Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет информатики и радиоэлектроники

УДК 004.93

Каминский
Александр Викторович

ОБЛАЧНАЯ СИСТЕМА РАСПОЗНАВАНИЯ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ ПО ВОКАЛЬНОМУ ЗАПРОСУ

ДИССЕРТАЦИЯ
на соискание степени магистра

по специальности 7-06-0611-05 – Компьютерная инженерия. Хранение и обработка данных

_____

Научный руководитель:
Петровский Николай
Александрович
к.т.н., доцент каф. Встраиваемых
вычислительных систем БГУИР

_____

Минск 2026

ВВЕДЕНИЕ

Методы цифровой обработки сигналов, машинного обучения и интеллектуального анализа данных находят широкое применение в задачах распознавания речи, обработки изображений и рекомендационных систем. Особую актуальность приобретают технологии, позволяющие извлекать смысловую информацию из неструктурированных данных и обеспечивать удобное взаимодействие пользователя с цифровыми сервисами.

Одним из направлений развития таких технологий является автоматизация поиска мультимедийного контента. В условиях постоянного увеличения объёма цифровых музыкальных коллекций традиционные методы поиска по названию произведения, исполнителю или тексту становятся недостаточными. Даже системы распознавания музыки по аудиофрагменту, такие как Shazam, предполагают наличие оригинальной записи. На практике же пользователи нередко помнят лишь мелодию композиции, не обладая сведениями о её названии или авторе. В подобных ситуациях возникает задача поиска музыкального произведения по вокальному запросу, известная как Query by Singing or Humming (QbSH), при которой пользователь напевает фрагмент мелодии, а система должна определить соответствующую композицию в базе данных.

Сложность решения данной задачи обусловлена вариативностью человеческого исполнения. Напев может отличаться по темпу, тональности, точности интонирования и длительности звучания отдельных нот. Кроме того, вокальный сигнал подвержен шумам и акустическим искажениям. В результате система поиска должна быть устойчива к временным и тональным искажениям, а также обеспечивать высокую скорость обработки запросов при работе с большими музыкальными базами данных.

Современные исследования в области поиска по напеву используют методы частотного анализа, хроматические представления, алгоритмы динамического выравнивания по времени и нейросетевые модели формирования эмбеддингов. Однако проблема компромисса между точностью сопоставления и вычислительной эффективностью остаётся актуальной, особенно при практической реализации интерактивных систем.

В рамках данной магистерской работы рассматривается задача разработки архитектуры системы поиска мелодий по вокальному запросу, сочетающей методы компактного векторного представления мелодий и алгоритмы точного временного выравнивания.

Целью работы является разработка и исследование программной системы поиска музыкальных произведений по вокальному запросу пользователя.

Для достижения поставленной цели в работе необходимо решить следующие задачи:

1. Провести анализ существующих методов поиска мелодий по вокальному запросу и определить требования к разрабатываемой системе.

2. Обосновать выбор методов предварительной обработки аудиосигнала и способов представления мелодий.

3. Разработать архитектуру системы поиска мелодий по вокальному запросу с применением компактных векторных представлений и алгоритма динамического выравнивания по времени для уточнения результатов поиска.

4. Реализовать программную систему поиска мелодий по вокальному запросу.

5. Провести экспериментальное исследование точности и вычислительной эффективности предложенного подхода.

Таким образом, работа направлена на разработку прикладного решения задачи поиска музыкальных произведений по напеву с учётом требований устойчивости, масштабируемости и практической применимости.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью исследования является разработка и научное обоснование метода поиска музыкальных произведений по вокальному запросу, который обеспечивает преобразование голосового сигнала в формализованное представление и последующий поиск соответствующей мелодии.

Актуальность разработки программного средства для распознавания мелодии по вокальному запросу обусловлена стремительным ростом объёма цифрового музыкального контента и необходимостью создания эффективных инструментов поиска музыкальных произведений. Такие системы позволяют осуществлять поиск по напетой мелодии, что особенно важно в случаях, когда пользователю неизвестны название произведения или его исполнитель.

Разработка данного программного средства способствует повышению точности и удобства музыкального поиска как в научных исследованиях, так и в прикладных мультимедийных сервисах.

