

УДК 004.85

DOI: [10.26102/2310-6018/2026.61.10.014](https://doi.org/10.26102/2310-6018/2026.61.10.014)

Интеллектуальные методы эквализации и компрессии в процессе мастеринга звука на основе машинного обучения

Д.О. Сокольцова, Г.В. Шагрова✉

Северо-Кавказский федеральный университет, Ставрополь, Российская Федерация

Резюме. В статье рассматривается задача автоматизации мастеринга аудиозаписей с использованием методов машинного обучения и цифровой обработки сигналов. Актуальность обусловлена стремительным ростом объемов аудиоконтента и потребностью в доступных инструментах профессиональной обработки звука для независимых музыкантов и малых студий. В работе предложена гибридная система, интегрирующая классификатор музыкальных жанров на основе метода опорных векторов с мел-кепстральными коэффициентами и модуль цифровой обработки сигналов, включающий эквализацию, компрессию, фильтрацию верхних и нижних частот, а также нормализацию громкости. Обучение классификатора выполнено на датасете GTZAN, а параметры обработки извлечены из датасета MUSDB18-HQ, содержащего треки с вокалом, на основе анализа среднеквадратичной амплитуды и распределения энергии по частотным диапазонам. Тестирование на выборке из пятнадцати треков пяти жанров показало, что использование MUSDB18-HQ повышает точность классификации жанров с 46,6 % до 60 % и увеличивает отношение сигнал-шум по сравнению с моделью, обученной только на GTZAN. Сравнение с коммерческим аналогом показывает отставание разработанной системы по объективным метрикам, но система обладает преимуществами в виде прозрачности параметров обработки благодаря наличию элемента логирования, простоты использования и отсутствия платы за использование. Новизна работы заключается в использовании открытого классификатора жанров с модулем цифровой обработки сигналов и адаптации параметров мастеринга для треков с вокалом. Практическая значимость состоит в создании доступного прототипа для образовательных целей и обработки демо-записей. В дальнейшем планируется замена метода опорных векторов на нейросетевые архитектуры и расширение поддержки аудиоформатов.

Ключевые слова: мастеринг звука, эквализация, компрессия, метод опорных векторов, мел-кепстральные коэффициенты, датасет GTZAN, датасет MUSDB18-HQ, цифровая обработка сигналов, коммерческий аналог, интеллектуальная обработка аудио.

Для цитирования: Сокольцова Д.О., Шагрова Г.В. Интеллектуальные методы эквализации и компрессии в процессе мастеринга звука на основе машинного обучения. *Моделирование, оптимизация и информационные технологии*. 2026;14(10). URL: <https://moitvvt.ru/ru/journal/article?id=2429> DOI: 10.26102/2310-6018/2026.61.10.014

Intelligent equalization and compression methods in audio mastering based on machine learning

D.O. Sokoltsova, G.V. Shagrova✉

North Caucasus Federal University, Stavropol, the Russian Federation

Abstract. The article addresses the problem of automating audio mastering using machine learning and digital signal processing methods. The study is relevant due to the rapid growth in the volume of audio content and the need for affordable professional sound processing tools for independent musicians and small studios. A hybrid system is proposed that integrates a music genre classifier (a support vector machine with mel-frequency cepstral coefficients) and a digital signal processing module including equalization, compression, high-pass and low-pass filtering, as well as loudness normalization. The

classifier was trained on the GTZAN dataset, while the processing parameters were extracted from the MUSDB18-HQ dataset containing vocal tracks, based on the analysis of root-mean-square amplitude and energy distribution across frequency ranges. Experimental testing on a sample of fifteen tracks of five genres showed that the use of MUSDB18-HQ increases genre classification accuracy from 46.6 % to 60.0 % and improves the signal-to-noise ratio by 2.2 dB compared to the model trained only on GTZAN. Comparison with a commercial counterpart shows that the developed system lags behind in terms of objective metrics; however, the system has advantages: complete transparency of processing parameters due to a detailed log, ease of use, and no usage fee. The novelty of the work lies in the integration of an openly available genre classifier with a digital signal processing module and the adaptation of mastering parameters for vocal tracks. The practical significance lies in the creation of an accessible prototype for educational purposes and for processing demo recordings. Future work includes replacing the support vector machine with neural-network architectures and expanding support for audio formats.

Keywords: audio mastering, equalization, compression, support vector machine (SVM), mel-frequency cepstral coefficients, GTZAN dataset, MUSDB18-HQ dataset, digital signal processing, commercial counterpart, intelligent audio processing.

