model to extract relevant audio features, (2) generating quality indices by comparing the processed signal to a reference signal, and (3) mapping these indices to Mean Opinion Scores (MOS) using subjective dataset training. Objective audio quality metrics are typically designed for specific applications; however, their usage often extends beyond their original domains. For instance, PEAQ, ViSQOL, and PEMO-Q were primarily developed for evaluating audio CODEC, yet studies by [10] and [11] have utilised PEAQ and ViSQOL, respectively, to assess AI-generated audio. Similarly, [12] employed PEAQ to evaluate time-scale modifications of audio. Metrics such as HASQI and HAAQI, initially intended for hearing aid applications, have also been applied to assess recording distortions [13]. A plausible explanation [14] for the field-dependent performance of these metrics lies in the training datasets and the underlying auditory models used in their development. For instance, studies such as [9] and [15] demonstrate that PEAQ can be adapted to novel applications by recalibrating with domain-specific training datasets.

This study categorises digital audio degradations that are commonly found in recording studio signal chains into four distinct types, distinguishing between noise - defined differently from the degradation in other acoustic applications - such as CODEC artefacts or hearing aid processing distortions. These categories include (1) content-unrelated pitched noise, (2) content-unrelated broadband noise, (3) content-related and inharmonic noise, and (4) temporal noise. This study builds on the extensive subjective evaluations that have already been conducted to validate these metrics. Rather than reassessing their alignment with subjective perception, the focus is on evaluating their suitability as indicators in a digital recording studio context. Building on previous work, this study examines the strengths and limitations of each metric in the context of music production. The findings help clarify how degradations impact audio quality and could potentially support the use of background monitoring metrics to quickly flag issues and target fidelity levels in practical workflows.

Section 2 provides a detailed description of the experimental methodology and the characteristics of the noise types. Section 3 presents an analysis and discussion of the findings.

2. METHOD

2.1. Degradation simulations

This section expands on the discussion of studio degradations, detailing the definitions, origins, and methods for simulating common types of noise within the signal chain. The study assumes a typical home studio setup consisting of a few microphones, a computer, and an audio interface, all built using relatively budget-friendly equipment and operating in a room without proper acoustic treatment. This scenario underscores how susceptible modern recording setups are to various types of degradation and motivates a study of their effects on audio quality.

2.1.1. Hum

Hum is a distinctive type of degradation categorised as a content-unrelated pitched noise, caused by power line interference. As noted by [16], the origins of this degradation are often linked to inadequate power-supply stabilisation in vacuum-tube-era recording equipment, or insufficient shielding of sensitive microphone cables. Structurally, the hum typically consists of a fundamental frequency (the so-called "ground loop" - 50 Hz in Europe and 60 Hz in the United States) and its harmonic components. A spectrogram derived from [17] illustrates this phenomenon, showing hum disturbances at 50 Hz and 150 Hz in a representative example. To simulate potential hum-related degradations, this study replicates the structural characteristics of hum, while varying its loudness levels. The signal-to-noise ratio (SNR) is employed as a parameter to manipulate the intensity of the hum degradations.

2.1.2. Hiss

In content-unrelated broadband noise, "hiss" in an audio recording is a steady background noise, often resembling a soft, continuous "s" sound. According to the [18], random noise must be expected in the signal chain because it is inherent in all-electric devices. The source of the hissing noise in the signal chain might be hard to locate due to its variety, such as the hiss of a microphone [19] or an overworking transducer [20]. Given the random nature of hissing noise, additive white Gaussian noise is frequently utilised in simulations, as demonstrated in previous research [21]. In this study, white noise is used to replicate the characteristics of hissing noise, with its loudness levels adjusted to simulate varying intensities. The SNR controls the strength of the white noise relative to the source signal.

2.1.3. Clipping

Digital clipping occurs when the amplitude of an audio signal exceeds the maximum threshold of the digital system, causing the over-threshold samples to be truncated to the threshold value. This process introduces nonlinear distortion, including alterations to the original signal and the generation of new harmonic components. Both [13, 22] have identified clipping as a significant source of audio degradation, with listening tests rating its impact as "very annoying". Annoyance can be rated using ISO 15.666. In this study, digital clipping is categorised as content-related and inharmonic noise. The quantity of clipping is based on a peak to threshold factor so that quieter signals are exposed to the same relative amount of clipping as louder signals would be, for any given percentage of degradation.

