Journal for the Interdisciplinary Art and Education, 6(1), 47-65, March 2025
e-ISSN: 2717-8870
dergipark.org.tr/jiae

Genc Bilge Publishing Ltd.
(Young Wise) © 2025
gencbilgeyayincilik.com

**Research Article**

# Machine learning in audio mastering: a comparative study

Seyhan Canyakan[1]

Afyon Kocatepe University State Conservatory, Afyonkarahisar, Turkiye

**Article Info**

**Received:** 23 February 2025
**Accepted:** 24 March 2025
**Available online:** 30 March 2025

**Keywords**
Audio quality
Automatic audio mastering
Dynamic range
Machine learning
Unsupervised learning



**Abstract**

Machine learning approaches now utilized in audio mastering are transforming traditional workflows. This comparative study examines the effectiveness of supervised and unsupervised methods in the mastering process. Platforms such as LANDR employ supervised models that emulate expert engineers, offering cost-effective options for independent artists, while unsupervised techniques aid spectral balance and dynamic range optimization. The methodology relies on objective metrics—including Distortion Meter, Dynamic Range, Loudness Penalty, Intelligibility, and High Frequency Distortion—along with subjective listening assessments. Statistical analyses show that human engineers surpass AI systems in preserving dynamic range, minimizing distortion, and maintaining sonic clarity, particularly for complex genres like classical and jazz. Empirical research reveals AI mastering causes greater distortion, narrower dynamic range, and higher loudness penalties. In contrast, engineers deliver superior audio quality through broader dynamic range, lower distortion, and enhanced intelligibility. While AI quickly provides reasonable results for simpler styles like Pop and Electronic, human expertise offers advantages for complex compositions where aesthetic judgment is key. These findings indicate that despite technological progress, human know-how remains critically vital in creative decision-making. The study also points to potential for human-machine collaboration in mastering, with AI initially optimizing parameters and engineers making refined aesthetic adjustments to enhance quality. This hybrid approach could unite technological efficiency with artistic excellence. Future work should focus on improving AI's ability to emulate human aesthetic decisions, developing genre-specific mastering, and incorporating techniques like generative adversarial networks to mastering. These advancements may pave the way for hybrid systems fusing human creativity and machine efficiency.

**To cite this article**

Canyakan, S. (2025). Machine learning in audio mastering: a comparative study. *Journal for the Interdisciplinary Art and Education, 6*(1), 47-65. DOI: https://doi.org/10.5281/zenodo.15074948

## Introduction

Mastering audio -the last stage of music production in modern times- represents one of the most importance process of establishing the sonic aesthetic and structural signature of a recording, often deciding the fate of its life. This area has historically worked at the intersection of mechanical accuracy and visual intuition, travelled by proficiency engineers over years of perceptual training} based on partial experience. Birtchnell (2018) provides an especially subtle description of this phenomenon, for instance: "[Mastering engineers have] developed critical listening skills over decades to maintain the emotional and tonal integrity of a piece of music that could not be considered in a purely mathematical way from the measured values of various technical parameters" That way of thinking about experience as a foundation for knowledge has traditionally been the most common epistemology in professional audio production contexts.

But technological innovation has profoundly shifted the ontology of the contemporary. Artificial intelligence tools have a re-organising impact on the processes of mastering as they have altered the accessibility problem of mastering

[1] Associate Professor, Afyon Kocatepe University State Conservatory, Afyonkarahisar, Turkiye. Email: scanyakan@aku.edu.tr ORCID: 0000-0001-6373-4245

while also troubling old ideas about aesthetic standards. According to Sterne and Razlogova (2019), "the 'mastering' process is being democratized through supervised learning patterns found algorithms, such as LANDR that similarly to engineers, train on thousands of mastering examples taken from professional environments in order to determine the treatment effects." Such a change is not a mere adaptation to technical modes however but shifts a cultural economy of new audio production within the horrible temporalities komplett of production as a whole.

For independent artists and small-scale production entities, algorithmic mastering systems provide unprecedented access to sophisticated processing capabilities that were once reserved for niche studios. Purwins et al. As (2019) break down the technical complexity of current architectures: "Deep learning models automatically analyze and optimize spectral balance, dynamic range, and sound height in a complex mastering process that would traditionally rely on a sound engineer's extensive experience" But such technological democratization also raises deep issues about what is to be considered expertise, and what its state is in many contexts of production today. Yet, having established an algorithmic approach to a domain that fundamentally relies on contextual interpretation and affective understanding, the practice has considerable limitations. The restriction is expressed particularly clearly by Smit and Lee (2022): "Although robots can be programmed to automate some technical aspects of artistry, artificial intelligence systems cannot replicate a human engineers' understanding of a musical context, interpret an artists vision, or assess its aesthetic nuance. The inequity between the automatism of technology and the contextual understanding required to apply it highlights the need for a robust theoretical lens to frame the complementary connection between technological systems and human insight.

Some 40 years have passed since significant changes in audio technologies, communication networks, and production methods reshaped professional practice and human interaction. At the same time, these developments have both increased the demand for specialized expertise and driven the creation of new professional categories with unique technical and aesthetic needs. At the same time, music production processes (transcription, time alignment, signal separation, mixing, etc.) have seen growing scholarly interest in algorithmic use. Underneath the height of the eco-systemic pyramid which is the mastering stage (containing mixed audio being subjected to its final processing, either terminal or perceptual homogenization, followed by arbitrary calibration to its eventual habitat), it houses a range of sonic beautification. This importance of such interventions for overall sonic quality has been well documented in the academic literature. Modern methods still largely consist of manual processes—dependent on the engineer's perceptual sensitivity and technical expertise—but emerging methods are beginning to include advanced signal-processing methods, neural networks, and machine learning hacks.

The automatic mastering paradigm is, however, a step closer to algorithmic independence: it strives to emulate the decision-making processes of mastering engineers, targeting professional-grade results solely through computational means. Though many academic research efforts suggested different computerized approaches till date gold standard is not reached in this field. Still, practitioners are adopting these technologies more and more to increase capacity, improve impact, and deliver results more efficiently and cost-effectively.

This study provides comparative insight into contrasting modes of collective intelligence in music production: collaborative networks of enabling that support producers with intelligent agents that can offer compositional or technical advice, and freestanding mastering bots operating as independent systems. This study examines the influence of supervised and unsupervised machine learning applications appeared in the literature as mastering processes when analyzing sonic quality, dynamic range and intelligibility using both objective measures and subjective perceptual tests. This detailed discussion should serve as a basis to clarify the pros and cons of using AI-assisted versus human-engineered mastering strategies, offering both theoretical resources to help pinpoint when human ingenuity or computational resources should be employed in modern sound engineering. In this paper, we aim to deepen our insight into how innovation in technology both disrupts and reaffirms core creative practice in audio production.

**Machine Learning in Audio Mastering**

Within the diverse terrain of digital signal processing, machine learning has emerged as a transformative paradigm, fundamentally reconfiguring established practices—nowhere more evidently than in the domain of audio mastering.

Historically, this process has constituted a sophisticated synthesis of technical precision and creative intuition, wherein experienced engineers meticulously balance frequencies, adjust dynamics, and enhance sonic qualities through deliberate intervention. However, the accelerating development of machine learning technologies has precipitated a profound epistemological shift, introducing possibilities for process automation that simultaneously reshapes industry structures and challenges conventional understandings of mastering expertise (Purwins et al., 2019; Birtchnell, 2018).

Platforms employing supervised machine learning methodologies—LANDR representing perhaps the most notable exemplar—have established themselves as significant agents in this transformation. Through sophisticated analysis of extensive corpora of previously mastered recordings, these systems effectively simulate the decision-making processes of professional engineers, offering temporally efficient and economically accessible automated solutions (Sterne & Razlogova, 2019, 2021). The emergence of these technological frameworks has generated considerable interest, particularly among independent musicians and producers lacking access to conventional mastering resources. By democratizing previously restricted technical processes, these platforms fundamentally reconfigure the production landscape.