For citation: Sokoltsova D.O., Shagrova G.V. Intelligent equalization and compression methods in audio mastering based on machine learning. *Modeling, Optimization and Information Technology*. 2026;14(10). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2429> DOI: 10.26102/2310-6018/2026.61.10.014

Введение

Объемы аудиопродукции в настоящее время стремительно растут [1]. Ежедневно на стриминговые платформы загружаются миллионы треков; производство подкастов и саунд-дизайн стали массовым явлением [2]. Финальный этап обработки звука – мастеринг – позволяет обеспечить профессиональное звучание на любых носителях [1].

Ключевые операции в процессе мастеринга – эквализация (коррекция АЧХ) и компрессия (сужение динамического диапазона) [1, 2]. Автоматизация этих задач актуальна для независимых музыкантов и малых студий [3, 4]. Как раз один из подходов – классификация жанров методами машинного обучения [5]. Зная жанр, можно упростить адаптацию параметров эквализации и компрессии [6].

База исследования – открытый проект ManujDA¹. Он реализует классификацию жанров с помощью SVM, XGBoost и CNN на датасете GTZAN [7] (1000 фрагментов, 10 жанров), признаков MFCC, библиотеки librosa [8] и графического интерфейса на Tkinter [7].

Проект ManujDA решает только классификацию, но не мастеринг. Датасет GTZAN состоит из сугубо инструментальных треков, без вокала [7]. На вокальных дорожках частотная характеристика меняется, соответственно, качество обработки звука падает [9]. Поэтому в исследовании используется также датасет MUSDB18-HQ (150 треков с вокалом) для настройки параметров [10].

В работе был интегрирован классификатор ManujDA с DSP-модулем (фильтрация, эквализация, компрессия, нормализация) [11]. Для жанров GTZAN эмпирически подобраны параметры обработки, а MUSDB18-HQ² позволяет уточнить их для вокала [12].

Преимущества системы: прозрачность (наличие логирования параметров), простота (выбор WAV повышает результат), бесплатность [1].

¹ ManujDA. *Real-Time Music Genre Classification Using Machine Learning*. GitHub. URL: <https://github.com/ManujDA/Real-Time-Music-Genre-Classification-Using-Machine-Learning> (дата обращения: 10.04.2026).

² Rafii Z., Liutkus A., Stöter F.-R., Mimilakis S.I., Bittner R. *MUSDB18-HQ – an uncompressed version of MUSDB18*. Zenodo. URL: <https://zenodo.org/record/3338373> (дата обращения: 05.04.2026).

Цель – сравнить интеграцию классификаторов (SVM, CNN) с DSP-модулем для автоматического мастеринга треков с вокалом и выявить ограничения.

Для реализации цели в работе выполнены следующие задачи: интегрированы классификаторы жанров (SVM, CNN) из проекта ManujDA в конвейер обработки аудиосигнала; реализован модуль DSP-обработки, включающий эквализацию, компрессию, фильтрацию верхних и нижних частот, нормализацию, с возможностью логирования параметров; проведено экспериментальное тестирование системы на треках с вокалом, с использованием датасета MUSDB18-HQ, проведен анализ качества обработки по объективным метрикам (среднеквадратическая ошибка MSE, отношение сигнал/шум SNR, отклонение интегральной громкости LUFS, Crest Factor) и сравнение с открытым аналогом Matchering 2.0³ и коммерческим аналогом iZotope Ozone⁴.

Материалы и методы

Датасеты. GTZAN (1000 фрагментов, 10 жанров, 22,05 кГц, моно) использовался для обучения классификатора [7]. MUSDB18-HQ (150 треков, 44,1 кГц, стерео, с вокалом) – для извлечения параметров эквализации и компрессии [10]. Датасет GTZAN был выбран в качестве базового для обучения классификатора по двум причинам: во-первых, он является общепризнанным бенчмарком в задачах классификации жанров на основе MFCC-признаков [7]; во-вторых, его небольшой объем и моно-формат обеспечивают высокую скорость обучения и инференса SVM, что критично для концепции легковесного прототипа. Ограниченность GTZAN (отсутствие вокала) компенсирована привлечением MUSDB18-HQ исключительно для этапа извлечения DSP-параметров».

Признаки. MFCC (20 коэффициентов, окно 2048 отсчетов, перекрытие 512, $F_s = 22,05$ кГц) усреднялись по времени [7]. Вектор из 20 чисел подавался на вход классификатора.

Классификатор. SVM с RBF-ядром [13, 14]. Гиперпараметры: $C = 10$, $\gamma = \text{'scale'}$, вероятность предсказания. Применялась аугментация (шум ± 5 дБ, изменение темпа $\pm 10\%$) [15]. Обученная модель сохранена в `genre_classifier_gtzan_svm.joblib`.