2.1.4. Glitch

Glitches are a common form of temporal noise encountered in music production, characterised by brief interruptions in the digital audio stream, often referred to as "dropouts" [5]. These interruptions produce audible artefacts, such as cracks or pops, during recording or playback. Glitches typically arise from failures in real-time data communication during digital audio processing on a computer. Specifically, audio systems rely on buffers for transferring audio samples between the operating system (OS) and the digital-to-analogue converter. When the OS fails to deliver processed samples to the audio buffer within the required timeframe [23], dropouts occur, resulting in glitches. Improper configuration of audio input and output settings is a common cause of such failures.

In this study, glitches are categorised as a form of temporal noise. To simulate this degradation, dropout events are introduced by randomly setting a specified percentage of audio data packets to zero, replicating the effects of failed buffer delivery in a digital audio system. These simulated glitches mimic real-world conditions where dropouts disrupt entire audio buffer segments. The number of glitches is controlled as a parameter, representing the frequency of such events in the audio clip. The position of each glitch is chosen randomly. At a sample rate of 48 kHz and a block size of 256 samples, each glitch lasts approximately 0.05 seconds. Fig.1 illustrates a simulated glitch scenario, where the disruption is visible as a discontinuity in the waveform.

It is important to note that this study focuses solely on unintentional glitches as a form of audio degradation. Creative uses of glitch effects, such as in the "glitch music" genre[24], fall outside the scope of this discussion.

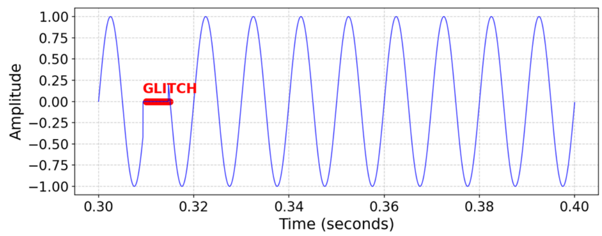

Figure 1: *Simulated glitch in a 100 Hz test tone at a 48 kHz sample rate with a 256-sample buffer segment.*

2.2. Evaluation process

This section describes the design of a system to simulate degradations in music production, along with an evaluation process applying various objective metrics. The subsequent content details the experimental preparation, including the selection of audio tracks, the choice of evaluation metrics, the software environment and the methodology employed for the evaluation.

2.2.1. Audio examples selection

Twelve audio tracks (list in Table.1) used for testing are drawn from a subset of tracks from the MUSDB18 dataset [25]. The MUSDB18 is renowned for its extensive collection of pop/rock genres, which is well-suited for our study purposes. In both tests, only the final mix product, "mixture.wav", is evaluated. To enhance computational efficiency, the first 8-second excerpts are extracted from the file for processing.

Table 1: *The Audio tracks for testing.*

$TrackTitle$	$Genre$
$MusicDelta - Beatles$	$Singer/Songwriter$
$MusicDelta - Britpop$	Pop
$MusicDelta - Country1$	$Country$
$MusicDelta - Country1$	$Country$
$MusicDelta - Country1$	$Country$
$MusicDelta - Country2$	$Country$
$MusicDelta - Disco$	Pop
$MusicDelta - Gospel$	Pop
$MusicDelta - Grunge$	$Rock$
$MusicDelta - Hendrix$	$Rock$
$MusicDelta - Punk$	$Rock$
$MusicDelta - Reggae$	$Rock$
$MusicDelta - Rock$	$Rock$
$MusicDelta - Rockabilly$	$Rock$

2.2.2. *Metric Implementation and Evaluation Setup*

Five objective metrics algorithms are utilised in this study: gst-PEAQ (both basic and advanced versions), aligned to the ITU-R BS.1387-1, as implemented in [26]; PEMO-Q developed by [7]; ViSQOL was introduced by [8]; and for the implementation of HAAQI, the study chose the implementation from [27]. Due to the differing scales of these metrics, their outputs are normalised to [0,1] before generating the comparative plots presented in the results section.

Metrics were computed offline on a Linux operating system, primarily using Python scripts, with PEMO-Q executed via MATLAB. Processing was done at 48 kHz and 16-bit precision, as required by PEAQ and ViSQOL. All metrics used floating-point arithmetic to ensure consistent accuracy. HAAQI was run in normal hearing mode without adjustments for hearing loss.

2.2.3. *Evaluation procedure and test design*

The evaluation process uses objective metrics to compare degraded audio with reference signals, analysing the impact of signal chain degradations on audio quality. To simulate cumulative effects, the audio mixture undergoes a multi-step degradation procedure and CODEC compression [28]. The lossy (MP3) versions serve as references. To reduce the influence of loudness level, both levels of degraded and reference tracks are matched to -14 LUFS, ensuring consistent levels throughout the evaluation. LUFS, a standardised loudness measurement unit defined by EBU R 128, is adopted for audio normalisation. The -14 LUFS level aligns with the loudness standards commonly used by popular streaming platforms, providing a consistent baseline for comparison. The evaluation process is demonstrated in Fig.2.