Для достижения поставленной цели необходимо решить следующие задачи:

– изучить существующие подходы и алгоритмы распознавания мелодий по вокальному запросу, включающие методы обработки и анализа аудиосигналов;

– проанализировать методы выделения мелодической линии из вокального сигнала и разработать способ представления мелодии в виде формализованной последовательности признаков;

– разработать алгоритм сопоставления вокального запроса пользователя с музыкальными произведениями в базе данных;

– спроектировать и реализовать программное средство для распознавания мелодии по вокальному запросу и провести его экспериментальную оценку.

Объектом исследования являются музыкальные сигналы (аудиоданные).

Предметом исследования являются алгоритмы и методы обработки музыкальных сигналов, а также способы их представления для последующего анализа и поиска музыкальных фрагментов.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа соответствует приоритетному направлению научной, научно-технической и инновационной деятельности Республики Беларусь «Цифровые технологии и искусственный интеллект», включая подпункты «технологии искусственного интеллекта», «математические теории и методы, высокопроизводительные вычислительные системы, цифровые модели и двойники».

Личный вклад соискателя

Результаты, представленные в диссертации, а также положения, выносимые на защиту, получены автором самостоятельно. Роль научного руководителя, доцента Н. А. Петровского, заключалась в формулировании цели и задач исследования, определении возможных направлений их решения и обсуждении результатов, полученных соискателем. В совместных публикациях вклад автора определяется содержанием результатов, изложенных в диссертации.

Апробация диссертации и информация об использовании результатов

Основные результаты диссертационной работы докладывались и обсуждались на двух конференциях: XV Международная научная конференция «Информационные технологии и системы» (Минск, БГУИР, 2025), 62-я научная конференция аспирантов, магистрантов и студентов БГУИР (Минск, БГУИР, 2026).

Опубликование результатов диссертации

По теме диссертации опубликована 2 печатные работы в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертационная работа состоит из введения, общей характеристики исследования, трёх глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен аналитический обзор существующих методов и алгоритмов поиска мелодий по вокальному запросу. Рассматриваются основные этапы решения задачи Query by Humming, включая методы частотного анализа, извлечения музыкальных признаков, выравнивания временных последовательностей и сравнения мелодий. Проанализированы как классические подходы, такие как динамическое выравнивание по времени и скрытые марковские модели, так и современные нейросетевые методы формирования векторных представлений мелодий. На основе проведённого анализа определены ключевые проблемы и сформулированы требования к системе поиска, которые используются при дальнейшем проектировании архитектуры и выборе алгоритмических решений. Во второй главе рассматриваются архитектурные и алгоритмические решения, лежащие в основе разрабатываемого метода поиска мелодий по вокальному запросу. Обосновывается выбор общей структуры конвейера поиска, методов предварительной обработки аудиосигнала и способов представления мелодий в виде компактных векторных описаний. Особое внимание уделено двухэтапному подходу к поиску, который сочетает быстрый поиск в пространстве эмбеддингов с последующим уточнением результатов с использованием алгоритма динамического выравнивания по времени. Представленные решения

формируют концептуальную основу для последующей программной реализации и экспериментальной оценки эффективности системы. В третьей главе описывается программная реализация предложенного метода системы поиска мелодий по вокальному запросу, а также результаты экспериментального исследования её эффективности. Рассматривается структура программных модулей, применяемые технологии и параметры обработки аудиосигналов, а также особенности реализации поиска в пространстве эмбеддингов и алгоритма динамического выравнивания по времени. Приводятся результаты тестирования метода на наборе музыкальных данных, проводится анализ точности поиска и вычислительной эффективности предложенного подхода.

Проверка на уникальность

Проведена экспертиза диссертационной работы на корректность использования заимствованных материалов с применением сетевого ресурса «Антиплагиат» (адрес доступа: https://antiplagiat.ru) в on-line режиме.

В результате проверки установлена корректность использования заимствованных материалов. Оригинальность диссертационной работы составляет 97,68%.

