ТЕХНИЧЕСКИЕ НАУКИ / TECHNICAL SCIENCE

УДК 004.032.26:659.148 https://doi.org/10.33619/2414-2948/118/14

МЕТОДИКА РАЗРАБОТКИ АУДИОДИЗАЙНА ДЛЯ РЕКЛАМЫ СРЕДСТВАМИ ГЕНЕРАТИВНЫХ НЕЙРОСЕТЕЙ

©Каршакова Л. Б., *ORCID: 0000-0003-2158-2508, SPIN-код: 2813-4762, канд. техн. наук, Российский государственный университет им. А. Н. Косыгина (Дизайн. Искусство. Технологии), г. Москва, Россия, lkarshak@mail.ru*
©Нагай С. С., *ORCID: 0009-0005-5173-6688, Российский государственный университет им. А. Н. Косыгина (Дизайн. Искусство. Технологии), г. Москва, Россия, lana.nagaina@yandex.ru*
©Павлинов А. М., *ORCID: 0009-0005-7717-9891, Российский государственный университет им. А. Н. Косыгина (Дизайн. Искусство. Технологии), г. Москва, Россия, aleksanderpavlinov@yandex.ru*

METHOD FOR DEVELOPING AUDIO DESIGN FOR ADVERTISING USING GENERATIVE NEURAL NETWORKS

©Karshakova L., *SPIN-code: 2813-4762, ORCID ID: 0000-0003-2158-2508, Ph.D., The Kosygin State University of Russia (Technologies. Design. Art), Moscow, Russia, lkarshak@mail.ru*
©Nagay S., *ORCID: 0009-0005-5173-6688, The Kosygin State University of Russia (Technologies. Design. Art), Moscow, Russia, lana.nagaina@yandex.ru*
©Pavlinov A., *ORCID: 0009-0005-7717-9891, The Kosygin State University of Russia (Technologies. Design. Art), Moscow, Russia, aleksanderpavlinov@yandex.ru*

Аннотация. Рассматривается применение генеративных нейросетевых инструментов для звукового сопровождения рекламы с акцентом на русскоязычный контент. Проведен сравнительный анализ инструментов синтеза речи (YandexTTS, Play.ht) и музыки (Udio, Suno), выявлены их ключевые характеристики, преимущества и ограничения. Особое внимание уделено проблемам интеграции звуковых компонентов с видеорядом, включая вопросы синхронизации, баланса громкости и эмоционального соответствия. Предложен поэтапный алгоритм создания звукового сопровождения: от анализа видеоряда и подготовки текстовых промптов до генерации, сведения и тестирования финального аудиопродукта. Исследование выявило, что, несмотря на высокий потенциал генеративных технологий, сохраняются проблемы непредсказуемости результатов, особенно при создании музыкального фона. Предложенная методика подтверждает возможность автоматизации звукового дизайна в рекламе с необходимостью баланса между технологиями и творческим контролем. Результаты работы имеют практическое значение для специалистов в области звукового дизайна, маркетинга и цифрового производства, предлагая эффективный способ автоматизации создания аудиоконтента. Статья будет полезна исследователям, изучающим применение искусственного интеллекта в креативных индустриях.

Abstract. The paper discusses the application of generative neural network tools for sound accompaniment of advertising with a focus on Russian-language content. A comparative analysis of speech synthesis tools (YandexTTS, Play.ht) and music (Udio, Suno) is conducted, and their key characteristics, advantages, and limitations are identified. Special attention is paid to the challenges of integrating sound components with video content, including issues of synchronization, volume balance, and emotional matching. A step-by-step algorithm for creating sound accompaniment is

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

proposed, from analyzing the video content and preparing text prompts to generating, mixing, and testing the final audio product. The study revealed that, despite the high potential of generative technologies, there are still problems with the unpredictability of the results, especially when creating a musical background. The proposed methodology confirms the possibility of automating sound design in advertising, with the need for a balance between technology and creative control. The results of the work have practical significance for specialists in sound design, marketing, and digital production, offering an effective way to automate the creation of audio content. The paper will be useful for researchers studying the application of artificial intelligence in creative industries.

Ключевые слова: синтеза речи, генерация музыки, реклама, звуковой дизайн.

Keywords: speech synthesis, music generation, advertising, sound design.