The study assumes that an ideal metric should accurately and sensitively reflect changes in audio quality as degradation levels vary (effectiveness), while remaining stable under fixed degradation conditions and across audio samples (robustness). Therefore, the evaluation process includes two individual tests.

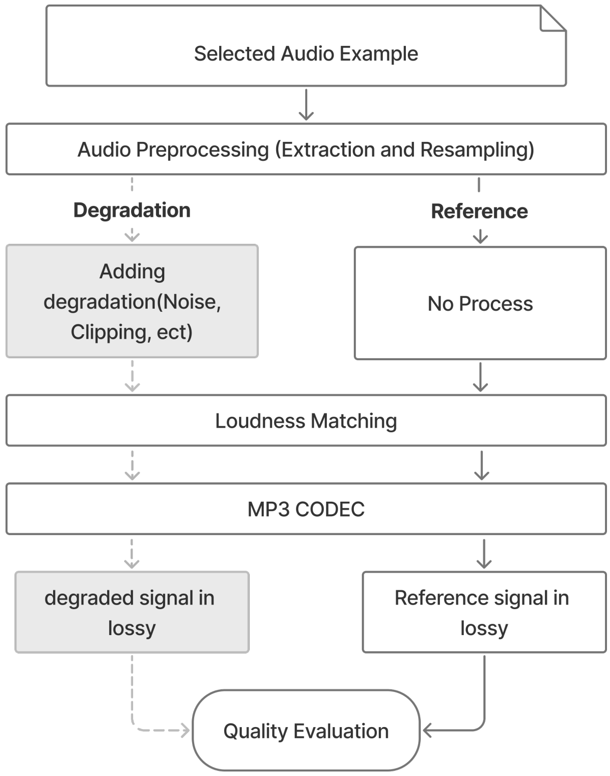

Figure 2: *Audio-quality evaluation pipeline for clean reference vs. degraded signal*

Effectiveness Test: the effectiveness test evaluates metric performance as audio degradations gradually increase. Each degradation type (hum, hiss, clipping, glitches) is applied in incremental steps: signal to hum and hiss vary from severe (1 dB SNR) to mild (80 dB SNR), clipping from 1% to 100% of samples, and glitches from 1 to 100 occurrences. Metrics are averaged across audio examples at each degradation level.

Robustness Test: the robustness test examines metric stability under fixed degradation conditions, introducing fluctuation through random factors such as white noise and glitch placement. Conditions include signal-to-hum and white noise at 60 dB, 1% clipping, and 5 glitch occurrences. Metrics are assessed over 100 iterations for each audio example to quantify consistency under the various conditions.

3. RESULTS

By comparing the metric results, this analysis identifies trends and limitations of each metric to quantify audio quality, highlighting their effectiveness and robustness under various degradation conditions.

3.1. Effectiveness test

This section explores audio-quality metrics across four degradation types: hum noise, white noise, hard clipping, and glitches. Metrics are evaluated against references to identify trends and limitations as degradations increase. The study observed similar trends across the audio examples, thus, the curves discussed in the main text represent the mean values across examples. Individual results for each audio example are provided in Appendix 6.

3.1.1. Hum noise degradation test

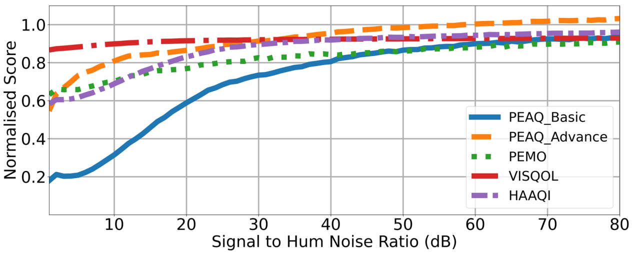

Figure 3: *Mean metric trends with rising signal to hum noise ratio, comparing degraded audio to lossy MP3 references.*

The metrics in Fig.3 rise with increasing signal-to-hum noise ratio and plateau between 30–50 dB, indicating improved quality as hum noise decreases, and eventually converge toward to 1.

ViSQOL show stable scores across all SNR levels. This early convergence indicates lower sensitivity to the hum noise. PEAQ Basic, PEMO and HAAQI show greater sensitivity at lower SNR; PEAQ Basic improves sharply beyond 20 dB, while PEMO and HAAQI increases gradually, both capturing incremental quality changes.

