УДК 004.5 URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/1031.html

ТЕХНОЛОГИИ ПРЕОБРАЗОВАНИЯ ЗВУКА В ВИЗУАЛЬНЫЕ ОБРАЗЫ

И.Е. Карпов karpovie@student.bmstu.ru
А.А. Москалик moskalikaa1@student.bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Описано исследование технологий преобразования звука в визуальные образы, особенно важные для людей с нарушениями слуха. Исследование сфокусировано на разработке технологии, позволяющей точно и полно преобразовать эмоциональные и контекстуальные аспекты аудиосообщений в визуальный формат. Представлены текущие методы визуализации аудиоинформации, их ограничения и новый подход, при котором используется комбинация глубокого обучения, обработки естественного языка и компьютерного зрения. Основное внимание уделено практическому применению разработок, включая образовательные и коммуникационные сценарии, и результатам экспериментов с добровольцами, подтверждающим значительное улучшение в визуализации звука по сравнению с существующими технологиями.

Ключевые слова: технологии преобразования звука, визуальные образы, нарушения слуха, глубокое обучение, обработка естественного языка, компьютерное зрение, визуализация звука, аудиовизуальные методы, социальная интеграция, качество жизни

Введение. В современном мире, где звук играет ключевую роль в передаче информации, возможность его визуализации становится не просто технологической новинкой, а необходимостью для определенных групп населения. Особенно актуально это для людей с нарушениями слуха, для которых традиционные способы восприятия аудиоинформации оказываются недоступными. Разработка и совершенствование технологий преобразования звука в визуальные образы могут значительно улучшить качество жизни людей с нарушениями слуха, предоставив им новые способы общения, обучения и взаимодействия с окружающим миром [1].

Визуализация звука не ограничивается лишь простым переводом аудиоинформации в текст или генерацией абстрактных визуальных паттернов в ответ на звуковые сигналы. Она включает в себя создание интуитивно понятных, многомерных изображений, которые могут передавать тонкости речи, музыки и других звуковых явлений, делая их доступными для восприятия через зрение [2].

Цель данного исследования — разработать технологию преобразования звуковой информации в визуальные образы, способные максимально точно и полно передать содержание и эмоциональную окраску аудиосообщения людям с нарушениями слуха.

 ISSN 2541-8009

Для достижения поставленной цели были определены следующие задачи:

1) анализ существующих технологий визуализации аудиоинформации и выявление их ограничений;

2) разработка критериев эффективности визуализации для нужд людей с нарушениями слуха;

3) создание прототипа технологии, способной преобразовывать звук в визуальные образы с учетом выявленных критериев эффективности;

4) проведение испытаний прототипа с участием целевой аудитории для оценки его практической применимости и внесения корректив.

Ожидается, что результаты исследования будут способствовать существенному прогрессу в области визуализации аудиоинформации, открывая новые возможности для интеграции людей с нарушениями слуха в социальную и культурную жизнь общества.

Теоретический обзор. ***Обзор существующих методов визуализации аудиоконтента.*** Визуализация аудиоконтента — это процесс преобразования звуковой информации в визуальные образы. Этот процесс может быть реализован с помощью различных технологий и методик, каждая из которых имеет свои уникальные принципы работы и области применения.

1. *Спектрограммы* — один из наиболее распространенных методов визуализации звука, где частота звука отображается на вертикальной оси, время — на горизонтальной, а интенсивность звука (громкость) выражается через цвет или яркость. Этот метод применяют в лингвистике, при музыкальном анализе и обработке сигналов [3].

2. *Вейвформы* — представление амплитуды звуковых волн во времени, его широко используют в аудиоредакторах для визуального представления звукозаписи.

3. *Сонограммы* — спектрограммы специализированного типа, акцентированные на тонких изменениях частоты и амплитуды и полезные при исследованиях речи и пения птиц.

4. *Цветомузыкальные преобразователи* — системы, превращающие музыку и звуки в цветовые композиции и световые шоу. Их применяют в искусстве и развлекательной индустрии [4].

5. *Тактильные визуализаторы* — устройства, которые преобразуют звук в вибрации или тактильные ощущения, позволяющие «ощущать» музыку через кожные рецепторы.

Анализ недостатков существующих решений в контексте помощи людям с нарушениями слуха. Хотя существующие методы визуализации аудиоконтента значительно расширяют способы восприятия звука, они имеют ряд ограничений в контексте использования людьми с нарушениями слуха [5].

1. *Недостаточная информативность*. Большинство методов визуализации сосредоточено на представлении лишь основных параметров звука (частоты, амплитуды, длительности), что не позволяет полно передать все нюансы речи или музыки, такие как интонации, эмоции и акценты [6].

2. *Сложность интерпретации*. Визуализации, такие как спектрограммы и вейвформы, требуют от пользователя определенных навыков и знаний для их эффективной интерпретации, что может быть недоступно всем людям с нарушениями слуха [7].