Звуковое сопровождение в цифровой рекламе играет не вспомогательную, а структурообразующую роль. Звук обеспечивает эмоциональную динамику, фокусирует внимание на ключевых моментах и формирует общее восприятие сюжета. Особенности рекламных форматов предъявляют к звуку ряд специфических требований, отличных от тех, что характерны для кино, радио или игровых проектов. Элементы звукового дизайна — дикторская речь, фоновая музыка, звуковые эффекты — должны быть лаконичны, точны по времени и синхронизированы с визуальными акцентами [1].

Возрастающий интерес к технологиям автоматического формирования звукового сопровождения находит отражение в динамике их внедрения в профессиональные и повседневные практики [2].

За последние годы наблюдается устойчивый рост объёмов использования синтезированной речи и сгенерированной музыки в цифровой рекламе, презентационном контенте, аудиоформатах и социальных платформах. Количество пользователей, регулярно обращающихся к генеративным платформам для создания речи или фоновой музыки, увеличивается ежегодно. Это свидетельствует не только о расширении сферы применения технологий, но и о формировании устойчивого спроса. Актуальность темы определяется не только технологическим прогрессом, но и фактическим объёмом интереса к инструментам генеративного звука. Широкое распространение подобных решений подтверждает их значимость и необходимость системного анализа подходов к их практическому использованию.

Целью настоящей работы является разработка и апробация методики создания звукового сопровождения для рекламного видеоролика с использованием генеративных инструментов. Под звуковым сопровождением понимается совокупность дикторской речи, музыкального оформления и других звуковых элементов, синтезированных на основе текстовых и параметрических инструкций.

История применения вычислительных методов в создании звука началась в середине XX века с экспериментов в области алгоритмической композиции. Первые попытки автоматизации музыкального процесса базировались на правилах классической гармонии и математических закономерностях. В 1957 году Макс Мэтьюз в Bell Labs разработал первую цифровую звуковую синтезирующую программу — Music, положив начало направлению компьютерной генерации звука. В последующие десятилетия развивались технологии синтеза речи, основанные на фонемных моделях и формантных алгоритмах, которые применялись преимущественно в телекоммуникациях и системах озвучивания [3].

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

С переходом к обучающимся алгоритмам — в частности, к нейросетям — появилась возможность не только воспроизводить заранее записанные звуки, но и создавать новые звуковые фрагменты на основе статистических закономерностей [4].

Особый импульс развитию генеративного звука дали технологии глубинного обучения. В 2016 г корпорация Google представила WaveNet — первую нейросетевую архитектуру, способную синтезировать речь с высокой степенью естественности. Вслед за этим появились системы, способные воспроизводить не только дикторский текст, но и музыкальные композиции, стилизованные под конкретный жанр или автора. К началу 2020-х годов появились сервисы генерации звука по текстовому описанию, а также алгоритмы, позволяющие имитировать интонации, эмоции и ритмику живого выступления.

На сегодняшний день в сфере автоматизации звукового сопровождения наблюдается активное развитие использование генеративных нейросетей [5].

На международной арене сформировался ряд решений, позволяющих не просто синтезировать речь, но и формировать полностью новый звуковой материал на основе текстовых или параметрических описаний. К таким системам относятся разработки ElevenLabs, Suno, Udio и Murf AI, Play.ht. Эти сервисы позволяют генерировать дикторскую речь с заданной интонацией, стилизованные под конкретные жанры музыкальные композиции, а также моделировать эмоциональное звучание и варьировать голосовые параметры с высокой точностью. Принципиальным преимуществом данных платформ является возможность обучения на пользовательских голосах, включая генерацию уникальных тембров [6].

Отечественные разработки в этой области пока существенно уступают по функциональности. Наиболее заметные среди них — Yandex SpeechKit и SaluteSpeech от Сбер. Эти инструменты обеспечивают синтез речи на русском языке, допускают базовую настройку скорости, пауз и интонации, однако их архитектура основана на использовании базовой технологий синтеза речи на основе текста (Text-to-Speech, TTS). Такие системы опираются на заранее обученные голосовые модели и не обладают возможностью генерации новых голосов, создания уникальных тембров или синтеза музыкального сопровождения. Их использование представляется особенно актуальным в условиях ограничений на международные платформы и необходимости соблюдения нормативных требований при разработке цифрового продукта.