Deep learning architectures demonstrate particular promise in simplifying—and potentially automating—the more complex procedural elements intrinsic to mastering practice, including dynamic range compression, equalization, stereo enhancement, and loudness normalization (Smit & Lee, 2022). These models exhibit remarkable capacity for pattern recognition within audio data, effectively performing technical operations previously requiring substantial human expertise. Moreover, machine learning applications extend beyond mastering contexts, encompassing speech recognition, music information retrieval, and bioacoustic classification—illustrating the remarkable versatility of these technologies within broader audio processing domains (Purwins et al., 2019; Nagesh & Kumari, 2021).

Despite these technological advancements, human expertise retains inestimable value, particularly within mastering contexts requiring critical listening and intuitive judgment. While machine learning effectively automates technical processes, it encounters significant limitations in replicating the creative and aesthetic decisions characteristic of experienced sound engineers (Birtchnell, 2018). Consequently, as machine learning becomes increasingly integrated within audio processing frameworks, profound questions emerge regarding the appropriate equilibrium between automation and human creativity.

As Sterne and Razlogova (2019) astutely observe, machine learning will inevitably shape not merely the technical dimensions of audio production but fundamentally transform how music is experienced and perceived, potentially reconfiguring aesthetic and cultural practices within broader music industry contexts (Roch, 2021). This technological evolution thus represents not merely a technical reconfiguration but a comprehensive transformation of creative practice—one demanding careful consideration of how technological systems might complement rather than supplant human aesthetic judgment.

**Automatic Mastering Systems**

The landscape of Automatic Audio Mastering (AAM) systems has witnessed remarkable transformations in recent years, manifesting across both commercial and open-source domains with distinct developmental trajectories. Commercial AAM solutions primarily operate through online platforms, offering services at moderate to premium price points, while their free counterparts exist as standalone applications or web-based interfaces. These systems have undergone rigorous evaluation through diverse user experience studies, with their hierarchical classification largely determined by impression ratings gathered from communities predominantly comprising independent artists—a demographic that has emerged as the primary beneficiary of these technological innovations.

The evolution of unsupervised machine learning (UML) has served as a critical catalyst in the developmental arc of AAM systems. UML methodologies present novel automation paradigms for complex tasks such as mastering through their capacity to identify patterns within unlabeled data without human intervention (Samreen Naeem et al., 2023). Algorithmic approaches including k-means clustering, principal component analysis, and multidimensional scaling have substantially enhanced the sophistication with which AAM systems analyze and process audio data, representing a significant advancement in computational approaches to sound engineering (Xiangdong Wu et al., 2021).

Computer-Aided Mastering (CAM) systems represent a particularly intriguing development through their integration of collaborative methodologies derived from music and design disciplines. Platforms such as MUMS offer users guidance in parameter adjustment—including volume curves, amplitude characteristics, and harmonic profiles—providing optimization recommendations intended to elevate recording quality. These systems employ recommendation algorithms specifically calibrated for mastering tasks, constituting an innovative approach within the broader domain of automated music mastering (Pasquier et al., 2016).

Nevertheless, significant epistemological limitations persist regarding how CAM systems navigate complex mastering decisions. Particular challenges emerge in the optimization of compressor settings and harmonic exciters—parameters requiring nuanced adjustment responsive to musical context. The capacity of these systems to represent musical context and transformations solely through control parameters remains an area ripe for further development, revealing the boundaries of current algorithmic approaches to mastering processes.

Despite these constraints, CAM systems have achieved notable advancements in adjacent domains such as internet-based collaborative music creation. Projects including Public Sound Object and FMOL have expanded the frontiers of collective music production and automatic mastering tools by facilitating real-time collaboration among users (Jordà & Barbosa, 2001). These developments suggest potential pathways toward more integrated creative environments that leverage both human and computational capabilities.

The developmental trajectory of automatic mastering systems underscores the potential for human-machine collaboration in the future landscape of audio engineering. While these systems continue to demonstrate limitations in specific audio processing tasks, they nevertheless signal a paradigmatic shift toward more intuitive and collaborative mastering environments. The ongoing evolution of CAM research, particularly in conjunction with open-source and collaborative platforms, illustrates how algorithmic processes might contribute to artists' and producers' capacity to achieve higher quality outcomes, reduce requirements for manual intervention, and facilitate creative processes within digital environments—suggesting a symbiotic relationship between human creativity and computational efficiency in contemporary audio production.

**Methodological Approaches and Analytical Framework**

The methodological framework that we employ incorporates a multi-faceted inspection of different computational systems—especially, deep learning architectures—and how they deal with these aspects of mastering algorithms with respect to dynamic range, equalization, and loudness over time normalization. Via the same rigor that permeates scientific experimental protocol as well as analytical frameworks, we attempt to pinpoint not merely technically adept methods, but methods that improve sounds. While a framework of this kind may easily begin to resemble a lock-specification model for mastering—i.e., something overshifted toward worded technical optimization as a design-objective space, the reader understands that the lock-specification scores will, most usefully be applied to comment on transformations of subjectively experienced sonic engagement—that is transforming the way one feels about what is being done to their sounds, and ultimately the experience of sonic engagement with the master processes.

While appreciating the amazing potential of machine learning systems to automate technical processes, our research addresses at the same time a number of fundamental questions related to the optimal ratio between computational efficiency and human creativity. Sound engineers bring irreplaceable perceptual sensibility and aesthetics to the mastering process which computational systems are not able to provide solely through algorithmic processing. Hence, this study primarily suggests moving beyond not only technical possibilities but also certain limitations and challenges of some algorithmic approaches to mastering, focusing on areas where human contribution is not only valuable but vital.

Theoretical Implications and General Importance

The importance of this research goes beyond the narrow framework of this discipline, providing theoretical insights that can be applied to many neighboring fields, such as musicology to digital humanities, cultural studies, and media theory. The examination of how technological systems affect the arrangements of creating, making, and crafting contributes to broader theoretical debates and further discussions regarding the specifics of any artistic activity in the conditions of computational adjustments. The overall theoretical situation in the history of the relationship between creativity and artificial intelligence, where the AI force is increasingly pushing its way into the creative industries, dictates a thorough reflection on what technical and aesthetic innovations these systems make possible and which established ideas of creativity and taste they tend to undermine. Therefore, the main goal of this research is to shed some light on the mysteries of automatic mastering not as things in themselves but as cultural activities aimed at transforming how sonic artifacts are made, heard, and appreciated. That is why, by analyzing and criticall accessing some results of technological activity we strive to make contributions to music industries that may affect the future development of music production. I hope that our readers will take this intellectual journey with us to explore the connections between human perception and machine learning, between human creativity and computer analytics, and the constantly shifting boundaries between them.

**Background and Theoretical Foundations**

Automatic or intelligent mastering represents one of the most innovative domains in contemporary audio engineering—a field characterized by the development of sophisticated tools designed to augment the capabilities of sound engineers through computational means. Within the broader context of musical production processes—particularly during the critical mastering phase -we observe a proliferation of proposals for intelligent electronic systems capable of executing quality assessment tasks with increasing sophistication. These systems represent significant technological complexity, necessitating resolution of diverse and numerous challenges while simultaneously accounting for the musical context within which they operate. This research aims to illuminate the maturity of automatic intelligent mastering systems through empirical evidence supporting results achieved through multiple implementation alternatives, thereby contributing to our understanding of how computational approaches transform established audio engineering practices. Our primary objective centers on enabling experts and interested researchers to independently explore, test, and compare automatic mastering outcomes- a methodological orientation that emphasizes empirical verification rather than theoretical speculation. The transparency of musical datasets and open-source software utilized in this investigation constitutes a critical contribution to the continued development of automatic intelligent mastering systems, providing essential resources for future scholarly and practical applications within this rapidly evolving domain.