3.1.2. Hiss noise degradation test

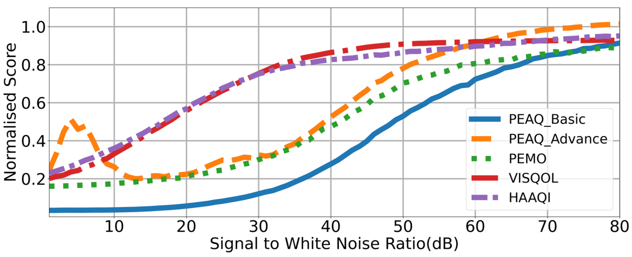

Figure 4: *Mean metric trends with rising signal to white noise ratio, comparing degraded audio to lossy MP3 references.*

All metrics in Fig.4 exhibit an upward trend as the signal-to-white noise ratio increases, indicating that low SNR conditions result in significant signal degradation. Additionally, the metrics tend to plateau at higher SNR levels above 60 dB, where the noise becomes negligible and has minimal impact on perceived audio quality.

The most critical region for evaluation occurs at middle-range SNR values(30dB to 60dB), where the metrics show notable variability in their response to moderate levels of degradation. As SNR increases, PEMO-Q and PEAQ Basic show a sharp quality improvement between 40 to 60 dB, while HAAQI and ViSQOL exhibit a more gradual rise starting around 30 dB SNR. PEAQ Advanced displays an anomalous peak at 0 to 10 dB SNR, indicating potential algorithmic limitations in low-SNR conditions.

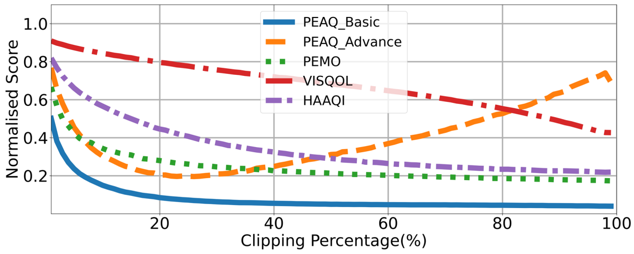

Figure 5: *Mean metric trends with rising percentage of clipping, comparing degraded audio to lossy MP3 references.*

3.1.3. Clipping degradation test

All metrics in Fig. 5 exhibit a downward trend as the clipping percentage increases, highlighting the severe distortion introduced even at levels below 5%. PEAQ Basic and PEMO-Q drop sharply and plateau around 20%, while HAAQI declines more gradually, stabilising near 50%. This stabilisation across metrics likely occurs as the signal becomes heavily distorted and unrecognisable. Notably, PEAQ Basic's early decline aligns with findings from [9], which reported its limitations in handling non-linear distortion. In contrast, ViSQOL demonstrates a more gradual decline, reflecting its moderate sensitivity to clipping distortion across all levels. However, this behaviour underscores its limited effectiveness in accurately identifying and assessing the impact of severe clipping artefacts. Supported by [11], which reports that ViSQOL performs poorly when applied to datasets containing clipping distortion. Furthermore, PEAQ Advanced exhibits an unintended trend after reaching its lowest point. This anomaly indicates its limitations when handling extreme clipping conditions as well.

3.1.4. Glitch degradation test

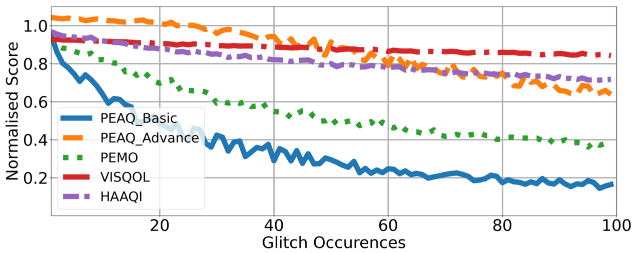

Figure 6: *Mean metric trends with rising occurrence of glitch, comparing degraded audio to lossy MP3 references.*

All metrics in Fig.6 generally decline as the number of glitch occurrences increases. This trend is expected, as an increasing number of glitches introduces severe interruptions in the audio signal, leading to a readily perceived degradation in quality. Due to the randomness of glitch occurrences, all metrics exhibit a certain degree of fluctuation as the number of glitches increases.