Для анализа был выбран один набор, полностью доступный из России; другой — ориентированный на международный сегмент (Таблица 1). Необходимым условием было наличие поддержки русского языка для дикторской озвучки, что исключало использование некоторых зарубежных решений с узкой языковой моделью. Помимо синтаксической корректности, проверялась также интонационная выразительность, отсутствие артефактов и соответствие стандартной дикторской манере подачи информации. Кроме того, принимались во внимание такие параметры, как простота интеграции, наличие бесплатного тарифа или возможности экспортировать аудиофайлы без звуковых меток, а также юридическая допустимость использования результатов генерации в образовательных и некоммерческих целях. На основе этих критериев были выбраны два комплекта инструментов, каждый из которых включает средство синтеза речи и средство генерации музыкального сопровождения.

Первый набор инструментов для генерации звукового сопровождения ориентирован на решения, полностью доступные из России без использования средств обхода блокировок. В качестве генератора речи выбран сервис YandexTTS, а в качестве инструмента музыкального сопровождения — Udio, как один из немногих генеративных музыкальных сервисов,

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

свободно работающих на территории РФ. При этом отмечается отсутствие на российском рынке отечественного программного обеспечения для нейросетевой генерации музыки.

Второй набор инструментов представляет собой альтернативное решение, ориентированное на международный рынок. В его состав входят два компонента: Play.ht для генерации дикторской речи и Suno для создания музыкального сопровождения. Play.ht — это облачный сервис синтеза речи, разработанный в США, активно применяемый в задачах конвертации текста в аудио. В своей архитектуре он использует современные нейросетевые модели, включая многослойные трансформеры с возможностью настройки акцента, интонации и скорости речи. В отличие от многих аналогов, Play.ht поддерживает разнообразие голосов, включая эмоциональные и профессиональные дикторские пресеты, а также предоставляет гибкий контроль над параметрами синтеза. Несмотря на англоязычную направленность сервиса, он обладает поддержкой русского языка, хотя и с некоторыми ограничениями по качеству. Главным преимуществом Play.ht стало качество воспроизведения: синтезированная речь отличается высокой естественностью, плавной интонацией и точной артикуляцией, что особенно важно в рекламных сценариях.

В качестве генератора музыкального сопровождения выбран Suno. Модель использует многослойные трансформерные архитектуры, обученные на крупномасштабных музыкальных датасетах, что позволяет синтезировать композиции высокого качества, адаптированные под заданную жанровую и эмоциональную специфику. Suno предоставляет возможность выбора длительности, стиля, инструментального состава и даже общего настроения композиции, что делает его удобным для построения звуковой среды в рекламных продуктах.

Таблица 1

СРАВНЕНИЕ ИНСТРУМЕНТОВ

Критерий	*YandexTTS (Россия)*	*Play.ht (Международный)*	*Udio (Музыка, доступен в РФ)*	*Suno (Музыка, международный)*
Тип контента	Синтез речи	Синтез речи	Генерация музыки	Генерация музыки
Поддержка русского	Да (оптимизировано)	Да (ограниченно)	Да (ограниченно)	Да (ограниченно)
Качество звучания	Высокая разборчивость, стандартные голоса	Высокая естественность, эмоциональные пресеты	Хорошее качество, жанровое соответствие	Высокое качество, богатые настройки
Настройки параметров	Скорость, паузы, базовые голоса	Акцент, интонация, эмоции, скорость	Жанр, темп, длина	Жанр, настроение, инструменты, темп
Интеграция	Веб-интерфейс, API для разработчиков	Облачный сервис, API	Веб-интерфейс	Веб-интерфейс
Юридические аспекты	Разрешено коммерческое использование	Требует проверки лицензии	Доступен в РФ	Ограничения для РФ
Пример применения	Рекламные ролики, озвучка текстов	Профессиональная дикторская озвучка	Фоновая музыка для видео	Сложные музыкальные композиции

Предпринятая параллельная реализация на базе двух решений позволяет выявить особенности интеграции звукового материала: от точности тайминга до сложности

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

редактирования и адаптации под видеоряд. Современные авторы рекомендуют начинать проект с референс-сцены, задающей общую звуковую палитру ролика. В проекте использовался заранее созданный видеоролик. Каждая сцена имела фиксированную длительность и была сгенерирована с применением инструментов автоматизированного видеодизайна, что исключало возможность последующего редактирования структуры или хронометража [7].