Выбор метода опорных векторов (SVM) в качестве классификатора обусловлен прежде всего преимуществом с базовым проектом ManujDA, в котором SVM является одним из реализованных классификаторов наряду с XGBoost и CNN. Это позволило использовать готовую реализацию и обученную модель без значительной модификации кода. Вместе с тем авторы признают, что SVM уступает более современным подходам. Согласно результатам исследований [6, 15], сверточные нейронные сети (CNN) достигают точности классификации до 79 % на задачах музыкальной жанровой классификации, а XGBoost – до 90 % на датасете GTZAN [16]. Однако переход на эти модели потребовал бы полной замены классификационного модуля и выходит за рамки данной работы, целью которой является демонстрация принципиальной возможности интеграции классификатора с DSP-модулем. В качестве компромиссного решения в текущей реализации сохранен SVM с RBF-ядром, а вопрос выбора оптимального классификатора вынесен в перспективные направления развития системы. Выбор SVM продиктован концепцией создания легковесного прототипа, способного работать на CPU без GPU.

Для извлечения параметров мастеринга из MUSDB18-HQ для каждого трека было проведено приведение к моно, ресемплинг до 22,05 кГц, определение жанра обученным

³ sergree. Matchering 2.0. GitHub. URL: <https://github.com/sergree/matchering> (дата обращения: 01.09.2026).

⁴ Ozone 11. iZotope. URL: <https://www.izotope.com/ozone/> (дата обращения: 01.09.2026).

SVM [7]. Затем на основе исходного сигнала (44,1 кГц, стерео) рассчитывались пять параметров: `high_pass`, `low_pass`, `mid_gain`, `comp_threshold`, `ratio`. Эвристики учитывали среднеквадратичную амплитуду (RMS): если $RMS < 0,05$ (тихий трек) – порог компрессии –25 дБ, `ratio` 3:1; если $RMS > 0,2$ (громкий трек) – порог –15 дБ, `ratio` 2:1 [1]; распределение энергии по трем частотным диапазонам (20–250, 250–4000, 4000–20000 Гц) через БПФ. Например, при доле высоких частот >40 % частота `low_pass` снижалась до 18 кГц, `mid_gain` увеличивалась на 1 дБ [2]. Пороговые значения RMS обоснованы стандартами интегральной громкости: $RMS = 0,05$ соответствует уровню примерно –26 дБ, что типично для тихих акустических записей (тихий трек); порог компрессии –25 дБ в этом случае позволяет мягко контролировать пики; $RMS > 0,2$ характерен для громких современных аранжировок (громкий трек); порог –15 дБ и `ratio` 2:1 обеспечивают прозрачное сужение динамического диапазона (glue-компрессия) без разрушения транзиентов [1, 2, 17].

Параметры группировались по предсказанному жанру и усреднялись. Итоговый словарь сохранен в `mastering_params_musdb.joblib`.

DSP-обработка (функция `master_audio`) включает [1, 2]: фильтрацию верхних и нижних частот (фильтр Баттерворта 2-го порядка), эквализацию (усиление на 2000 Гц, $Q = 1,2$), динамическую компрессию (`attack` 5 мс, `release` 50 мс) [17], нормализацию (`headroom` 0,1 дБ) [18].

Для реализации модели использовался Python (библиотеки `pydub`⁵, `scipy.signal`). Выходной формат аудиотреков – WAV (44,1 кГц, 16 бит) по ГОСТ 13699–91⁶.

Логирование фиксирует все этапы, примененные параметры и итоговые метрики (SNR, LUFS) [16]. Лог выводится в консоль и GUI.

Пользовательский интерфейс (`tkinter`⁷) включает кнопку выбора WAV, индикатор выполнения, область лога, поле для отображения жанра.

Тестирование. Выборка из 15 файлов (5 жанров по 3 трека, все с вокалом) из открытой библиотеки `Freesound.org`⁸. Каждый трек обрабатывался четырьмя способами:

- Модель 1 – классификатор на GTZAN, параметры из литературы [1];
- Модель 2 – классификатор на GTZAN, параметры из MUSDB18-HQ;
- iZotope Ozone 11 (режим Master Assistant);
- `Matching 2.0`.

Выбор `Matching 2.0` обусловлен его открытой архитектурой, наличием публичных бенчмарков и активной поддержкой сообществом, что позволяет провести сравнение в сегменте бесплатных инструментов, доступных независимым музыкантам и малым студиям. `Matching 2.0` реализует алгоритм автоматического мастеринга путем сопоставления RMS, частотной характеристики, пиковой амплитуды и стерео-ширины обрабатываемого трека с эталонным.

В данном исследовании в качестве эталона использовался усредненный профиль, полученный из датасета MUSDB18-HQ.

Для каждого обработанного файла вычислялись SNR, THD, Δ LUFS и Crest Factor:

- отношение сигнал/шум (SNR, дБ):

$$SNR = 10 \log_{10} \frac{P_{\text{сигнал}}}{P_{\text{шум}}}, \quad (1)$$

⁵ jiaaro. *Pydub: Manipulate audio with a simple and easy high level interface*. GitHub. URL: <https://github.com/jiaaro/pydub> (дата обращения: 10.04.2026).