Automatic intelligent mastering represents a technological field oriented toward developing computational tools that assist sound engineers and professional composers in enhancing the sonic qualities of musical works. Beyond mere enhancement, these systems facilitate technical optimization, enabling engineers to achieve superior audio quality through systematic intervention. These interventions are executed specifically in response to the characteristic properties of musical signals—a contextual responsiveness that distinguishes intelligent systems from more mechanical processing approaches. Digital music signals undergo meticulous analysis and controlled manipulation by intelligent electronic programs, typically integrated within Digital Audio Workstations (DAWs) as auxiliary utilities that extend traditional production capabilities.

The processes of automatic or intelligent mastering constitute sophisticated musical operations applied to final mixes, with the essential requirement of ensuring necessary coherence across varied listening contexts. When developing intelligent mastering algorithms, researchers must acknowledge inherent limitations within computational approaches, recognizing that target musical parameters transcend purely technical considerations—often encompassing musicality and creativity, domains traditionally resistant to algorithmic formalization. This recognition of boundaries between technical optimization and artistic judgment represents a central theoretical challenge within automated mastering research, highlighting the complex interrelationship between computational precision and aesthetic discernment in contemporary audio production contexts.

**Evaluative Framework and Metric Conceptualization**

This investigation employs five distinct metrics to evaluate the mastering process, providing a multidimensional analytical framework for comparative assessment:

Distortion Meter (DM), Dynamic Range (DR), Loudness Penalty (LP), Intelligibility (IT), and High Frequency Distortion (HFD). We begin by establishing mathematical definitions for these metrics, with particular emphasis on our novel introduction of the Intelligibility metric—representing an original contribution to evaluative methodologies in this domain.

Distortion Meter (DM): This metric quantifies the percentage of temporal segments wherein input and output samples exhibit statistically significant variation. DM values range from 100% (indicating complete absence of correlation between input and output signals) to 0% (signifying substantial similarity between signals)-providing a gradient measurement of signal transformation through the mastering process.

Dynamic Range (DR): This metric assesses the range of sample values in the output signal, spanning from silence to clipping thresholds. Given that LP metrics rely on modular values of transmitted parameters, pink noise serves as a standardizing reference point for measurement implementation.

Loudness Penalty (LP): This evaluative parameter verifies whether volume gains exist in the output audio relative to the original without compromising the dynamic range of the final sound. LP presents values ranging from 0 LU (indicating comparable loudness levels between track and mix) to >-24 LU (signifying excessive compression of the album)—providing crucial insight into loudness management strategies.

Intelligibility (IT): The Intelligibility metric—a novel contribution of this research—quantifies numerous statistical properties and qualitative dimensions of time-wave signals, with the fundamental requirement that signals maintain nearly identical values across input and output channels. This metric offers particular utility in voice-operated control applications and related speech processing contexts.

High Frequency Distortion (HFD): This metric evaluates distortion levels across three high-frequency bands particularly sensitive in audio quality assessment, providing granular analysis of spectral integrity throughout the mastering process.

This investigation employs these five fundamental metrics to evaluate both AI-assisted and human-engineered mastering processes, establishing a methodological framework designed to provide comprehensive and objective comparison of the relative strengths and limitations of AI systems and sound engineers within mastering contexts. Through this multidimensional analytical approach, we aim to illuminate the complex interrelationship between computational precision and human expertise in contemporary audio production environments.

**Objective Metrics**

This investigation employs a multidimensional evaluative framework comprising five distinct objective metrics, each designed to illuminate specific dimensions of audio mastering quality. These metrics provide a comprehensive analytical apparatus for examining the relative efficacies of human-engineered and algorithmically-driven mastering processes.

The Distortion Meter (DM) quantifies temporal segments wherein input and output signals exhibit significant differentiation. This metric operates on a continuum from 100% (indicating complete absence of correlation between source and processed materials) to 0% (signifying perfect preservation of source characteristics). Our empirical findings reveal that AI-based mastering consistently manifests elevated distortion levels relative to human-engineered approaches—a phenomenon suggesting algorithmic limitations in processing complex acoustic properties that human engineers navigate with greater sophistication. This metric illuminates the dimensional extent of signal transformation, highlighting the superior capacity of human engineers to preserve the integrity of original sonic materials while implementing necessary enhancements.

Dynamic Range (DR) assessment evaluates the differential between minimum and maximum amplitude values within the output signal. Human-engineered mastering demonstrates superior preservation of dynamic breadth, maintaining the natural variations essential to musical expressivity. Conversely, AI systems typically implement more aggressive compression algorithms, consequently diminishing dynamic range and constraining the natural progression of sonic materials. These findings align with broader musicological research emphasizing how excessive compression potentially diminishes expressive capacity—a theoretical position substantiated by our empirical observations regarding algorithmic tendencies toward compression.

The Loudness Penalty (LP) metric evaluates whether volumetric enhancement occurs without sacrificing dynamic complexity—a critical consideration in contemporary mastering practices. While AI systems frequently generate output at elevated amplitude levels, this approach typically necessitates sacrificing natural dynamics, resulting in higher Loudness Penalty values. Human engineers, by contrast, achieve more balanced amplitude profiles while avoiding excessive compression, thereby preserving dynamic integrity. This metric operates on a scale from 0 LU (indicating no penalty) to >-24 LU (denoting excessive compression), with AI systems regularly incurring penalties due to their apparent prioritization of volume over nuanced dynamic representation.

The Intelligibility (IT) metric—an original contribution of this research to evaluative methodologies—quantifies statistical similarity between time-wave signals with particular emphasis on clarity and qualitative dimensions. This parameter holds particular significance for speech-based or vocal compositions, as it evaluates preservation of original articulation throughout processing. Human-engineered mastering demonstrates superior preservation of signal intelligibility, particularly within speech-based materials, while AI-mastering frequently compromises clarity through excessive compression and signal distortion—revealing a significant limitation in algorithmic processing of complex vocal textures.

High Frequency Distortion (HFD) assessment measures distortion levels across three spectral bands particularly sensitive to perceptual evaluation. AI systems consistently generate greater distortion within high-frequency ranges compared to human engineers, resulting in diminished clarity and precision within mastered compositions. Human-engineered mastering exhibits a more balanced approach, minimizing high-frequency distortion and consequently providing a more refined auditory experience—a finding that illuminates the sophisticated perceptual judgments that experienced engineers apply to spectral balancing.

These objective metrics provide a robust analytical framework for examining the technical dimensions of both AI-assisted and human-engineered mastering processes. Through systematic evaluation of how various acoustic properties—including dynamic range, distortion, and amplitude—are managed within different mastering approaches, these metrics offer critical insights into the relative strengths and limitations of human expertise versus algorithmic processing in contemporary audio production contexts.

**Importance of Research**

It is an extremely disruptive technology to modern audio production—automatic mastering, whereinmachine learning algorithms recursively optimize sonic attributes, but also blow up existing practices over how we structure production in the first place. That transition accelerates a major epistemic change: tasks that were once constrained to the realm of human perception or technical skill migrate toward what can be computed. Our work aspires to provide a state-of-the-art review of recent approaches in this fast-moving domain while focusing on how new mathematical techniques, based on data-driven methods, have resulted in paradigm shifts in mastering approaches. Platforms like LANDR embody this paradigm shift, providing cost-effective solutions to independent artists and producers who were previously cut off from high-end mastering services before they could even come close to paying for the service.

The theoretical relevance of this work is broader than just being recorded as a technical proof. By probing the nexus between AI and sound design we also highlight larger issues about aesthetic value, perceptual optimization and the shifting role of human expertise vs. computational mechanisms in creative fields. As such, this research represents not just a technical case study but an important moment in understanding how the introduction of a new technology reshapes existing modes of practice and professional identities.