Первым шагом на этом этапе стало разбиение ролика на логически обособленные фрагменты. Каждая сцена рассматривалась как носитель отдельного визуального сообщения, требующего индивидуального подхода к речевому сопровождению. Одновременно с этим формировалась общая концепция музыкального оформления, призванного обеспечить целостность восприятия и связность между сценами, несмотря на их различное визуальное наполнение. Задачи, поставленные перед звуковым сопровождением, касались как технической, так и содержательной стороны. Озвучка должна была быть максимально лаконичной, информационно насыщенной и стилистически согласованной с видеорядом. Дикторский текст проектировался с учётом ограничений по длительности сцены и необходимости интонационного акцента на ключевых словах. Музыкальная подложка, в свою очередь, создавалась в виде единой композиции, сопровождающей весь ролик, и должна была поддерживать общий ритм и эмоциональное движение, не перегружая восприятие. Важной особенностью этапа стал также учёт внешних технических и платформенных ограничений. Предполагалось, что итоговое видео будет воспроизводиться преимущественно на мобильных устройствах или в браузерах социальных сетей, где качество акустической передачи ограничено. Это требовало соблюдения определённого баланса громкости, чёткости дикции и структурной ясности аудиоряда. Результатом данного этапа стало чёткое определение звуковой задачи для каждой сцены, выбор подхода к построению музыкального фона и формирование ограничений, в рамках которых должна была выстраиваться вся дальнейшая методика.

После анализа видеоряда и постановки целей следующим шагом стало формирование исходных данных для генерации. Звуковое сопровождение в проекте предполагает два ключевых компонента: дикторскую озвучку и музыкальную подложку. Каждый из них требует подготовки отдельного текстового входа, служащего основой для последующей генерации. Создание дикторского текста осуществлялось на основе смыслового наполнения каждой сцены. При этом учитывалась ограниченная продолжительность эпизодов, что обязывало к предельной лаконичности формулировок. Фразы подбирались таким образом, чтобы не превышать трёх-четырёх секунд звучания, при этом содержать в себе чёткое рекламное послание и передавать основную идею сцены. Дополнительно принималась во внимание стилистика речи — она должна была оставаться нейтральной, но выразительной, соответствовать ожиданиям целевой аудитории и быть интонационно приближенной к профессиональной дикторской подаче. Тексты проверялись на фонетическую простоту, отсутствие сложных или многосложных слов, способных затруднить синтез. Особое внимание уделялось ритмической структуре: избегались длинные паузы, сбивчивость и лексические конструкции, плохо воспринимаемые в устной форме. Все формулировки проходили этап редактуры, направленный на достижение максимальной чёткости произношения при использовании генеративных инструментов синтеза речи.

Параллельно с этим разрабатывался входной запрос для генерации музыкального сопровождения. В отличие от речевого фрагмента, музыкальная композиция создавалась как единое целое, охватывающее весь видеоролик. Описание содержало ключевые характеристики: жанр, эмоциональный вектор, темп, инструментальный состав.

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

Формулировки подбирались таким образом, чтобы быть интерпретируемыми генеративной моделью, но не допускать чрезмерной вариативности результатов. На данном этапе была сформирована полная языковая основа для последующей генерации аудиофайлов: текст дикторской озвучки для каждой сцены и описание музыкальной композиции как фонового трека ко всему ролику. Это обеспечило информационную чёткость на входе и позволило повысить точность и предсказуемость результатов синтеза.

После подготовки текстовых формулировок озвучки следующим этапом методики стала их генерация с использованием синтезаторов речи. Для достижения целевых параметров были выбраны два инструмента: YandexTTS в составе первого набора и Play.ht во втором. Эти генераторы обеспечивали преобразование текстовых данных в аудиофайлы с заданными характеристиками голоса, тембра, скорости и интонации. При этом особое внимание уделялось поддержке русского языка, так как дикторские фрагменты ролика изначально создавались для русскоязычной аудитории. Сначала происходила поочерёдная генерация всех пяти реплик — по одной на каждую сцену. Текст каждой реплики вводился в систему отдельно, с указанием желаемых параметров озвучивания. В YandexTTS на этом этапе использовалась настройка диктора, близкого к стандартному «универсальному» голосу с нейтральной интонацией. В Play.ht осуществлялся выбор из доступных голосов с русской фонетической поддержкой, в том числе с возможностью предварительного прослушивания и ручной регулировки скорости. Также был рассмотрен вариант с генерацией уникального тембра по заранее записанному образцу голоса.

Критерием успешной генерации служили: наличие чёткой дикции; интонационной выразительности; соблюдения допустимой продолжительности фразы; правильно расставленные ударения; соответствия синтезированной речи рекламному стилю подачи. Результаты, не отвечающие требованиям, отклонялись, а генерация повторялась с изменением параметров: например, варьировалась скорость или подбирался другой голос из доступного пула. Такое повторение могло потребоваться как из-за особенностей фонетики текста, так и из-за ограничений самого генератора.