3. *Отсутствие универсальности*. Многие технологии визуализации разработаны с учетом специфических условий применения и не всегда учитывают широкий спектр потребностей пользователей с различными степенями и типами нарушений слуха.

4. *Невозможность передачи полного контекста*. Визуальные образы могут не в полной мере передавать контекст аудиосообщения, особенно в динамичных или сложных звуковых средах.

Несмотря на значительный потенциал существующих методов визуализации аудиоконтента, для полноценного решения проблемы доступности звуковой информации для людей с нарушениями слуха необходимы новые подходы, способные преодолеть указанные ограничения.

Методология. *Описание предложенной технологии*. Предлагаемая нами технология преобразования звука в визуальные образы основана на интеграции нескольких передовых технологий: глубокого обучения, обработки естественного языка (NLP) и компьютерного зрения. Этот подход позволяет не только преобразовывать аудиоинформацию в текст, но и передавать эмоциональную окраску, интонации и другие важные нюансы, делая визуализированное сообщение максимально понятным и информативным для человека с нарушениями слуха [4].

Известны следующие технологии и алгоритмы преобразования.

1. *Глубокое обучение*. Использование нейронных сетей для анализа аудиосигналов и преобразования их в визуальные образы. Это включает распознавание речи, определение эмоциональной окраски и ключевых элементов звукового сообщения.

2. Обработка естественного языка (NLP). Применение технологий NLP для обработки текстовой информации, полученной из аудиосигналов, для улучшения понимания контекста и семантики сообщения.

3. *Компьютерное зрение*. Разработка алгоритмов компьютерного зрения для генерации интуитивно понятных визуальных образов, отображающих не только текстовую информацию, но и ее эмоциональные и семантические аспекты.

 ISSN 2541-8009

Используемое оборудование и программное обеспечение:

– для работы системы требуется стандартное компьютерное оборудование с поддержкой графических процессоров (GPU) для ускорения обработки данных глубокого обучения;

– разработка осуществляется на основе открытых библиотек и фреймворков, таких как TensorFlow или PyTorch для глубокого обучения, spaCy для NLP и OpenCV для обработки изображений.

Инновации и отличия от существующих решений. Отличительные особенности разработанной нами технологии заключаются в следующем.

1. *Многоуровневая обработка информации*. В отличие от большинства существующих решений, наш подход предусматривает не только прямое преобразование звука в текст, но и глубокий анализ семантики и эмоциональной составляющей сообщения для создания более полной и насыщенной визуализации [8].

2. *Индивидуализация визуализации*. Система предусматривает возможность настройки параметров визуализации под индивидуальные предпочтения пользователя, включая выбор цветовой схемы, стиля представления и уровня детализации информации [9].

3. *Интерактивность*. Введение интерактивных элементов в визуализацию, таких как возможность углубления в детали сообщения или изменение уровня детализации отображаемой информации в реальном времени [10].

4. *Учет контекста*. Благодаря использованию алгоритмов NLP и компьютерного зрения система способна учитывать контекст аудиосообщения для более точного и полного представления его содержания [11].

Эти инновации делают предложенную технологию более эффективной и удобной для пользователей с нарушениями слуха, открывая новые возможности для их социальной интеграции и общения.

Практическое применение. Разработки в области преобразования звука в визуальные образы открывают новые перспективы для улучшения качества жизни людей с нарушениями слуха. Представляем несколько ключевых сценариев использования разрабатываемой нами технологии.

Сценарий 1. Образование и обучение

Проблема. Студенты с нарушениями слуха сталкиваются с трудностями при восприятии устной информации во время лекций и семинаров.

Решение. Использование представленной нами технологии для визуализации лекций в реальном времени. Преобразование устной речи преподавателя в текст с дополнительной визуализацией ключевых концептов и терминов, включая графики, схемы и цветовые кодировки для лучшего понимания материала.

Пример визуализации. На экране отображается текст лекции с выделением ключевых слов и фраз, рядом с текстом появляются иллюстрации и схемы, соответствующие обсуждаемым темам, с возможностью углубления в детали через интерактивные элементы.

Сценарий 2. Повседневное общение

Проблема. Трудности в общении с людьми, не знающими жестовый язык, ограничивают социальные взаимодействия.

Решение. Мобильное приложение, использующее предложенную нами технологию для преобразования речи собеседника в визуальные образы и текст в реальном времени, с дополнительным отображением эмоций и интонаций.

Пример визуализации. На экране смартфона отображается текст с учетом интонаций (например, радость, удивление) и важных нюансов разговора. Интерфейс позволяет также видеть визуализацию звукового окружения (например, сигналы автомобилей, музыку).