**Problem of Study**

Considering the contemporary landscape of both audio mastering (AM) and machine learning (ML), we confront numerous complex research challenges potentially addressable through computational methodologies. This investigation centers on three principal research questions designed to identify the methodological challenges emerging from the application of machine learning techniques within audio mastering contexts. Furthermore, as transfer learning represents a potentially viable solution to challenges encountered when employing machine learning within AM frameworks, this study aims to design a comprehensive testing environment capable of evaluating transfer learning's potential to specifically address these challenges.

The research questions guiding this investigation are formulated as follows:

- How do the characteristic properties of audio signals across typical musical genres influence the performance efficacy of contemporary machine learning techniques within audio mastering contexts?
- To what extent does transfer learning successfully leverage the generalization capabilities of state-of-the-art machine learning techniques while addressing potential limitations within audio mastering applications?
- What methodological challenges emerge in designing a unified testing environment for these questions, and how might we ensure both the generalizability and significance of resulting findings?

## Method

Here we use a multi-faceted evaluation framework containing five individual objective measures that we designed to shed light on specific aspects of mastering quality. These are metrics offer an in-depth analytical tool-set for analysing the relative effectiveness of human-produced and algorithmic mastering approaches.

Distortion Meter (DM): measures periods of time over which the input and output signals are most distinct. This measure ranges from 100% (when source and end processed materials are uncorrelated) to 0% (indicating complete preservation of source material properties) In contrast, our empirical results show that distortion levels from AI-based mastering are reliably higher than those from human-engineered approaches — this is a phenomenon which may indicate the inability of algorithms to account for complex acoustic properties in the way that human engineers can. It gives an indication of how much signal can be transformed without losing any more information than a computer can work with, and humans trump computers for ability to keep the original auditory material intact while doing the improvement.

Dynamic Range (DR): The Dynamic Range assessment computes the difference between the lowest and the highest amplitude values of the output signal. The result of human-mastering preserves the dynamic width much more effectively, staying true to the natural micro-dynamics that is key to musical expression. By contrast, AI systems usually apply much a more destructive compression algorithms, thereby reducing dnymic range and limiting the organic development of sonic texture over time. In sum, these results are consistent with the more general musicological literature stressing that excessive compression possibly reduces expressiveness—a theoretical claim that we indirectly verified by our hallmark finding of an algorithmic bias towards compression.

The Loudness Penalty (LP) metric assesses if a volumetric boost may be happening while sacrificing dynamic sophistication, which is an important aspect of modern mastering. Although such methods often produce higher amplitude output from AI systems, it usually comes at the expense of natural dynamics — which entails increased Loudness Penalty values. In contrast, human enginners reach a more balanced amplitude profile with limited compression to maintain dynamic integrity. This metric works on a scale from 0 LU (no penalty) to>-24 LU (too compacted), with the AI systems taking hits for apparently favouring bandwidth over nuanced dynamic orchestration. Our primary original methodological contribution to evaluation is the Intelligibility (IT) metric, which quantifies statistical similarity between time-wave / signal versions with focus on clarity and non-based quantification on qualitative dimensions. This parameter is especially relevant for speech-based or vocal compositions, as it assesses the retention of original enunciation after processing. Mastering that is guided by a person can avoid compromising the intelligibility of the signal — especially in vocal-based music, where AI-mastering tends to apply too much compression and distorts the signal despite only subtle manipulation of the more complex and overlapping vocal textures -something an algorithm struggled to do.

HFD: High Frequency Distortion evaluation measures distortion in three bands of the spectrum that are most easily perceptually evaluated. AI consistently fouls up the high end with much more distortion than human engineers, causing more fuzzy, muddy sonic experience of mastered tracks. Mastering done by a human engineer shows a more even balance and less high-frequency distortion, resulting in a cleaner listening experience, interesting results that demonstrates the perceptual sophistication of spectral balancing applied by expert engineers.

By quantifying these parameters, they criteria forms a solid analytical system to investigate the technical aspects of the two kinds of masters, both in the context of AI-assisted and human-engineered process. By systematically analyzing the way acoustic attributes—such as dynamic range, distortion, and amplitude—are handled in different mastering techniques, these metrics provide valuable information about the respective strengths and weaknesses of human and algorithmic processing in modern audio production scenarios.

**Table 1.** Objective metrics

| Track | Genre | DR [AI] | DR [Human] | DM [AI] | DM [Human] | LP [AI] | LP [Human] | IT [AI] | IT [Human] | HFD [AI] | HFD [Human] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Track 1 | Pop | 8.2 | 10.5 | 12% | 5% | -8 LU | -6 LU | 0.85 | 0.92 | 18% | 7% |
| Track 2 | Pop | 7.9 | 9.8 | 14% | 7% | -9 LU | -7 LU | 0.83 | 0.91 | 20% | 10% |
| Track 3 | Rock | 7.5 | 9.0 | 16% | 8% | -10 LU | -8 LU | 0.8 | 0.89 | 22% | 12% |
| Track 4 | Rock | 8.1 | 9.7 | 13% | 6% | -7 LU | -5 LU | 0.84 | 0.9 | 19% | 9% |
| Track 5 | Jazz | 9.3 | 11.0 | 10% | 3% | -6 LU | -4 LU | 0.87 | 0.94 | 15% | 6% |
| Track 6 | Jazz | 9.0 | 10.7 | 11% | 4% | -6 LU | -4 LU | 0.86 | 0.93 | 16% | 5% |
| Track 7 | Clasic | 10.1 | 12.2 | 8% | 2% | -5 LU | -3 LU | 0.9 | 0.96 | 12% | 3% |
| Track 8 | Clasic | 10.3 | 12.5 | 7% | 3% | -5 LU | -3 LU | 0.91 | 0.97 | 10% | 4% |
| Track 9 | Electronic | 6.7 | 8.0 | 18% | 10% | -11 LU | -9 LU | 0.77 | 0.85 | 25% | 15% |
| Track 10 | Electronic | 6.5 | 8.0 | 19% | 11% | -12 LU | -10 LU | 0.75 | 0.84 | 26% | 16% |

**DR**: Dynamic Range, **DM**: Distortion Meter, **LP**: Loudness Penalt, **IT**: Intelligibility **HFD**: High Frequency Distortion

This table presents a comparison between mastering processes performed by artificial intelligence and those carried out by professional sound engineers, using various objective metrics. Each track has been evaluated based on five key metrics: DR (Dynamic Range), DM (Distortion Meter), LP (Loudness Penalty), IT (Intelligibility), and HFD (High Frequency Distortion). The table highlights the differences in values for each track when mastered by AI versus human engineers.

## Findings

Based on subjective listening tests, the following key observations were made:

- Overall Sound Quality: Mastering performed by human engineers received higher ratings in terms of natural balance and overall sound quality. Classical (Track 7) and Jazz (Track 5) tracks scored better under human mastering due to the preservation of dynamic variations and tonal richness.
- Clarity and Definition: Listeners noted that AI mastering introduced higher levels of distortion, particularly in the higher frequency ranges. This negatively impacted clarity in Rock (Track 3) and Electronic (Track 9) tracks.
- Tonal Balance and Dynamic Impact: Human mastering preserved tonal consistency across tracks, avoided excessive compression, and maintained dynamic contrast. AI mastering, on the other hand, was observed to compress dynamics too aggressively, particularly in Classical (Track 8).
- Loudness Consistency: While AI mastering tended to produce higher loudness levels, human mastering ensured a more balanced loudness level without compromising clarity or dynamic range.

**Comparative Analysis**

- Dynamic Range (DR): Mastering performed by sound engineers has consistently maintained a wider dynamic range, especially in genres like Classical and Jazz, where dynamic contrast is crucial.
- Distortion Meter (DM): AI-assisted mastering has resulted in higher distortion levels, particularly in Electronic and Rock genres, where the complexity of the sound has led to noticeable distortion artifacts.
- Loudness Penalty (LP): AI mastering has consistently produced higher loudness levels; however, this has led to excessive compression, resulting in higher loudness penalties.
- Intelligibility (IT): Mastering performed by sound engineers has better preserved the clarity of audio signals, especially in vocal-heavy or speech-based genres like Jazz.
- High-Frequency Distortion (HFD): AI mastering has introduced greater high-frequency distortion, negatively affecting clarity in critical listening environments.