В процессе работы с каждым инструментом велся каталог сгенерированных фрагментов, где указывались параметры, использованные при генерации, а также субъективная оценка качества. Это позволило не только осуществить выбор наиболее удачных версий, но и подготовить основу для последующего анализа качества TTS-инструментов в рамках сравнительного подхода. По завершении этапа был сформирован набор звуковых реплик, полностью соответствующих сценам видеоряда, готовых к интеграции с музыкальной подложкой. Эти файлы стали основой для последующего комбинирования в структуре итогового звукового сопровождения.

Этапы реализации проекта представлены в Таблице 2. Одной из ключевых особенностей звукового оформления проекта стала ориентация на единую музыкальную композицию, сопровождающую весь видеоряд. Музыкальный трек генерировался с использованием двух разных инструментов: Udio и Suno. Каждый из них применялся в составе своего набора, что позволило провести параллельную генерацию на основе идентичного текстового описания. Входной запрос (промт) для генерации включал указание жанровой направленности, желаемого настроения, а также продолжительности композиции. Кроме того, обращалось внимание на то, чтобы композиция не содержала ярко выраженных мелодических скачков, способных вступать в конкуренцию с дикторской речью. Генерация осуществлялась интерактивно: после отправки запроса производился анализ результата с точки зрения его совместимости с видеорядом и дикторской озвучкой. Особое внимание уделялось темпу и ритмическому рисунку, так как расхождения между аудио и визуальной

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

динамикой могли привести к искажённому восприятию сцен. Если полученный результат не удовлетворял требованиям композиция отправлялась на повторную генерацию с откорректированным описанием. Подобные итерации являлись важной частью методики, позволяющей достичь необходимого соответствия без вмешательства в структуру видео. Также учитывались технические параметры итогового аудиофайла — битрейт, формат, возможность бесшовного воспроизведения. Композиции, содержащие шумы или артефакты отклонялись. В ходе работы фиксировались характеристики каждого генеративного трека, что в дальнейшем позволило сравнивать не только художественную, но и технологическую сторону полученных результатов. На данном этапе была получена основная музыкальная подложка проекта — единая композиция, синхронизируемая с видеорядом и дикторскими репликами.

На следующем этапе сгенерированные звуковые компоненты — дикторские реплики и единая музыкальная композиция — объединяются в целостную звуковую структуру, синхронизируемую с видеорядом. Основная задача заключается в проверке совместимости компонентов между собой, а также в выявлении фрагментов, требующих корректировки.

При сведении аудио сначала производится наложение речевых реплик на музыкальную подложку. Для каждой сцены дикторская фраза располагается внутри отведённого временного интервала, при этом обеспечивается чёткость звучания на фоне музыки. На данном этапе важным аспектом становится звуковой баланс: дикторская речь должна быть разборчивой, но не агрессивной, а музыкальный фон — достаточно насыщенным, чтобы поддерживать эмоциональное восприятие, но не доминировать. В случае несоответствия этим требованиям производится повторная генерация одного из компонентов, так как существуют исследования, говорящие о нецелесообразности нейросетевого вмешательства в уже готовую запись из-за высокой вероятности усиления артефактов. Однако в рамках разработанной методики приоритетом считается повторная генерация именно музыкальной подложки, а не дикторской озвучки. Это обусловлено несколькими причинами. Во-первых, синтез речи в современных TTS системах даёт стабильно воспроизводимые результаты при условии корректно подобранного текста и параметров генерации. Во-вторых, даже незначительная вариация дикции может нарушить восприятие ключевых смыслов, особенно в условиях коротких сцен. В-третьих, корректировка речи требует изменения текста, что может повлечь нарушение логики рекламного посыла. В то же время музыка, будучи более гибким и вариативным материалом, допускает многократную генерацию без ущерба для структуры ролика. Изменения в ритме, тембре или эмоциональной окраске могут существенно улучшить синергетический эффект между визуальной и звуковой частями. На практике было установлено, что повторная генерация фоновой композиции требуется чаще всего при наличии следующих проблем: сбитый ритм, неправильный временной интервал, наличие синтетических артефактов, чрезмерная насыщенность или, наоборот, фоновая разреженность. Такие дефекты особенно критичны в условиях короткой продолжительности ролика, где каждая секунда звука оказывает значительное влияние на восприятие сцены. После сбора композиции проводилась проверка всей аудиодорожки в контексте видеоряда. Проверялась не только техническая сшивка, но и эмоциональное соответствие, логика сценического развития и отсутствие конфликтов между звуковыми слоями. При необходимости производилась дополнительная балансировка громкости и выравнивание пауз.