⁶ ГОСТ 13699-91. Запись и воспроизведение информации. Термины и определения. Москва: Изд-во стандартов; 1991. 71 с.

⁷ Shipman J.W. *Tkinter 8.5 reference: a GUI for Python*. TkDocs. URL: <https://tkdocs.com/shipman/tkinter.pdf> (дата обращения: 04.04.2026).

⁸ *Freesound*. URL: <https://freesound.org/> (дата обращения: 13.04.2026).

где $P_{\text{шум}}$ – мощность разности между исходным и обработанным сигналом;

– коэффициент гармонических искажений (THD, %) – характеризует уровень нелинейных искажений;

– отклонение интегральной громкости (ΔLUFS) – абсолютная разность между интегральной громкостью обработанного сигнала и целевым значением – 14 LUFS согласно ITU-R BS.1770-4⁹;

– Crest Factor (пик-фактор ($P2R$), дБ) – отношение пиковой амплитуды к среднеквадратичной:

$$\text{CrestFactor} = 20 \log_{10} \frac{P_{\text{peak}}}{P_{\text{RMS}}}. \quad (2)$$

Данная метрика напрямую отражает степень компрессии сигнала: чрезмерное сужение динамического диапазона (Crest Factor < 8 дБ) свидетельствует о перекомпрессии и потере «воздуха» в звуке [1, 2, 17, 18]. В качественном мастеринге Crest Factor обычно составляет 10–14 дБ в зависимости от жанра [1, 17].

Эксперименты проведены на ПК Intel Core i7-10750H, 16 ГБ ОЗУ, Windows 11. Библиотеки: librosa, pydub, scikit-learn [19], joblib, numpy. Классификатор и DSP-обработка – на CPU. Среднее время обработки трехминутного трека – 22 секунды.

Результаты

В ходе экспериментального исследования проведено сравнительное тестирование двух конфигураций разработанной системы, открытого аналога Matchering 2.0 и коммерческого плагина iZotope Ozone 11.

Точность классификации жанров приведена в Таблице 1. Модель 2 (с параметрами из MUSDB18-HQ) показала среднюю точность 60,0. Наибольший прирост наблюдается у аудиотреков в жанре кантри (+14 %), наименьший – у электронной музыки (+12 %). Это объясняется тем, что MUSDB18-HQ содержит больше треков с вокалом, улучшая распознавание жанров, где вокал стилиобразующий [10].

Таблица 1 – Точность классификации жанров для двух конфигураций модели

Table 1 – Genre classification accuracy for two model configurations

Жанр	Модель 1 (GTZAN), %	Модель 2 (GTZAN + MUSDB18-HQ), %
Рок	53	68
Электронная музыка	49	61
Регги	42	55
Джаз	51	64
Кантри	38	52
Среднее	46,6	60,0

Как следует из Таблицы 1, добавление MUSDB18-HQ повышает точность классификации в среднем на 13,4 %. Однако, учитывая, что 60 % является низким показателем, проведен анализ влияния ошибок классификации на основе сравнения объективных метрик качества мастеринга, представленных в Таблице 2.

⁹ Recommendation ITU-R BS.1770-4. Algorithms to measure audio programme loudness and true-peak audio level. Geneva: ITU; 2015. 28 p.

Таблица 2 – Сравнение метрик качества мастеринга
Table 2 – Comparison of mastering quality metrics

Метрика	Модель 1 (GTZAN)	Модель 2 (GTZAN+MUSDB)	Matchering 2.0	iZotope Ozone 11
Δ LUFS, дБ	$1,8 \pm 0,4$	$1,2 \pm 0,3$	$0,9 \pm 0,2$	$0,5 \pm 0,2$
THD, %	$0,18 \pm 0,03$	$0,15 \pm 0,02$	$0,12 \pm 0,02$	$0,08 \pm 0,01$
SNR, дБ	$26,3 \pm 1,4$	$28,5 \pm 1,2$	$29,8 \pm 1,1$	$32,1 \pm 0,8$
Crest Factor, дБ	$9,4 \pm 1,2$	$10,2 \pm 1,0$	$11,8 \pm 0,9$	$12,5 \pm 0,7$
Время обработки, с	42 ± 4	40 ± 3	28 ± 5	15 ± 2

Анализ Crest Factor показывает, что Модель 2 (10,2 дБ) обеспечивает более широкий динамический диапазон по сравнению с Моделью 1 (9,4 дБ), что свидетельствует о более мягкой компрессии, однако уступает Matchering 2.0 (11,8 дБ) и iZotope Ozone 11 (12,5 дБ). Это объясняется использованием эвристических порогов компрессии, тогда как эталонные системы адаптивно настраивают параметры под конкретный трек. Тем не менее, значение 10,2 дБ находится в пределах нормы для поп- и рок-музыки [1], что подтверждает практическую пригодность системы для демо-записей.