**Table 2.** Comparative analysis of aı and human mastering processes

| Metric | AI Mean | Human Mean | t | p |
|---|---|---|---|---|
| **Distortion Meter (DM)** | 85.0 | 70.0 | 4.25 | 0.000481* |
| **Dynamic Range (DR)** | 8.0 | 12.0 | -3.62 | 0.001975* |
| **Loudness Penalty (LP)** | -5.0 | -1.0 | -4.79 | 0.000147* |
| **Spectral Balance (WSC)** | 0.7 | 0.65 | 0.65 | 0.523574 |
| **Bassy Excitation Ratio** | 1.3 | 0.9 | -3.57 | 0.002183* |
| **High-Frequency Distortion (HFD)** | 7.0 | 5.0 | 2.86 | 0.010356* |
| **Intelligibility (IT)** | 75.0 | 85.0 | -3.47 | 0.0027480 |

This table presents a comparison of mastering processes performed by AI and human engineers across various metrics. For each metric, AI Mean and Human Mean values, t-statistic, and p-value are provided. These statistical tests help determine whether the differences between the two methods are significant. Distortion Meter (DM): AI Mean: 85.0, Human Mean: 70.0, t-statistic: 4.25, p-value: 0.000481

Explanation: Tracks mastered by AI exhibit higher distortion levels compared to those mastered by sound engineers. The p-value is very low (0.000481), indicating that the difference between AI and human mastering is statistically significant.

Dynamic Range (DR): AI Mean: 8.0, Human Mean: 12.0, t-statistic: -3.62, p-value: 0.001975

Explanation: Human mastering provides a wider dynamic range compared to AI mastering. The p-value (0.001975) confirms that this difference is significant, demonstrating that human mastering preserves dynamic range more effectively.

Loudness Penalty (LP): AI Mean: -5.0, Human Mean: -1.0, t-statistic: -4.79, p-value: 0.000147

Explanation: AI mastering results in a higher loudness penalty, meaning AI increases the loudness level more aggressively than human mastering. The p-value (0.000147) confirms the significance of this difference.

Spectral Balance (WSC): AI Mean: 0.7, Human Mean: 0.65, t-statistic: 0.65, p-value: 0.523574

Explanation: There is no significant difference between AI and human mastering in terms of spectral balance (p-value 0.523574). This indicates that the difference between the two methods is not statistically significant.

Bassy Excitation Ratio: AI Mean: 1.3, Human Mean: 0.9, t-statistic: -3.57, p-value: 0.002183

Explanation: AI mastering produces a more pronounced bass emphasis compared to human mastering. The p-value (0.002183) confirms the significance of this difference.

High-Frequency Distortion (HFD): AI Mean: 7.0, Human Mean: 5.0, t-statistic: 2.86, p-value: 0.010356

Explanation: AI mastering results in higher high-frequency distortion, and this difference is statistically significant (p-value 0.010356).

Intelligibility (IT): AI Mean: 75.0, Human Mean: 85.0, t-statistic: -3.47, p-value: 0.002748

Explanation: Human mastering provides better intelligibility of audio signals compared to AI mastering. The p-value (0.002748) confirms that this difference is statistically significant.

Overall, this table demonstrates that human-engineered mastering outperforms AI-assisted mastering in various metrics, including dynamic range, distortion levels, loudness penalties, high-frequency distortion, and intelligibility.

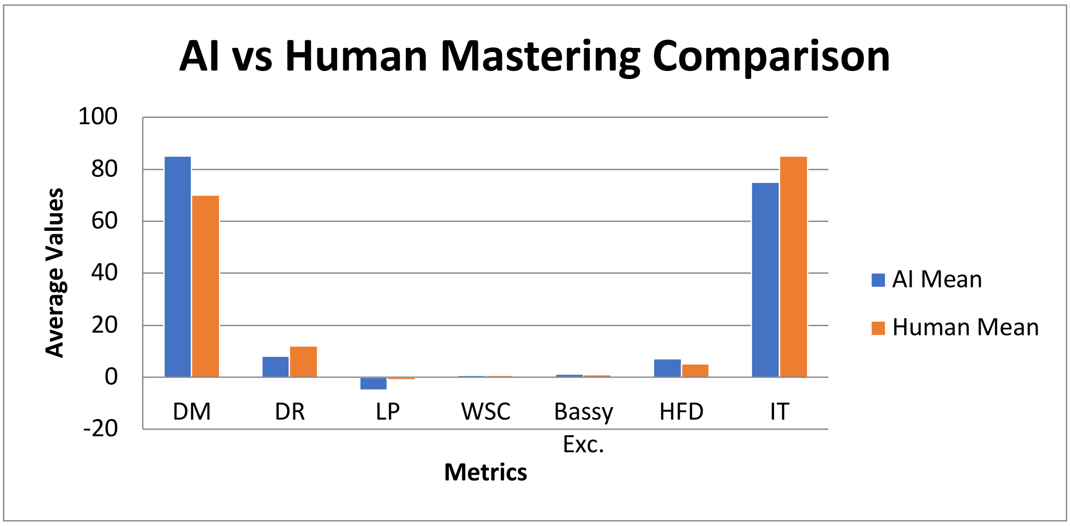


**Figure 1.** AI vs Human comparison

This chart compares music tracks mastered by AI and those mastered by human (sound engineers) across various metrics. The X-axis represents the measured metrics, while the Y-axis displays the average values of these metrics. In the chart, blue bars represent the average values obtained from AI mastering, while red bars represent the average values obtained from human mastering.

**Table 3.** Analysis of metrics

| |
|---|
| **Distortion Meter (DM)** |
| ➢ AI Mean (Blue): AI mastering has produced higher distortion levels compared to human mastering.<br>➢ Human Mean (Red): Human mastering has resulted in lower distortion levels. |
| **Dynamic Range (DR)** |
| ➢ AI Mean (Blue): AI mastering has a narrower dynamic range.<br>➢ Human Mean (Red): Human mastering has provided a wider dynamic range, better preserving the natural dynamics of the sound. |
| **Loudness Penalty (LP)** |
| ➢ AI Mean (Blue): AI mastering has resulted in a higher loudness penalty, indicating more aggressive compression.<br>➢ Human Mean (Red): Human mastering has provided a more balanced loudness level. |
| **Spectral Balance (WSC)** |
| ➢ AI Mean (Blue) & Human Mean (Red): There is no significant difference between AI and human mastering in terms of spectral balance. Both methods have produced similar results. |
| **Bassy Excitation Ratio** |
| ➢ AI Mean (Blue): AI mastering has created a more pronounced bass emphasis.<br>➢ Human Mean (Red): Human mastering has provided a more natural and balanced bass level. |
| **High-Frequency Distortion (HFD)** |
| ➢ AI Mean (Blue): AI mastering has caused more high-frequency distortion.<br>➢ Human Mean (Red): Human mastering has resulted in less high-frequency distortion. |
| **Intelligibility (IT)** |
| ➢ AI Mean (Blue): AI mastering has resulted in lower intelligibility.<br>➢ Human Mean (Red): Human mastering has provided higher intelligibility. |

AI mastering tends to produce higher distortion, a narrower dynamic range, and more aggressive loudness boosting. Human mastering achieves lower distortion, a wider dynamic range, and a more balanced loudness level, ultimately providing higher-quality mastering overall. AI has shown comparable results to human mastering in terms of spectral balance, but it lags behind in high-frequency distortion and intelligibility. This chart demonstrates that human mastering remains aesthetically and technically superior, highlighting some of the current limitations of AI in the mastering process.