Описанная последовательность шагов представлена на схеме методики генерации звука на Рисунке 1, где каждый этап — от анализа видеоряда до окончательной сборки — включён в единую логическую структуру.

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

Таблица 2

ЭТАПЫ ГЕНЕРАЦИИ ЗВУКОВОГО СОПРОВОЖДЕНИЯ

Этап	*Действия*	*Инструменты*	*Критерии качества*
Анализ видеоряда	Разбивка на сцены, определение длительности и ключевых сообщений	Редакторы видео	Четкость структуры, логичность
Подготовка текстов	Написание дикторских реплик и описаний для музыки	Текстовые редакторы	Лаконичность, фонетическая простота
Генерация речи	Синтез дикторского текста с настройкой параметров	YandexTTS, Play.ht	Разборчивость, интонационная точность
Генерация музыки	Создание фоновой композиции по текстовому описанию	Udio, Suno	Жанровое соответствие, ритмичность
Сведение аудио	Наложение реплик на музыку, балансировка громкости	Редакторы звука Audacity, Adobe Audition	Синхронность, отсутствие конфликтов
Тестирование	Проверка на устройствах	Различные плееры	Отсутствие артефактов, гармоничность

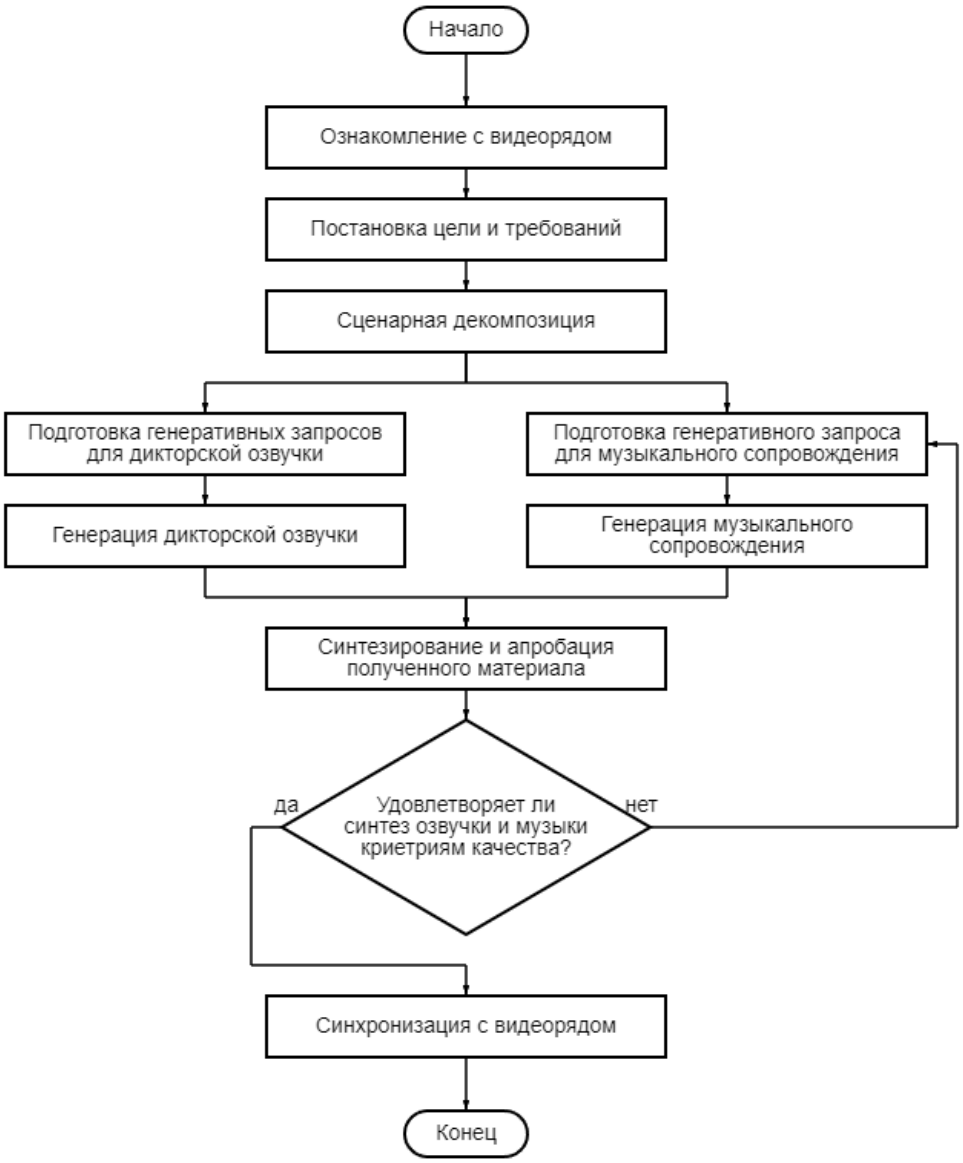

Рисунок 1. Схема генерации звукового сопровождения

Т. 11. №9 2025
https://doi.org/10.33619/2414-2948/118

а б

Рисунок 2. QR-коды на видеоролики со сгенерированным звуковым сопровождением: а — на основе YandexTTS и Udio; б — на основе Play.ht и Suno

Результаты и обсуждение

Несмотря на функциональную завершённость (Рисунок 2), предложенный подход обладает рядом ограничений и уязвимостей, требующих критической оценки. Одним из ключевых затруднений является высокая зависимость от стабильности и качества внешних генеративных систем. Платформы, используемые для синтеза речи и музыки, не всегда обеспечивают предсказуемый результат. Даже при одинаковых входных данных различные генерации могут отличаться по тембру, ритму и стилистике, что снижает повторяемость результатов и усложняет стандартизацию процесса. Особенно это касается генерации музыкального сопровождения, где отклонения от желаемого эмоционального настроя встречаются чаще всего. Как показала практика, для получения удовлетворительного результата зачастую требуется многократная генерация, сопровождаемая субъективной оценкой качества.

Дополнительные сложности возникают при необходимости точной синхронизации звуковых фрагментов с видеорядом. Несмотря на строгое соблюдение временных рамок, дикторская речь и музыка подчиняются своим ритмическим законам, что требует повышенного внимания при компоновке. Отсутствие возможности гибко управлять ритмом синтезированных звуков также ограничивает возможности тонкой настройки. Следует отметить и методологическую хрупкость этапа постановки задач. В условиях, когда видео не сопровождается заранее прописанным сценарием, формулировка задач для создания звукового ряда во многом носит субъективный характер. Это делает методику трудно формализуемой и слабо масштабируемой на другие проекты без участия квалифицированного специалиста.