PEAQ Basic and PEMO-Q emerge as the most sensitive metrics to glitch degradation. PEAQ Basic shows a steep decline in quality scores within the first 10 glitches, indicating a rapid response to early interruptions. However, PEAQ Basic and PEMO-Q exhibit noticeable fluctuations throughout the plot, where scenarios with more glitches occasionally achieve higher scores than

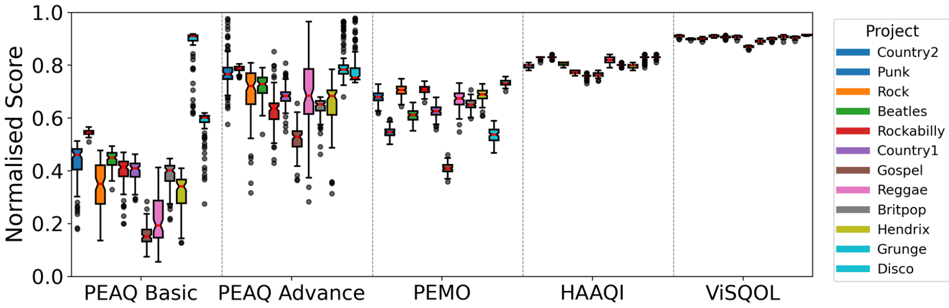

Figure 7: *The variability of audio quality metrics score under the conditions of fixed level degraded signal against the lossy MP3 reference across the audio samples.*

Table 2: Mean of each metric normalised score across each audio example.

Metric	Country2	Punk	Rock	Beatles	Rockabilly	Country1	Gospel	Reggae	Britpop	Hendrix	Grunge	Disco
PEAQ Basic	0.429	0.545	0.341	0.443	0.397	0.400	0.156	0.215	0.382	0.319	0.462	0.577
PEAQ Advance	0.773	0.786	0.695	0.719	0.629	0.683	0.520	0.694	0.639	0.650	0.763	0.783
PEMO-Q	0.679	0.547	0.704	0.609	0.707	0.625	0.410	0.671	0.651	0.687	0.517	0.733
HAAQI	0.797	0.828	0.829	0.805	0.773	0.759	0.762	0.815	0.800	0.798	0.826	0.830
ViSQOL	0.910	0.898	0.900	0.910	0.906	0.907	0.869	0.890	0.899	0.907	0.905	0.914

those with fewer glitches. The PEAQ Advanced metric demonstrates similar behaviour; additionally, while the reference is in a lossy condition, over-threshold values (scores exceeding 1.0) are observed. In contrast, the HAAQI and VISQOL Audio show a more stable decline. Whether this counterintuitive behavior reflects real-world conditions remains the subject of further work.

3.2. Robustness test

In addition to evaluating the effectiveness of the metrics, this study assesses their robustness. In this test, we define robustness as the degree of fluctuation in the metrics' outputs when exposed to controlled, but temporally randomised, noise conditions across the range of audio stimuli. Boxplots represent the distribution of scores, highlighting their spread and stability, along with the mean and average standard deviation (STD) values across all audio examples.

The setup assessed each metric's robustness under fixed degradation conditions: 60 dB SNR hum noise, 60 dB SNR white noise, 1% clipping and 5 glitch occurrences across audio examples over 100 iterations. Due to the randomness of white noise and glitches, the degradation affected the original signal slightly differently during each iteration. This randomness introduces variability in the metrics' responses, making it a critical factor for assessing their robustness.

The boxplot in Fig.7 shows score distributions over 100 iterations under fixed degradation, and per-audio example mean values. Table.2 and standard deviations Table.3. On average, PEAQ Basic (0.057) and Advanced (0.065) show the highest variability, while VISQOL (0.004) has the lowest standard deviation, indicating superior robustness, followed by HAAQI (0.007) and PEMO-Q (0.021).

The results also reveal perceptual differences across audio samples. All metrics assign lower scores to the gospel track, perhaps due to its prominent vocals, which make degradations more noticeable. ViSQOL shows stable variability across iterations but also little variation across examples, suggesting limited sensitivity to genre-specific differences.

3.3. Discussion

The effectiveness test evaluates how each metric responds to increasing degradation levels, revealing differences across degradation types. HAAQI shows a smooth, gradual trend that aligns well with linear degradation. PEAQ Basic and PEMO-Q also track changes effectively, but display more abrupt shifts, and fluctuations under glitches. ViSQOL responds to white noise and clipping but remains flat under hum and glitches. PEAQ Advanced shows inconsistent behaviour, including an unexpected peak with white noise and unusually high scores at severe clipping levels.

In the robustness experiment, metrics were evaluated under fixed but fluctuating degradations across audio examples. ViSQOL and HAAQI demonstrated the highest stability, with the lowest standard deviations across the degradation. Meanwhile, ViSQOL and HAAQI also showed minimal variation across examples, indicating limited sensitivity to genre differences.