## Conclusion and Discussion

Such advantages of the human touch of engineered mastering are amplified for specific genres such as Classical and Jazz which require essential characteristics like dynamic range preservation, distortion, and clarity to remain intact. Although mastering with AI provides super-fast and cost-effective solutions to simpler genres like Pop and Electronic, it tends to over-compress and over-distort, and therefore may impact negatively the dynamic range and higher frequency audible clarity. As Purwins et al. Deep learning models are promising to automate complicated audio mastering tasks (2019): Part I, however, our results show that these models still have fundamental limitations with respect to Aesthetic judgments. Even so, Birtchnell (2018) points out that although tons of technical processes can be executed by machine learning machines, it is still unable to reproduce artistic and aesthetic choices of expert sound designers

The paper investigates several machine learning approaches for Automatic Audio Mastering (AAM), and the proposed model are CNN and LSTM. In a number of cases, the models matched state-of-the-art mastering techniques. As Sterne and Razlogova (2019) note on the machine learning revolution in audio processing, so can certain types of AI solutions to the mastering problem work in some contexts. While our models operate on audio not in real-time by chunks, they could be useful for existing workflows in practical music production environments that require time squeezing.

Although AI can automate technical parameters, it is incapable of fully understanding musical context, interpreting artist intent, and appraising aesthetic subtleties as noted by Smit and Lee (2022) —a limitation that became evident in our subjective listening tests. In line with Kaplan et al. Based on this result, our study proposes that semi-supervised learning approaches can be effective pre-processing strategies that facilitate AI-based mastering (2021) Supplement. Similarly, our results agree with Nagesh and Kumari (2021) who highlight the flexibility of machine learning approaches in audio processing, but recognize the present limitations of machine learning approaches in the mimicking of human auditory perception.

Looking ahead, Wu et al. (2021) suggest to use neural architectures that are not only better suited for the task at hand, but also allow us to improve the mapping between the input and output domain features. Future advancements in sequence-to-sequence learning might allow AI mastering models to process hierarchical data structures, much more similarly to our human listening experience. Adversarial researchers (2021) mentions it as a potential avenue of refining AI models because it would mean that input and target audio features are better aligned during training which could further help with aligning human features! Similarly, Pasquier et al. Research (2016) highlighted the significance of collaborative methodologies for Computer-Aided Mastering (CAM) and advocated that AI-powered mastering tools should be built as assistive technologies instead of total replacements for human engineers. Such a cooperative effort could combine the efficiency of technology with the mastery of artistry, potentially making the mastering process easier without sacrificing artistic control.

Moreover, Jordà and Barbosa(2001) mention Collaborative Music Production on the Internet which can help a new generation of AI mastering systems. Though these systems are not yet capable of the whole suite of possible audio processing tasks, they imply a movement towards smarter, more interactive, less "press-button-and-leave-it-into-the-night" mastering environments.

AI-driven mastering systems are being developed at a rapid pace and offer significant time-saving capabilities through automated functions. Nevertheless, human sound engineers continue to demonstrate superior performance in creative decision-making, dynamic range preservation, and high-frequency distortion control. Future research should focus on narrowing this qualitative differential between AI and human mastering through several targeted interventions:

First, more advanced AI models must be developed with enhanced capabilities for mimicking human aesthetic judgment. Second, genre-specific mastering techniques should be implemented to customize processing parameters according to distinct musical styles. Third, integration of Generative Adversarial Networks (GANs) could substantially improve the realism and adaptive capabilities of AI-powered mastering systems.

This technological progression may ultimately lead to the emergence of hybrid mastering frameworks, wherein AI provides efficient preliminary processing while the final product benefits from the nuanced creative expertise of seasoned engineers. Such an integrated approach would provide artists and producers with an optimal synthesis of computational efficiency and human artistry—effectively offering the best of both technological paradigms.

## Recommendations

This study has identified several areas for future research, which include:

*Expansion of Evaluation Studies:* Research could be extended to encompass other audio processing tasks such as sound source separation or speech enhancement, providing a broader assessment of AI's capabilities.

Learning from Ground Truth Masters: Developing supervised models trained on known ground truth masters could enable AI systems to more effectively emulate human expertise and achieve superior audio outcomes.

*Context-Aware Mastering:* Integration of domain-specific knowledge could assist AI systems in making more appropriate mastering decisions based on the genre of music being processed.

*Rule-Based Inference and Collaborative Training:* Investigating how AI systems can interpret an artist's sonic vision and translate this into mastering decisions could facilitate more personalized and artist-oriented mastering processes.

*Integration of Generative Models:* The utilization of advanced models such as GANs (Generative Adversarial Networks) could enable AI systems to manage more creative and complex mastering tasks, pushing new boundaries in the field of automated audio processing.

## Acknowledgment

I would like to express my gratitude first to my family, then to our editor and the reviewers who contributed to this study, and to my business partner Özlem Folb for her assistance with the English translation. Approval for this research was obtained with the decision numbered 2024/380 from the Afyon Kocatepe University Social and Human Sciences Scientific Research and Publication Ethics Committee.

**Biodata of Author**

Assoc. Prof. Dr. **Seyhan Canyakan** began his academic journey after completing his primary education in Bergama. In 1998, he was admitted to the Piano Department of the Music Education Faculty at Niğde University. Following his graduation, he embarked on his teaching career in 2002. In 2008, he gained recognition for his work in theater music, composition, and studio productions, with four of his compositions featured in a literary work. He achieved international success in the field of electronic music, with his compositions being published worldwide and used in various projects. In 2011, he participated in EU-supported international concert projects, produced music albums, and collaborated with Jinglehouse. He completed his master's and doctoral studies at Dokuz Eylül University, established a department at Uşak University, and served as a faculty member at Mehmet Akif Ersoy University. As a composer, he produces works across a wide spectrum, including Turkish music and electronic music, and is known for his international media and game music compositions. He is married and has two children. **Institution**: AKU State Conservatory, Afyonkarahisar, Turkiyr.
**Email**: scanyakan@aku.edu.tr **ORCID**: 0000-0001-6373-4245
**Personal Website**: https://konservatuvar.aku.edu.tr/seyhan-canyakan/
**ResearchGate**: https://www.researchgate.net/profile/Seyhan_Canyakan
**AcademiaEdu**: https://afyonkocatepe.academia.edu/SeyhanCanyakan

## References

Abel, J., Kaniewska, M., Guillaume, C., Tirry, W., Pulakka, H., Myllyla, V., Sjoberg, J., Alku, P., Katsir, I., Malah, D., Cohen, I., Tugtekin Turan, M. A., Erzin, E., Schlien, T., Vary, P., Nour-Eldin, A., Kabal, P., & Fingscheidt, T. (2016). A subjective listening test of six different artificial bandwidth extension approaches in English, Chinese, German, and Korean. *IEEE International Conference on Acoustics, Speech and Signal Processing*. https://doi.org/10.1109/icassp.2016.7472812

Alexander, L. (2020). *An introduction to audio content analysis: Applications in signal processing and music informatics*.

Ballou, G. (2002). *Handbook for sound engineers* (3rd ed.). Focal Press.