Ещё одной уязвимой точкой методики является ограниченность доступных отечественных инструментов. При всей эффективности связки YandexTTS и Udio, последняя платформа формально не является отечественной, а YandexTTS — по своей сути не является полноценной генеративной моделью, приближаясь к синтетическому диктору со статичной базой голосов. Это ограничивает гибкость и адаптируемость при работе с нетипичными стилями и языковыми регистрами. В международных решениях, напротив, присутствует более широкий спектр параметров, но их использование сопряжено с юридическими, техническими и инфраструктурными рисками. Проблемы и возможные решения представлены в Таблице 3.

Исходя из вышеизложенного, несмотря на успешное применение методики в рамках конкретного проекта, она нуждается в дальнейшем развитии и адаптации для более широкой области задач. Проблемные зоны, выявленные в ходе реализации, служат основой для переосмысления архитектуры подхода, повышения его устойчивости и точности, а также выработки более гибких механизмов контроля качества на каждом этапе.

Таблица 3

ПРОБЛЕМЫ И РЕШЕНИЯ

Проблема	*Возможные решения*	*Пример*
Непредсказуемость результата	Многократная генерация с корректировкой параметров	Изменение текстового описания в Udio
Конфликты звуковых слоев	Приоритет повторной генерации музыки (не речи)	Настройка громкости в аудиоредакторе
Ограниченность отечественных инструментов	Комбинация YandexTTS с доступными международными сервисами	Использование Udio для музыки

Заключение

Разработанная и реализованная в рамках настоящей работы методика генерации звукового сопровождения демонстрирует значительный потенциал для практического применения в задачах, связанных с автоматизацией создания аудиоконтента для видеоформатов. Она включает в себя поэтапную структуру: от анализа видеоряда до синтеза звуковых компонентов с использованием генеративных инструментов, их сборки и финальной интеграции. Работа демонстрирует, что генеративные звуковые решения могут быть интегрированы в процессы цифрового медиапроизводства.

Список литературы:

1. Каршакова Л. Б. Потенциал звукового дизайна для виртуальных модных показов // Декоративное искусство и предметно-пространственная среда. Вестник РГХПУ им. С. Г. Строганова. 2024. №2–2. С. 146–155.