Based on the effectiveness and robustness test findings, HAAQI appears to be the most suitable metric for further research and practical audio quality assessment. It also offers the added benefit of being capable of assessing the impacts of degradation in hearing loss conditions.

4. CONCLUSION

This study assesses the effectiveness and robustness of objective audio quality metrics, specifically in the context of digital music production for pop/rock genre examples. Unlike previous metric comparison studies [9, 10, 11, 12, 13, 14, 15], we focus on assessing the suitability of metrics and comparing their utility in this particular context. In the result, HAAQI demonstrated the highest consistency in measuring degradation, aligning with findings from [9, 13]. In contrast, PEAQ Advanced showed some incon-

Table 3: Standard deviation (STD) of each metric normalised score across each audio example.

Metric	Country2	Punk	Rock	Beatles	Rockabilly	Country1	Gospel	Reggae	Britpop	Hendrix	Grunge	Disco
PEAQ Basic	0.076	0.009	0.086	0.032	0.056	0.040	0.041	0.089	0.052	0.069	0.079	0.063
PEAQ Advance	0.068	0.012	0.099	0.045	0.067	0.039	0.054	0.144	0.038	0.090	0.063	0.065
PEMO-Q	0.023	0.018	0.023	0.023	0.015	0.023	0.020	0.035	0.021	0.024	0.021	0.011
HAAQI	0.005	0.005	0.004	0.006	0.006	0.007	0.008	0.013	0.005	0.007	0.010	0.005
ViSQOL	0.004	0.002	0.004	0.003	0.004	0.004	0.004	0.005	0.004	0.004	0.004	0.001

sistencies. Whether these issues arise from inherent limitations or specific test conditions remains an area for further investigation. Building on these findings, the next phase of our research will use HAAQI as an indicator to explore a broader range of degradation mixtures. This will help identify particularly detrimental degradation combinations and inform the development of an quality metric plugin for real-time metering in post-production applications.

5. REFERENCES

[1] Roey Izhaki, *Mixing Audio: Concepts, Practices, and Tools*, Routledge, New York, 3 edition, Nov. 2017.

[2] Michael Schoeffler and Jürgen Herre, "The relationship between basic audio quality and overall listening experience," *J Acoust Soc Am*, vol. 140, no. 3, pp. 2101, Sept. 2016.

[3] Ric Viers, *The Sound Effects Bible : How to Create and Record Hollywood Style Sound Effects*, Michael Wiese Productions, Studio City, CA, 2008.

[4] AD Wilson and BM Fazenda, "Categorisation of distortion profiles in relation to audio quality," in *17th International Conference on Digital Audio Effects (DAFx-14)*, 2014.

[5] Curtis Roads, *The Computer Music Tutorial*, The MIT Press, Cambridge, Mass, 1995.

[6] Thilo Thiede, "PEAQ–The ITU Standardfor ObjectiveMeasuremenot f Perceived Audio Quality," *J. Audio Eng. Soc.*, vol. 48, no. 1, 2000.

[7] R. Huber and B. Kollmeier, "PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006, Conference Name: IEEE Transactions on Audio, Speech, and Language Processing.

[8] Andrew Hines, Eoin Gillen, Damien Kelly, Jan Skoglund, Anil Kokaram, and Naomi Harte, "ViSQOLAudio: An objective audio quality metric for low bitrate codecs," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. EL449–EL455, May 2015.

[9] James M. Kates and Kathryn H. Arehart, "The Hearing-Aid Audio Quality Index (HAAQI)," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 24, no. 2, pp. 354–365, Feb. 2016.

[10] Samarpreet Singh, Saurabh Singh Raghuvanshi, and Vinal Patel, "Design of medium to low bitrate neural audio codec," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*. IEEE, 2023, pp. 1–5.

[11] Tashi Namgyal, Alexander Hepburn, Raul Santos-Rodriguez, Valero Laparra, and Jesus Malo, "What You Hear Is What You See: Audio Quality Metrics From Image Quality Metrics," Aug. 2023, arXiv:2305.11582 [cs, eess].

[12] Timothy Roberts and Kuldip K. Paliwal, "An objective measure of quality for time-scale modification of audio," *The Journal of the Acoustical Society of America*, vol. 149, no. 3, pp. 1843–1854, Mar. 2021.