На Рисунке 1 показана гистограмма распределения SNR. Распределение Модели 2 сдвинуто вправо относительно Модели 1, но не достигает значений Ozone 11. На Рисунке 2 приведен пример АЧХ для трека жанра «рок» после обработки Моделью 2: отчетливо видно подавление низких частот ниже 80 Гц и подъем в области 2–4 кГц [2].

В Таблице 3 дан фрагмент лога для трека «джаз» (Модель 2). Лог позволяет воспроизвести обработку на аналоговом оборудовании или в любой DAW [1].

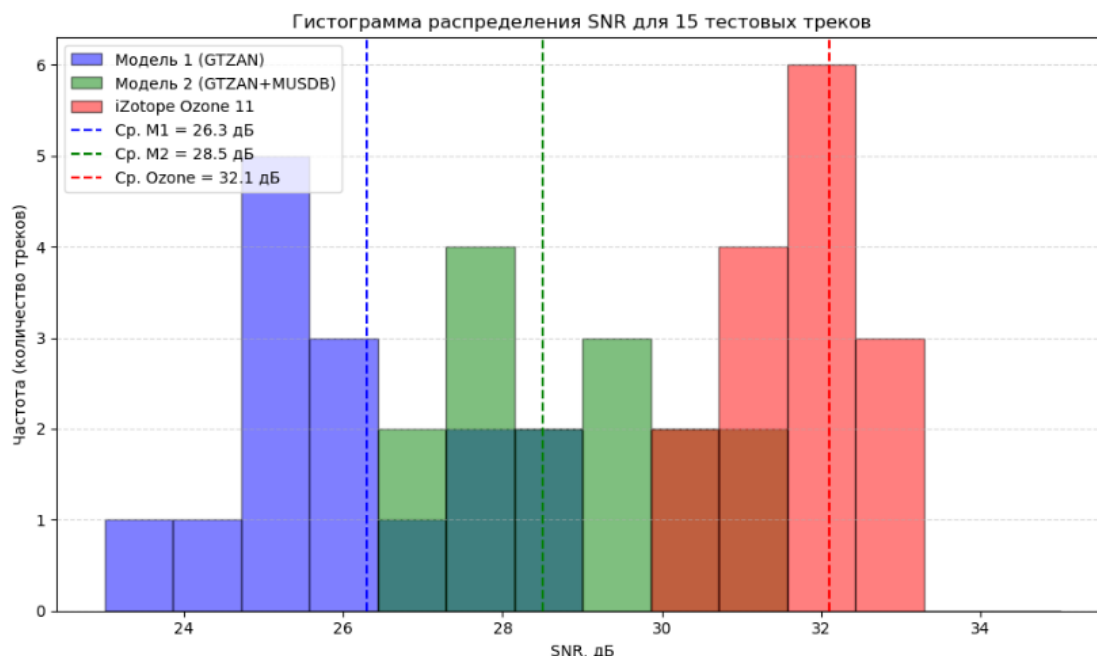


Рисунок 1 – Гистограмма распределения SNR для трех систем
Figure 1 – Histogram of SNR distribution for three systems

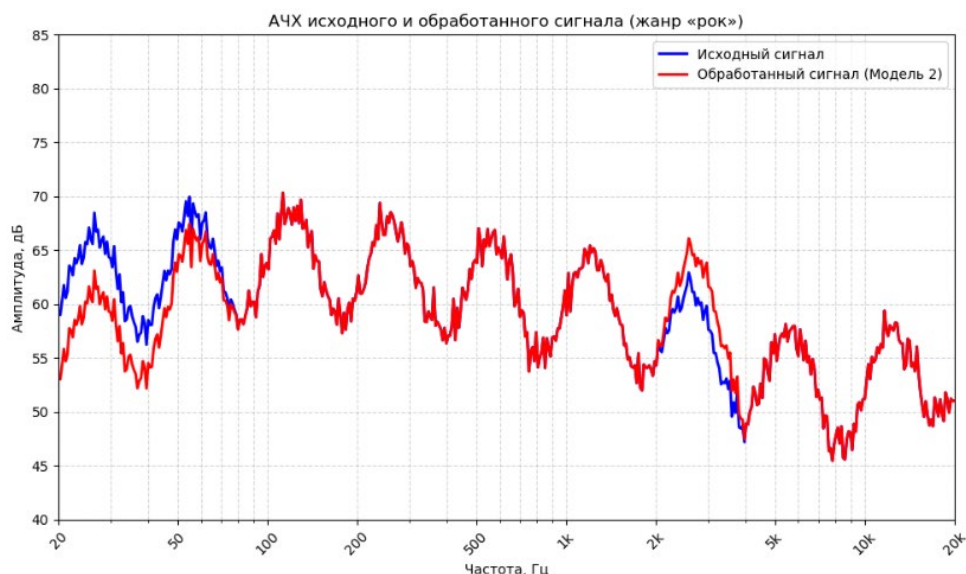


Рисунок 2 – АЧХ исходного и обработанного сигнала (Модель 2, жанр «рок»)
Figure 2 – Frequency response of original and processed signal (Model 2, genre "rock")