2. Абдуллаев У М. Методы и алгоритмы обработки звуковых сигналов // Бюллетень науки и практики. 2020. Т. 6. №6. С. 25-30. https://doi.org/10.33619/2414-2948/55/03

3. Мещеряков И. И. Основы звукового синтеза // Международный журнал гуманитарных и естественных наук. 2022. №11-1(74). С. 61-63. https://doi.org/10.24412/2500-1000-2022-11-1-61-63

4. Кудинов В. А., Водолад Д. В. Применение нейронных сетей для обработки звуковых сигналов в образовательном процессе // Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования. 2023. №4 (66). С. 67-77.

5. Семенюк В. В., Складчиков М. В. Разработка алгоритма распознавания эмоций человека с использованием сверточной нейронной сети на основе аудиоданных // Информатика. 2022. Т. 19. №4. С. 53-68. https://doi.org/10.37661/1816-0301-2022-19-4-53-68

6. Новикова П. А., Борзунов Г. И., Шиленко П. С., Тамашенко К. А. Информационное обеспечение автоматизированного проектирования аудиовизуальной среды // Дизайн. Материалы. Технология. Automation and control of technological processes and productions. 2025. №1(77). С. 208–213.

7. Каршакова Л. Б., Страшнов А. Ю., Нагай С. С., Павлинов А. М. Методика использования генеративных нейросетей для разработки рекламного видеодизайна // Бюллетень науки и практики. 2025. Т. 11. №8. С. 162-173. https://doi.org/10.33619/2414-2948/117/22

References:

1. Karshakova, L. B. (2024). Potentsial zvukovogo dizaina dlya virtual'nykh modnykh pokazov. *Dekorativnoe iskusstvo i predmetno-prostranstvennaya sreda. Vestnik RGKhPU im. S. G. Stroganova,* (2–2), 146–155. (in Russian).

2. Abdullayev, U. (2020). Methods and Algorithms Sound Signals Processing. *Bulletin of Science and Practice, 6*(6), 25-30. (in Russian). https://doi.org/10.33619/2414-2948/55/03

3. Meshcheryakov, I. I. (2022). Osnovy zvukovogo sinteza. *Mezhdunarodnyi zhurnal gumanitarnykh i estestvennykh nauk, (*11-1(74)), 61-63. (in Russian). https://doi.org/10.24412/2500-1000-2022-11-1-61-63

4. Kudinov, V. A., & Vodolad, D. V. (2023). Primenenie neironnykh setei dlya obrabotki zvukovykh signalov v obrazovatel'nom protsesse. *Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Seriya: Informatika i informatizatsiya obrazovaniya,* (4 (66)), 67-77. (in Russian).

5. Semenyuk, V. V., & Skladchikov, M. V. (2022). Razrabotka algoritma raspoznavaniya emotsii cheloveka s ispol'zovaniem svertochnoi neironnoi seti na osnove audiodannykh. *Informatika, 19*(4), 53-68. (in Russian). https://doi.org/10.37661/1816-0301-2022-19-4-53-68

6. Novikova, P. A., Borzunov, G. I., Shilenko, P. S., & Tamashenko, K. A. (2025). Informatsionnoe obespechenie avtomatizirovannogo proektirovaniya audiovizual'noi sredy. *Dizain. Materialy. Tekhnologiya. Automation and control of technological processes and productions,* (1(77)), 208–213. (in Russian).

7. Karshakova, L., Strashnov, A., Nagay, S., & Pavlinov, A. (2025). Methodology of using Generative Neural Networks for Developing Advertising Video Design. *Bulletin of Science and Practice, 11*(8), 162-173. (in Russian). https://doi.org/10.33619/2414-2948/117/22

Работа поступила
в редакцию 22.07.2025 г.

Принята к публикации
28.07.2025 г.

Ссылка для цитирования:

Каршакова Л. Б., Нагай С. С., Павлинов А. М. Методика разработки аудиодизайна для рекламы средствами генеративных нейросетей // Бюллетень науки и практики. 2025. Т. 11. №9. С. 138-148. https://doi.org/10.33619/2414-2948/118/14

Cite as (APA):

Karshakova, L., Nagay, S., & Pavlinov, A. (2025). Method for Developing Audio Design for Advertising using Generative Neural Networks. *Bulletin of Science and Practice, 11*(9), 138-148. (in Russian). https://doi.org/10.33619/2414-2948/118/14