[13] P. Kendrick, I. R. Jackson, T. J. Cox, B. M. Fazenda, and F. F. Li, "Perceived Audio Quality of Sounds Degraded by Non-linear Distortions and Single-Ended Assessment Using HASQI," *AES Journal*, vol. 63, no. 9, pp. 698–712, Sept. 2015, Publisher: Audio Engineering Society.

[14] Matteo Torcoli, Thorsten Kastner, and Jürgen Herre, "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1530–1541, 2021, Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[15] Pablo M. Delgado and Jürgen Herre, "Can We Still Use PEAQ? A Performance Analysis of the ITU Standard for the Objective Assessment of Perceived Audio Quality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2020, pp. 1–6, ISSN: 2472-7814.

[16] Andrzej Czyzewski, Andrej Ciarkowski, Andrzej Kaczmarek, Jozef Kotus, Maciej Kulesza, and Przemek Maziewski, "Dsp techniques for determining "wow" distortion," *Journal of the Audio Engineering Society*, vol. 55, no. 4, pp. 266–284, 2007.

[17] Matthias Brandt and Joerg Bitzer, "Detection of hum in audio signals," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)At: Tel Aviv, IsraelVolume: Proceedings of the 12th International Workshop on Acoustic Echo and Noise Control)*, 2010.

[18] Glen Ballou, *Handbook for Sound Engineers*, Taylor & Francis, 2008, Google-Books-ID: zsEBavFYZuEC.

[19] Joseph Coughlan-Allen, "Unknown Pleasure: interpretations of the mystery hiss in Feist's 2017 album," *Popular Music*, vol. 42, no. 3, pp. 310–327, Oct. 2023.

[20] Mead C Killion, "Noise of ears and microphones," *The Journal of the Acoustical Society of America*, vol. 59, no. 2, pp. 424–433, 1976.

[21] Timo Gerkmann, "Synthesis of Perceptually Plausible Multichannel Noise Signals Controlled by Real World Statistical Noise Properties," *Journal of The Audio Engineering Society*, Jan. 2017.

[22] Chin-Tuan Tan, B. Moore, and N. Zacharov, "The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals," *Journal of The Audio Engineering Society*, Nov. 2003.

[23] Ross Bencina, "Real-time audio programming 101: time waits for nothing," http://www.rossbencina.com/code/real-time-audio-programming-101-time-waits-for-nothing.

[24] Jan Torge Claussen, "Welcome to the glitch and make some noise: Understanding media through audio hacking," *Journal of Music, Technology & Education*, vol. 15, no. Exploring Audio and Music Technology in Education: Pedagogical, Research and Sociocultural Perspectives, pp. 7–26, Apr. 2022, Publisher: Intellect.

[25] Zafar Rafii, Antoine Liutkus, Fabian-Robert Stöter, Stylianos Ioannis Mimilakis, and Rachel Bittner, "Musdb18-hq - an uncompressed version of musdb18," Aug. 2019.

[26] Martin Holters and Udo Zölzer, "Gstpeaq–an open source implementation of the peaq algorithm," in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx), Trondheim, Norway*, 2015.

[27] "The Cadenza Project," https://cadenzaproject.github.io/.

[28] "LAME MP3 Encoder," https://lame.sourceforge.io/.

[29] Markus Erne, "Perceptual audio coders" what to listen for"," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

6. APPENDIX: EFFECTIVENESS TEST DETAILS

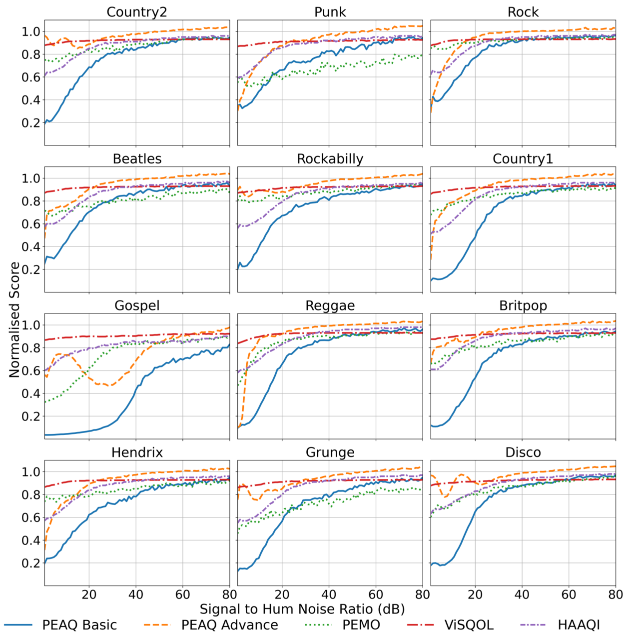

Figure 8: *Trends with rising signal-to-hum noise ratio, comparing degraded audio to lossy MP3 references.*