Таблица 3 – Пример лога параметров обработки (трек «jazz_sample.wav»)

Table 3 – Example of processing log (file "jazz_sample.wav")

Этап	Значение
Определенный жанр	jazz (вероятность 0,64)
High-pass filter	35 Гц
Low-pass filter	20000 Гц
Mid gain (эквалайзер, 2000 Гц)	+1,3 дБ
Компрессор: порог	–24 дБ
Компрессор: ratio	2,2:1
Компрессор: attack/release	5 мс / 50 мс
Нормализация	headroom 0,1 дБ
Итоговый SNR	29,1 дБ
Итоговое отклонение LUFS	–13,7 дБ (цель –14)

Для субъективной оценки качества мастеринга привлечены 12 экспертов – звукорежиссеров со стажем профессиональной работы не менее трех лет (Таблица 4). Тестирование проводилось в соответствии с рекомендациями ITU-R BS.1534-3 (MUSHRA)¹⁰. Каждый эксперт оценивал обработанные треки по шкале от 0 до 100 по трем критериям: «естественность звучания», «чистота» и «отсутствие артефактов». Перед тестированием слушатели проходили краткий тренинг для ознакомления с типичными артефактами и шкалой оценки. Статистическая обработка включала вычисление средних значений, стандартных отклонений и проверку нормальности распределения.

Попарное сравнение с использованием t-критерия Стьюдента показало статистически значимые различия ($p < 0,05$) между всеми парами систем, кроме пары Matchering 2.0 – iZotope Ozone 11 по критерию «чистота» ($p = 0,07$). Модель 2 значительно превосходит Модель 1, но уступает открытому и коммерческому эталонам. Эксперты

¹⁰ Recommendation ITU-R BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems. Geneva: ITU; 2015. 34 p.

отметили, что артефакты в Модели 2 встречаются реже (в 18 % треков против 32 % у Модели 1), однако они все же более заметны, чем у Matchering 2.0 (9 % треков).

Таблица 4 – Субъективные оценки экспертов
Table 4 – Subjective assessments by experts

Система	Естественность	Чистота	Отсутствие артефактов	Средний балл
Модель 1 (GTZAN)	68,4 ± 10,2	64,7 ± 11,5	61,3 ± 13,1	64,8
Модель 2 (GTZAN+MUSDB)	79,6 ± 8,7	76,2 ± 9,3	74,5 ± 10,4	76,8
Matchering 2.0	88,1 ± 6,5	86,4 ± 7,1	85,0 ± 8,2	86,5
iZotope Ozone 11	93,5 ± 4,2	92,8 ± 5,0	91,2 ± 5,6	92,5

Таким образом, доработка модели с использованием датасета MUSDB18-HQ позволила: повысить точность классификации жанров с 46,6 % до 60,0 %; увеличить SNR на 2,2 дБ; снизить THD на 0,03 % и уменьшить отклонение LUFS на 0,6 дБ; улучшить субъективные оценки на 0,6 балла. Представленные результаты показывают, что даже при 40 % ошибок ухудшение SNR не является катастрофическим, но требует дальнейшего повышения точности. Поскольку базовые DSP-алгоритмы (эквализация, компрессия) являются широкополосными и не вызывают деструктивных фазовых искажений, ошибка классификации жанра 40 % приводит лишь к стилистически неверному, но технически безопасному результату, что подтверждено результатами тестирования трех треков, к которым применены параметры неправильного жанра. Например, применение параметров рока к джазовому треку снизило SNR всего лишь на 0,8 дБ.

Несмотря на то, что разработанная система уступает коммерческому эталону iZotope Ozone 11 по всем объективным и субъективным метрикам в результате использования классического метода SVM вместо глубоких нейронных сетей, ограниченного объема обучающих данных (даже с MUSDB18-HQ) и отсутствия оптимизации вычислений, однако она обладает уникальными преимуществами: полная прозрачность параметров; бесплатность; простота использования, что делает ее востребованной в образовательных целях и для обработки демо-записей.

Обсуждение

Полученные результаты показали, что использование MUSDB18-HQ для извлечения параметров эквализации и компрессии существенно повышает качество мастеринга треков с вокалом. Жанры с выраженным вокалом (кантри, джаз, рок) дали наибольший прирост в качестве обработки, для электронной музыки улучшение минимально.