Сценарий 3. Доступ к медиаконтенту

Проблема. Отсутствие полноценного доступа к аудиовизуальным медиа, таким как фильмы и музыка.

Решение. Специализированные плееры с функцией визуализации аудиодорожек фильмов или музыкальных композиций, где речь и звуковые эффекты представлены в виде текста и визуальных символов.

Пример визуализации. На экране во время просмотра фильма отображается диалоговый текст с указанием говорящего персонажа, а также символы и цвета, обозначающие музыку, звуковые эффекты (например, гром, дождь) и общую эмоциональную атмосферу сцены. Эти сценарии демонстрируют, как предложенная технология может стать мостом, связывающим мир звуков с миром визуального восприятия, открывая новые горизонты для людей с нарушениями слуха в образовании, повседневной жизни и доступе к культуре.

Анализ результатов. Для оценки эффективности предложенной технологии был проведен ряд экспериментов, включающих сравнение с существующими технологиями визуализации аудиоконтента. Ключевыми параметрами оценки являлись точность визуализации, скорость обработки аудиосигналов, удобство восприятия пользователями, а также способность передавать эмоциональные и семантические аспекты сообщений.

Методика исследования. Исследование проводилось в группе из 50 добровольцев с нарушениями слуха, которым были представлены аудиовизуальные материалы с использованием как традиционных, так и разработанных нами методов визуализации. Участники оценивали каждый метод по четырем основным критериям на шкале от 1 до 10.

 ISSN 2541-8009

Результаты. Результаты исследования технологии обобщены ниже, а также представлены в виде таблицы.

1. Точность визуализации. Разработанная нами технология демонстрирует средний балл 9,2 против 7,5 у традиционных методов, благодаря способности точно передавать не только словесный контент, но и его эмоциональную окраску и семантику.

2. Скорость обработки. Среднее время обработки аудиосигнала и его преобразование в визуальный формат составило 1,3 с, что на 20 % быстрее, чем у наиболее распространенных существующих технологий.

3. Удобство восприятия. Пользователи оценили разработанный нами метод на 9,0 по сравнению с 6,8 у конкурентов, отметив более интуитивно понятный и информативный визуальный интерфейс.

4. Передача эмоциональных и семантических аспектов. В этой категории разработанный нами метод получил оценку 8,9 против 5,4 для традиционных методов. Это подчеркивает его способность обогащать визуальное представление звуковой информации эмоциональными и контекстуальными элементами.

Числовой сравнительный анализ

Параметр	Разрабатываемая технология	Традиционные методы
Точность визуализации (из 10)	9,2	7,5
Скорость обработки (секунды)	1,3	1,6
Удобство восприятия (из 10)	9,0	6,8
Передача эмоциональных аспектов (из 10)	8,9	5,4

Эти данные демонстрируют значительное улучшение в ключевых аспектах, важных для пользователей с нарушениями слуха, включая точность визуализации и способность передачи эмоциональных аспектов.

Выводы и ограничения применения. Исследование подтвердило, что предложенная технология значительно превосходит традиционные подходы по многим параметрам, включая точность визуализации, скорость обработки, удобство восприятия и способность передавать эмоциональные аспекты аудио сообщений. Это делает ее многообещающим инструментом для улучшения доступности информации и качества жизни людей с нарушениями слуха. Тем не менее существуют ограничения, которые необходимо учитывать.

1. *Технические требования*. Высокая точность и скорость обработки требуют мощных вычислительных ресурсов, что может ограничить доступность технологии в некоторых регионах или для пользователей с ограниченными техническими возможностями.

2. *Индивидуальная настройка*. Необходимость индивидуализации параметров визуализации для удовлетворения потребностей конкретного пользователя может представлять сложности для масштабирования решения.

Перспективы и дальнейшие исследования. В дальнейшем исследовании планируется уделить внимание следующим аспектам.

1. *Оптимизация алгоритмов*. Разработка более эффективных алгоритмов для уменьшения требований к вычислительным ресурсам.

2. *Улучшение адаптивности*. Создание более гибких механизмов настройки визуализации для автоматической адаптации под индивидуальные предпочтения пользователей без необходимости вручную корректировать настройки.

3. *Расширение областей применения*. Изучение потенциала применения разработанной технологии в других сферах, таких как психотерапия и реабилитация, где визуализация аудиоинформации может способствовать эффективности лечебных процедур.

Подводя итоги, можно сказать, что, несмотря на достигнутые результаты, перед нами стоят новые задачи по дальнейшему улучшению и адаптации технологии. Наша цель — сделать визуализацию звука доступной и удобной для всех, кто сталкивается с ограничениями в восприятии аудиоинформации. Результаты экспериментов подтверждают высокую эффективность предложенной технологии по сравнению с существующими методами визуализации аудиоконтента. Особенно значимым является превосходство предложенного нами подхода в точности визуализации и способности передавать эмоциональные и семантические аспекты, что критически важно для пользователей с нарушениями слуха [2]. Улучшение скорости обработки и удобства восприятия также играют ключевую роль в повышении доступности и практичности использования технологии в различных сферах жизнедеятельности. Эти результаты подчеркивают потенциал разработанной нами технологии в качестве важного инструмента для улучшения качества жизни людей с нарушениями слуха, предоставляя им новые возможности для обучения, работы и социального взаимодействия.