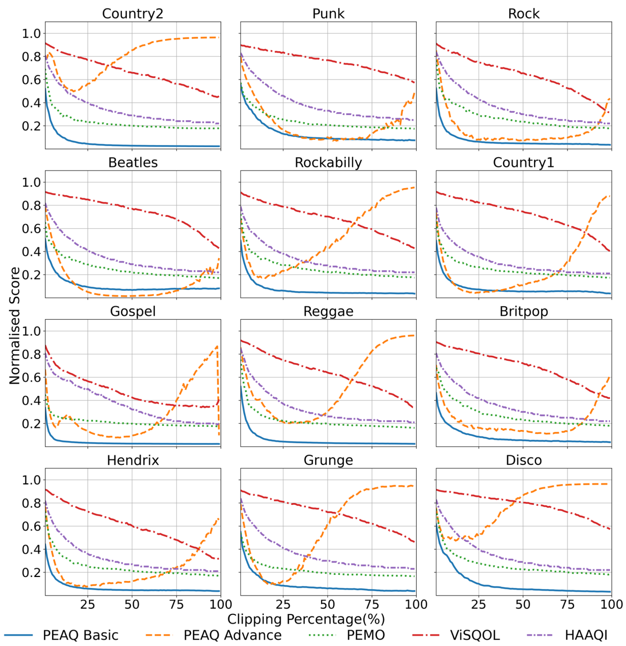

Figure 9: *Trends with rising clipping percentage of signal, comparing degraded audio to lossy MP3 references.*

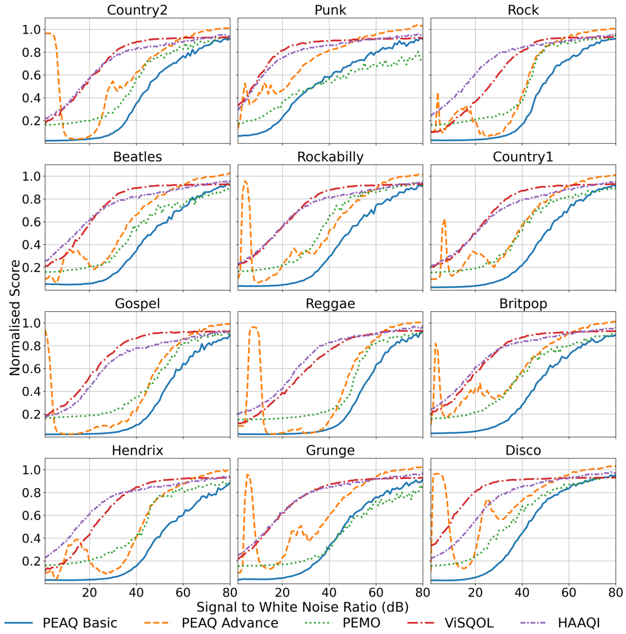

Figure 10: *Trends with rising signal-to-white noise ratio, comparing degraded audio to lossy MP3 references.*

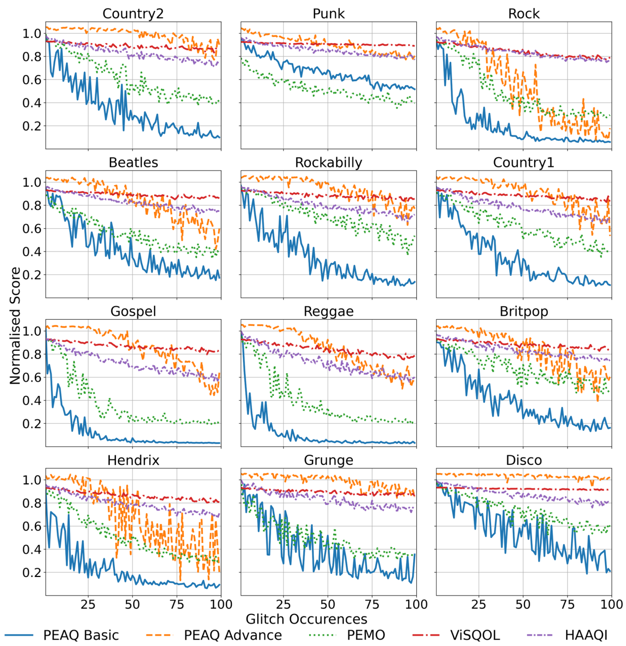

Figure 11: *Trends with rising glitch occurrences, comparing degraded audio to lossy MP3 references.*