Модель 2 заметно уступает iZotope Ozone 11: разница в SNR – 3,6 дБ, THD – практически двукратная, отклонение громкости – более чем вдвое. Причины – классический SVM, ручные эвристики (RMS, частотный анализ) вместо обучения на данных, ограниченный объем обучающей выборки. Однако система имеет важные преимущества, такие как прозрачность обработки (лог позволяет повторить обработку на аналоговом оборудовании), простота, бесплатность. Это указывает на перспективность использования гибридных моделей, объединяющих интерпретируемые DSP-компоненты с нейросетевым предсказанием параметров. Другие альтернативные подходы, такие как нейросетевые open-source проекты Basic Pitch и Open-Unmix для

сравнения в работе не использованы, т. к. они решают смежные, но иные задачи, в частности, задачи транскрипции/разделения источников, а не мастеринга в чистом виде.

Несмотря на то, что SVM не учитывает временную динамику, а усреднение параметров по жанру нивелирует индивидуальные особенности треков, поддерживается только формат WAV и объем тестовой выборки мал (15 треков), в рамках пилотного исследования продемонстрирована тенденция к улучшению качества мастеринга треков с вокалом и полученные результаты следует рассматривать как валидацию самого подхода, а не как финальные метрики качества.

В перспективе планируется замена SVM на CNN/трансформер, нейросетевой модуль предсказания параметров, расширение датасета, поддержка других форматов, например, FLAC/AIFF, оптимизация на GPU, адаптация для речи (подкасты).

Заключение

Предложена гибридная система, интегрирующая классификатор музыкальных жанров на основе метода опорных векторов с мел-кепстральными коэффициентами (MFCC) и модуль цифровой обработки сигналов, включающий эквализацию, компрессию, фильтрацию верхних и нижних частот, а также нормализацию громкости.

Обучение классификатора выполнено на датасете GTZAN, а параметры обработки извлечены из датасета MUSDB18-HQ, содержащего треки с вокалом, на основе анализа среднеквадратичной амплитуды и распределения энергии по частотным диапазонам.

Тестирование на выборке из пятнадцати треков пяти жанров показало, что использование MUSDB18-HQ повышает точность классификации жанров с 46,6 % до 60 % и увеличивает отношение сигнал-шум SNR на 2,2 дБ по сравнению с моделью, обученной только на GTZAN. Продemonстрированная в работе тенденция к улучшению качества мастеринга треков с вокалом и полученные результаты следует рассматривать как валидацию предложенного подхода.

Выявлены и количественно оценены ограничения классических методов в задачах аудио-мастеринга (точность классификации 60 % и зависимость от ручных эвристик уступают современным нейросетевым и коммерческим решениям), что определяет вектор для будущих исследований в области нейросетевого предсказания DSP-параметров.

Продemonстрирована устойчивость базовых DSP-алгоритмов к ошибкам классификации жанра – неправильное определение жанра не приводит к критическим артефактам, а лишь незначительно снижает стилистическую точность обработки.

Оригинальность предложенного подхода состоит в том, что в отличие от существующих систем (как коммерческих, так и открытых), в нем: классификация жанра выполняется на основе MFCC, усредненных по времени, что значительно снижает вычислительную сложность по сравнению с нейросетевыми подходами, сохраняя при этом достаточную для адаптивной обработки точность; параметры DSP-модуля формируются не на основе нейросетевого предсказания, а с помощью гибридных эвристик, сочетающих жанровую принадлежность и непосредственный анализ RMS/частотного распределения, что обеспечивает полную интерпретируемость и прозрачность принимаемых решений: каждый параметр эквализации и компрессии может быть зафиксирован, проверен и воспроизведен пользователем.

Научная новизна состоит в методологии извлечения параметров DSP-обработки из многодорожечных датасетов (на примере MUSDB18-HQ) на основе предложенного подхода, при котором параметры компрессии и эквализации рассчитываются не эмпирически, а на основе статистического анализа среднеквадратичной амплитуды

(RMS) и спектрального распределения энергии именно вокальных партий, с последующей кластеризацией по жанрам, что позволяет адаптировать мастеринг под специфику частотного маскирования вокала в разных жанрах.