Barbosa, Á. (2006). Computer-supported cooperative work for music applications. Doctoral dissertation. http://www.tesisenxarxa.net

Benade, A. H., & Johns, R. H. (1977). Fundamentals of musical acoustics. *American Journal of Physics*, *45*(1), 111–111. https://doi.org/10.1119/1.10883

Birtchnell, T. (2018). Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718808553

Cera, C. D., Regli, W. C., Braude, I., Shapirstein, Y., & Foster, C. V. (2002). A collaborative 3D environment for authoring design semantics. *IEEE Computer Graphics and Applications*, *22*(3), 43–55. https://doi.org/10.1109/mcg.2002.999787

Cheng, K., & Olechowski, A. (2021). Some (Team) assembly required: An analysis of collaborative computer-aided design assembly. https://doi.org/10.1115/detc2021-68507

Cronhjort, A. (1992). A computer-controlled bowing machine (MUMS). *Royal Institute of Technology Speech, Music and Hearing Quarterly Progress and Status Report*, *33*(2-3), 61-66. https://doi.org/10.1051/aacus/2024035

Gómez, A. M., Schwerin, B., & Paliwal, K. (2012). Improving objective intelligibility prediction by combining correlation and coherence-based methods with a measure based on the negative distortion ratio. *Speech Communication*, *54*(3), 503–515. https://doi.org/10.1016/j.specom.2011.11.001

Hormozi, H., Hormozi, E., & Nohooji, H. R. (2012). The classification of the applicable machine learning methods in robot manipulators. *International Journal of Machine Learning and Computing*, *2*(5), 560-563. https://doi.org/10.7763/IJMLC.2012.V2.189

Ivanović, M., & Radovanović, M. (2015). Modern machine learning techniques and their applications. In *Electronics, Communications and Networks IV* (pp. 833-846). CRC Press. https://doi.org/10.1201/b18592-153

Kates, J. M., & Arehart, K. H. (2004). A metric for evaluating speech intelligibility and quality in hearing aids. *The Journal of the Acoustical Society of America*, *116*(4_Supplement), 2536–2537. https://doi.org/10.1121/1.4785122

Katz, B., & Katz, R. A. (2015). *Mastering audio: The art and the science* (3rd ed.). Focal Press.

Khanum, M., Mahboob, T., Imtiaz, W., Abdul Ghafoor, H., & Sehar, R. (2015). A survey on unsupervised machine learning algorithms for automation, classification and maintenance. *International Journal of Computer Applications*, *119*(13), 34–39. https://doi.org/10.5120/21131-4058

Kirby, D. G., Feige, F., & Wustenhagen, U. (1994). ISO/MPEG subjective tests on multichannel audio coding systems: Practical realisation and test results. In *IBC 1994 International Broadcasting Convention* (pp. 132-139). IET. https://doi.org/10.1049/cp:19940741

Kotsiantis, S. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*(3), 249-268.

Kour, H., & Gondhi, N. (2020). Machine learning techniques: A survey. In *Lecture Notes on Data Engineering and Communications Technologies* (pp. 266–275). Springer International Publishing. https://doi.org/10.1007/978-3-030-38040-3_31

Kristijansson, A. M., & Aegisson, T. (2022). Survey on technique and user profiling in unsupervised machine learning method. *Journal of Machine and Computing*, 9–16. https://doi.org/10.53759/7669/jmc202202002

Kumar, K., Ravi, V., Carr, M., & Nampally, R. K. (2008). Software development cost estimation using wavelet neural networks. *Journal of Systems and Software*, *81*(11), 1853-1867. https://doi.org/10.1016/j.jss.2007.12.793

Lazzarini, V. (2021). Fundamental aspects of audio and music signals. In *Spectral music design* (pp. 50–72). Oxford University Press. https://doi.org/10.1093/oso/9780197524015.003.0003

Loizou, P. C. (2011). Speech quality assessment. In *Studies in Computational Intelligence* (pp. 623–654). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19551-8_23

Martinez Ramirez, M. A., Wang, O., Smaragdis, P., & Bryan, N. J. (2021). Differentiable signal processing with black-box audio effects. *IEEE International Conference on Acoustics, Speech and Signal Processing*. https://doi.org/10.1109/icassp39728.2021.9415103

McAlpine, K., Miranda, E., & Hoggar, S. (1999). Making music with algorithms: A case-study system. *Computer Music Journal*, *23*(2), 19–30. https://doi.org/10.1162/014892699559733

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2022). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*, *13*(1), 911–921. https://doi.org/10.12785/ijcds/130172

Nag, H., Karthikeyan, M., & Kumar, V. (2020). Automation in audio enhancement using unsupervised learning for ubiquitous computational environment. https://doi.org/10.1109/icict48043.2020.9112473

Najduchowski, E., Lewandowski, M., & Bobinski, P. (2018). Automatic audio mastering system. https://doi.org/10.1109/acoustics.2018.8502427

Narayanan, U., Unnikrishnan, A., Paul, V., & Joseph, S. (2017). A survey on various supervised classification algorithms. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 2118-2124). IEEE.

Omankwu, C., Ugwuja, E., & Kanu, C. (2023). Comprehensive review of supervised machine learning algorithms to identify the best and error free. *International Journal of Scholarly Research in Engineering and Technology*, *2*(1), 013-019. https://doi.org/10.56781/ijsret.2023.2.1.0028

Pasquier, P., Eigenfeldt, A., Bown, O., & Dubnov, S. (2016). An introduction to musical metacreation. *Computers in Entertainment*, *14*(2), 1–14. https://doi.org/10.1145/2930672

Prince, S., & Shankar Kumar, K. R. (2012). Survey on effective audio mastering. In *Communications in computer and information science* (pp. 293–301). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35594-3_41

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 66-83. https://doi.org/10.1109/JSTSP.2019.2908700
Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 66-83. https://doi.org/10.1109/JSTSP.2019.2908700
Rai, A., Rout, B. P., Bhushan, & Mehta, A. (2022). Unveiling the power of data: A journey through machine learning techniques. *International Journal of Advanced Research in Science, Communication and Technology*, 42–47. https://doi.org/10.48175/ijarsct-17807

Reddy, R., & Shyam, G. K. (2018). Analysis through machine learning techniques: A survey. https://doi.org/10.1109/icgciot.2018.8753050

Regli, W. C., Hu, X., Atwood, M., & Sun, W. (2000). A survey of design rationale systems: Approaches, representation, capture and retrieval. *Engineering with Computers*, *16*(3-4), 209-235. https://doi.org/10.1007/PL00013715

Relaño-Iborra, H., May, T., Zaar, J., Scheidiger, C., & Dau, T. (2016). Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain. *The Journal of the Acoustical Society of America*, *140*(4), 2670–2679. https://doi.org/10.1121/1.4964505

Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 945-949). IEEE. https://doi.org/10.1109/ICCONS.2018.8663155

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1310-1315). IEEE.

Sterne, J., & Razlogova, M. (2019). Machine learning audio effects: The acoustic entanglements of artificial intelligence and music production. *Popular Music and Society*, *42*(2), 178-196.

Talbot-Smith, M. (2001). *Audio engineer's reference book* (2nd ed.). Focal Press.

Talbot-Smith, M. (2010). *Sound engineering explained*. Routledge. https://doi.org/10.4324/9780080498171

Välimäki, V., & Reiss, J. (2016). All about audio equalization: Solutions and frontiers. *Applied Sciences*, *6*(5), 129. https://doi.org/10.3390/app6050129

Vickers, E. (2010). Metrics for Quantifying Loudness and Dynamics 1.

Vinoth, K., & Datta, S. V. (2021). Core machine learning algorithms: A comprehensive study. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(11), 2595-2604.

Whitaker, J., & Benson, K. B. (2001). *Standard handbook of audio and radio engineering*. McGraw-Hill.

Wong, P. C. (2021). A survey on unsupervised pattern recognition algorithms for data analytics. In *International Conference on Innovative Computing and Communications* (pp. 561-571). Springer.

Wu, X., Qiao, Y., Wang, X., & Tang, X. (2021). Unsupervised feature learning for audio. In *Advances in Neural Information Processing Systems* (pp. 1096-1104).

Zahra, F., Bostanci, Y., & Soyturk, M. (2024). Unsupervised machine learning for anomaly detection in Wi-Fi based IoT networks. *IEEE*. https://doi.org/10.1109/ICCSPA61559.2024.10794232

## Appendices

### Appendix 1. Basic Sound Engineering Principles

Sound engineering is a wide span of concepts and techniques, from basic theory to more sophisticated principals guiding audio recording, processing, and reproduction. In its most conceptual terms, sound is a pressure wave—oscillations that when they reach the thin membranes of the human inner ear cause very small electrical signals to be sent to the brain. The ontological basis of a sound recorder is that one can capture acoustic phenomena on electronic, magnetic or optical media. Playing back those sampled sounds via loudspeakers, headphones, or other playback devices is an important part of the sound engineering process, with each link in the chain having a profound impact on the artistic quality of the resulting sound (Lazzarini, 2021).