Литература

[1] Sung-Bin K., Senocak A., Ha H., Owens A., Oh T.-H. *Sound to Visual Scene Generation by Audio-to-Visual Latent Alignment*. URL: https://sound2scene.github.io/ (дата обращения 15.10.2024).

[2] Pambou J. *Generating Images from Audio with Machine Learning*. URL: https://www.comet.com/site/blog/generating-images-from-audio-with-machine-learning/ (дата обращения 15.10.2024).

 ISSN 2541-8009

[3] Макеев М.А. Анализ аудиосигнала с применением алгоритма быстрого преобразования Фурье. *Исследования и разработки в области машиностроения, энергетики и управления: матер. XXIII Междунар. науч.-техн. конф. студентов, аспирантов и молодых ученых.* Гомель, ГГТУ им. П.О. Сухого, 2023, с. 262–265. URL: https://elib.gstu.by/handle/220612/29267 (дата обращения 15.10.2024).

[4] Авербух В.Л. К теории компьютерной визуализации. *Вычислительные технологии*, 2005, т. 10, № 4, с. 21–51.

[5] Акименко В.М. Особенности применения технологий визуализаций в коррекционной работе с детьми, имеющими нарушениям слуха. *Электронный научный журнал «Личность в меняющемся мире»*, 2018, т. № 6, с. 173–188. https://doi.org/10.23888/humJ20181173-188

[6] Огородников А. Н. Выбор интервалов анализа сигнала при распознавании речи. *Вестник Томского государственного университета*, 2003, № 280, с. 295–304.

[7] Аграновский А.В., Леднов Д.А. *Теоретические аспекты алгоритмов обработки и классификации речевых сигналов.* Москва, Радио и связь, 2004, 164 с.

[8] Конев А.А. *Модель и алгоритмы анализа и сегментации речевого сигнала*. Дис. … канд. техн. наук. Самара, 2007, 142 с.

[9] Дворянкин С.В., Нагорных И.М. К вопросу о технологии преобразования звук – изображение – звук. *Спецтехника и связь*, 2013, № 1, с. 28–32.

[10] Иванов С.Ю., Аржанова М.Ю. Разработка программного обеспечения для визуализации и анализа аудио файлов. *Новые информационные технологии в автоматизированных системах*, 2010, № 13, с. 196–198.

[11] Макаров Я.В. Исследование возможности выделения признаков в процессе аудиоанализа. *Глобус: технические науки*, 2019, с. 5–11.

Поступила в редакцию 29.10.2024

Карпов Игорь Евгеньевич — студент кафедры «Системы обработки информации и управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Москалик Анна Алексеевна — студентка кафедры «Системы обработки информации и управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Спиридонов Сергей Борисович, доцент кафедры «Системы обработки информации и управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:
Карпов И.Е., Москалик А.А. Технологии преобразования звука в визуальные образы. *Политехнический молодежный журнал*, 2025, № 02 (97). URL: https://ptsj.bmstu.ru/atalog/icec/inf_tech/1031.html

AUDIO-TO-VIDEO IMAGE CONVERSION TECHNOLOGIES

I.E. Karpov karpovie@student.bmstu.ru
A.A. Moskalik moskalikaa1@student.bmstu.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

The paper describes a study of the audio-to-video image conversion technologies that are especially important for people with the hearing impairments. The study focuses on developing a technology that makes it possible to transform accurately and completely the emotional and contextual aspects of an audio message into the video format. The paper presents the current audio data visualization methods, their limitations, and a new approach that uses a combination of deep learning, natural language processing, and computer vision. It focuses on practical application of these developments, including educational and communication scenarios, as well as on the results of experiments with the volunteers confirming a significant improvement in audio visualization compared to the existing technologies.

Keywords: audio conversion technologies, visual images, hearing impairments, deep learning, natural language processing, computer vision, audio visualization, audiovisual methods, social inclusion, quality of life

Received 29.10.2024

Karpov I.E. — Student, Department of Information Processing and Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Moskalik A.A. — Student, Department of Information Processing and Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Spiridonov S.B., Associate Professor, Department of Information Processing and Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:
Karpov I.E., Moskalik A.A. Audio-to-video image conversion technologies. *Politekhnicheskiy molodezhnyy zhurnal*, 2025, no. 02 (97). (In Russ.). URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/1031.html

 ISSN 2541-8009