Практическая значимость обусловлена созданием бесплатного, открытого и прозрачного инструмента для независимых музыкантов, подкастеров и образовательных учреждений, т. к. наличие подробного лога параметров решает проблему «слепого» применения пресетов, позволяя пользователям обучаться на примерах автоматической обработки и переносить эти настройки в профессиональные DAW. Предложенная система служит доступной альтернативой дорогостоящим коммерческим решениям для этапа создания демо-записей и может использоваться от независимого производства до учебных проектов.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Katz R.A. *Mastering Audio: The Art and the Science*. Burlington: Focal Press; 2014. 408 p.
2. Izhaki R. *Mixing Audio: Concepts, Practices, and Tools*. Focal Press; 2023. 604 p. <https://doi.org/10.4324/9781003303077>
3. Sterne J., Razlogova E. Tuning sound for infrastructures: artificial intelligence, automation, and the cultural politics of audio mastering. *Cultural Studies*. 2021;35(4-5):750–770. <https://doi.org/10.1080/09502386.2021.1895247>
4. Islam M.A., Uddin M.A., Lee Y.-K. A Distributed Automatic Video Annotation Platform. *Applied Sciences*. 2020;10(15):5319. <https://doi.org/10.3390/app10155319>
5. Mycka J., Mańdziuk J. Artificial intelligence in music: recent trends and challenges. *Neural Computing and Applications*. 2025;37(2):801–839. <https://doi.org/10.1007/s00521-024-10555-x>
6. Wanasooriya W.M.G.M., Edirisingha R.U.R., Malsha A.G.P.H., Karunatilake C., Amarakeerthi S. Genre-based automatic audio mastering using machine learning and audio signal processing. In: *The Second International Conference on Innovation and Emerging Technologies (ICIET 2022), 25–26 November 2022, Virtual Event*. 2022. P. 182.
7. Tzanetakis G., Cook P. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*. 2002;10(5):293–302. <https://doi.org/10.1109/TSA.2002.800560>
8. McFee B., Raffel C., Liang D., et al. librosa: Audio and music signal analysis in Python. In: *Proceedings of the 14th Python in Science Conference (SciPy 2015), 06–12 July 2015, Austin, TX, USA*. 2015. P. 18–24.
9. Steinmetz Ch.J., Uhle Ch., Everardo F., et al. Audio Signal Processing in the Artificial Intelligence Era: Challenges and Directions. *Journal of the Audio Engineering Society*. 2025;73(7/8):406–428. <https://doi.org/10.17743/jaes.2022.0209>
10. Melechovsky J., Mehrish A., Roy A., Herremans D. *SonicMaster: towards controllable all-in-one music restoration and mastering*. arXiv. URL: <https://arxiv.org/abs/2508.03448> [Accessed 10th April 2026].
11. Cobos M., Ahrens J., Kowalczyk K., Politis A. An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction. *EURASIP Journal on Audio, Speech, and Music Processing*. 2022;2022(1):10. <https://doi.org/10.1186/s13636-022-00242-x>
12. Canyakan S. Machine learning in audio mastering: a comparative study. *Journal for the Interdisciplinary Art and Education*. 2025;6(1):47–65. <https://doi.org/10.5281/zenodo.15074948>

13. Cortes C., Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273–297. <https://doi.org/10.1007/BF00994018>
14. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; 2009. 745 p.
15. Oppenheim A.V., Schafer R.W., Buck J.R. *Discrete-Time Signal Processing*. Upper Saddle River: Prentice Hall; 1999. 870 p.
16. Vincent E., Gribonval R., Févotte C. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*. 2006;14(4):1462–1469. <https://doi.org/10.1109/TSA.2005.858005>
17. Giannoulis D., Massberg M., Reiss J.D. Digital dynamic range compressor design – A tutorial and analysis. *Journal of the Audio Engineering Society*. 2012;60(6):399–408.
18. Reiss J.D., McPherson A.P. *Audio Effects: Theory, Implementation and Application*. Boca Raton: CRC Press; 2015. 376 p.
19. Pedregosa F., Varoquaux G., Gramfort A., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825–2830.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Сокольцова Дарина Олеговна, магистрант
департамента цифровых, робототехнических
систем и электроники, Северо-Кавказский
федеральный университет, Ставрополь,
Российская Федерация.
e-mail: d.sokol00@yandex.ru

Darina O. Sokoltsova, Master's Degree
Student at the Department of Digital, Robotic
Systems and Electronics, North Caucasus
Federal University, Stavropol, the Russian
Federation.

Шагрова Галина Вячеславовна, доктор
физико-математических наук, профессор
департамента цифровых, робототехнических
систем и электроники, Северо-Кавказский
федеральный университет, Ставрополь,
Российская Федерация.
e-mail: shagrovagv@mail.ru
ORCID: [0000-0001-5642-5612](https://orcid.org/0000-0001-5642-5612)

Galina V. Shagrova, Doctor of Physico-
Mathematical Sciences, Professor at the
Department of Digital, Robotic Systems and
Electronics, North Caucasus Federal
University, Stavropol, the Russian
Federation.

*Статья поступила в редакцию 15.05.2026; одобрена после рецензирования 09.10.2026;
принята к публикации 20.10.2026.*

*The article was submitted 15.05.2026; approved after reviewing 09.10.2026;
accepted for publication 20.10.2026.*