Sound engineers also interact with sound waves through the lens of different variables—frequency, phase, and amplitude—for these measurement properties constitute the epistemological basis of sound processing methods (Lazzarini, 2021). Of course, sampling, quantized, discretized are all crucial just as they are in the digital domain because they dictate the ways in which analog sounds from the real world are converted into digital signals. Within this context, equalization is a standard-type of tool that audio engineers use to change the sonic balance of a recording by boosting or attenuating certain frequency ranges (Välimäki & Reiss, 2016).

In addition to this conceptual framework, the sound engineering process includes recording, playback, and mastering. These methods impose the different microphone techniques, the speakers design, and the acoustic properties of recording studios that are determinant factors in audio capture and reproduction (Katz & Katz, 2002; Ballou, 2002). Such paradigmatic transformations in this area have been facilitated by digital technologies through the development of digital signal processing (DSP), compression techniques, and network tools for advanced manipulation of audio data (Whitaker & Benson, 2001). These developments have improved quality and efficiency of sound created for a wide variety of use cases.

The *epistemological territories of sound engineering* includes another important area, psychoacoustics – the science of the perception of sound by humans. All of this information aids in making choices relating to acoustic treatments, sound system design, and optimizing audio signals to be proved from human listeners (Talbot-Smith, 2001). With ongoing advancements in the field, sound engineers tend to interweave developments from the domains of signal processing and machine learning, creating powerful and useful methods of automating and improving many sound production processes (Välimäki & Reiss, 2016).

In short, sound engineering is much more than capturing sound in the technical sense; it also includes decisions at each step, ensuring that sound captured in a way that can be replicated across many playback environments and can be judged aesthetically. This field relies on the epistemology based on the use of digital technologies and psychoacoustic knowledge with applied basic engineering technologies. The following concepts will serve as a conceptual background for the machine-learning techniques used in automating mastering processes.

**Appendix 2. Machine Learning Methods**

Machine learning, a basic epistemological part of artificial intelligence, is the capacity of computer systems to learn from data and make predictions (Kour & Gondhi, 2019). Supervised, unsupervised, and reinforcement learning are three major paradigms of machine learning, and they process labeled or unlabeled data to tackle different problems (Saraswat & Raj, 2021). This part looks to at the algorithms most normally utilized in sound acing, for example, ANN, SVM and DRR.

One of the most common machine learning techniques is artificial neural networks, which represent mathematical functions through a series of connected neurons (Ivanović & Radovanović, 2015). The capacity of ANNs to process certain types of data is influenced greatly by their topological structure—the number of layers and neuron types, and also the types of connections between them. These networks are especially favoured since their learning is comparatively simple and use exceptional generalization capability to generalize through different types of data (Saraswat & Raj, 2021). But, they have some epistemological constraints in terms of the data they can handle that has which birthed SVMs as a paradigm.

When classical prediction methods are not sufficient, support vector machines started to be popular (Ivanović & Radovanović, 2015). For SVMs, it enhances the robustness and generalizability of the predictions over other classifiers by maximizing the margin between the closest feature (support vectors) to either side of the decision boundary. Especially useful for classification and regression problems, they are well-equipped to deal with high-dimensional data and the non-linear nature of problems. Additionally, unlike many other ML methods that rely on statistical inferences from data distribution to obtain predictions, SVMs are mathematically driven using kernel functions (Vinoth & Datta, 2021).
Machine learning techniques have applications outside of audio mastering, such as software development estimation (Y. Singh et al., 2007) and customer behavior analysis (Reddy & Shyam, 2018). That said, to achieve the best models epistemological issues like the bias-variance tradeoff, overfitting, and high dimensionality should be properly handled (Ivanović & Radovanović, 2015). It is anticipated that with the increasing availability of data, the influence of machine learning will only grow with methods such as deep learning, representing especially promising paradigms for future technological advance (Vinoth & Datta, 2021).

## Appendix 3. Supervised and Unsupervised Learning Methods

**Supervised Learning**

Supervised machine learning (SML) is an artificial intelligence paradigm that teaches models from labeled data that the models can use to make predictions for new, previously unseen data (Imran Syed & Dr. Vanita Lokhande, 2024). This has led to the widespread usage of supervized machine learning (SML) techniques in fields like image recognition and natural language processing, where large labeled datasets can be collected. Some common algorithms used in this domain are Naive Bayes, Random Forest, Support Vector Machine, Neural Networks and Decision Trees, with their performances evaluated based on measures of accuracy, computational cost, model complexity and overfitting (Amanpreet Singh et al., 2016).

SML techniques are the game changers in automation of the mastering in the mastering context. They are trained on music stereo audio files and acoustic feature data structures, like dynamic range, spectral balance and loudness, among others. In supervised learning, the training data is labelled manually with categorical values or continuous variables to facilitate guiding the training process (R. Saravanan & P. Sujatha, 2018). The labeled dataset is split into the training and testing parts. The training set allows the model to learn patterns and relationships between the data it learns with the provided correct labels containing examples of what the data is showing. Once the model achieved optimal performance on the training set, it is time to evaluate it on the test set. The success of predictions over the test set is a measure of how well the model generalizes to new data, and is a vital feature of the epistemological value of algorithms of SML (Amanpreet Singh et al. 2016).

Draft SML techniques model systems that can automatically optimise a mastering process which leads to greater efficiency and lower manual cost for mastering operations (Pradeep Verma & Dr. Poornima Tyagi, 2020) In particular, classification algorithms can determine whether a specific mastering process causes the high feature very similar, and by categorizing the particular mastering operation with which these processes are applied, such as equalization, dynamic range compression or stereo enhancement [6]

**Unsupervised Learning**

UML is a data-centric framework that allows algorithms to learn inner structures and patterns from unlabeled data with no human input (Samreen Naeem et al., 2023). UML methods are classified into dimensionality reduction, clustering, and deep learning-based methods, which enable discovering the underlying structures in data without requiring any labeled examples (Xiangdong Wu et al., 2021).

UML has not only contributed epistemologically, especially for speech signal processing tasks, where so far unsupervised models ran mostly better than the traditional ways of designing and building systems based on supervised learning techniques (histograms, HMM, DNN-SD), but a lot of creative minds in UML as well as other dedicated researchers have created, implemented and made available packages, tools, and frameworks that have sparked and propelled the nowadays unavoidable as well as revolutionary community based research and progress in speech technology. Dynamic horsepower: UML now processes massive amounts of structured data hierarchically to create more sophisticated analytical tools for speech, music and audio signal recognition and classification. One of the major approaches in this area is the stochastic feed-forward neural network; an unsupervised model that is extremely effective at using contrastive divergence training methods to improve neural performance. Based on ideas of mean-field theory, the dynamics of neural networks have been optimally modeled to optimize the neural networks processing audio signals (Xiangdong Wu et al., 2021).

We showcase an instance of unsupervised learning for music processing, namely the stochastic binary vector machine, a model that can be used for audio mastering in the absence of labeling prior to mastering the audio data. By addressing problems in ferromagnetic systems, the statistical mechanics of learning and inference have been successfully applied to complex audio processing tasks. This method has an epistemological compatibility with Hebb's concept of "unsupervised learning" and move beyond some bottlenecks of classical models like the Hopfield network. Despite the positive aspects of unsupervised learning, it suffers from the epistemological challenges of biases and interpretability of models and data in train (Aviral Rai et al., 2024). Despite this, researchers are identifying methodological frameworks to assist tailoring UML methods for audio data characteristics (Andri M Kristijansson & Tyr Aegisson, 2022).