ОСНОВНОЕ СОДЕРЖАНИЕ

В первой главе проводится аналитический обзор существующих подходов и алгоритмов поиска мелодий по вокальному запросу. Рассматриваются основные этапы решения задачи Query by Humming, включая методы частотного анализа, извлечения музыкальных признаков, выравнивания временных последовательностей и сравнения мелодий. Анализируются классические алгоритмы, такие как динамическое выравнивание по времени и скрытые марковские модели, а также современные нейросетевые методы формирования векторных представлений мелодий. На основе проведённого обзора формулируются ключевые проблемы и требования к системе поиска, которые используются в дальнейшем при проектировании архитектуры и выборе алгоритмических решений.

Во второй главе рассматривается построение конвейера поиска мелодий по вокальному запросу. Описываются и обосновываются основные этапы предлагаемого подхода: формирование признакового представления мелодии, получение эмбеддингов, быстрый отбор кандидатов и последующее сопоставление музыкальных фрагментов. Центральное место занимает двухэтапная схема поиска, в которой извлечение ближайших кандидатов в пространстве эмбеддингов дополняется повторным ранжированием с использованием алгоритма динамического выравнивания по времени. Предложенные решения задают основу исследуемого метода и используются далее при его практической реализации и экспериментальной оценке.

В третьей главе проводится экспериментальная оценка предложенного метода поиска мелодий по вокальному запросу. Рассматривается практическая реализация ключевых этапов обработки аудиосигналов, извлечения эмбеддингов, поиска ближайших кандидатов и их повторного ранжирования с использованием динамического выравнивания по времени. Основное внимание уделяется метрикам качества и производительности, позволяющим оценить эффективность разработанной архитектуры применительно к задаче поиска музыкальных фрагментов. На основе проведённых экспериментов анализируются точность поиска, вычислительные затраты и влияние отдельных этапов метода на итоговые результаты.

ЗАКЛЮЧЕНИЕ

В ходе выполнения магистерской диссертации были исследованы существующие подходы к решению задачи поиска мелодий по вокальному запросу (Query by Humming), включая методы, основанные на преобразовании Фурье, хроматических признаках, алгоритме динамического выравнивания по времени (DTW) и нейросетевых архитектурах. Анализ показал, что классические методы обладают высокой интерпретируемостью и устойчивостью при работе с небольшими базами данных, однако их вычислительная сложность и ограниченная способность к обобщению затрудняют применение в современных масштабируемых музыкальных системах.

В рамках работы был разработан и реализован программный комплекс для поиска мелодий по вокальному запросу с использованием нейросетевого подхода. В качестве общего представления MIDI-мелодий и пользовательских напевов использовались хроматические признаки, позволяющие привести различные типы музыкальных данных к единому формату. Для извлечения эмбеддингов была реализована модель на основе архитектуры ResNet18, обучаемая с применением контрастивной функции потерь InfoNCE.

В процессе исследования были реализованы:

– алгоритмы преобразования MIDI- и аудиоданных в хрома-представление;

– методы аугментации и подготовки обучающих данных;

– механизм формирования эмбеддингов для музыкальных последовательностей;

– поиск похожих мелодий в векторном пространстве;

– дополнительное ранжирование результатов с использованием DTW.

Проведённые эксперименты показали, что использование эмбеддингов позволяет существенно сократить вычислительные затраты по сравнению с полным попарным DTW-сравнением всех композиций базы данных. При этом дополнительное применение DTW к ограниченному числу кандидатов повышает качество ранжирования итоговых результатов поиска.

В ходе экспериментальных исследований были получены значения метрик Accuracy и Mean Reciprocal Rank (MRR), подтверждающие работоспособность предложенного подхода. Также был выполнен анализ вычислительных затрат системы, включающий оценку использования оперативной памяти, загрузки процессора и количества потоков в процессе поиска мелодий.

Разработанное программное средство демонстрирует возможность эффективного применения методов глубокого обучения для решения задачи поиска мелодий по вокальному запросу. Предложенный подход обладает потенциалом дальнейшего развития и может быть расширен за счёт:

– использования более сложных архитектур нейронных сетей;

– увеличения объёма и разнообразия обучающих данных;

– оптимизации этапа ранжирования кандидатов;

– адаптации системы для работы с большими музыкальными каталогами и потоковыми сервисами.

Таким образом, поставленные в работе цели были достигнуты, а разработанная система подтверждает перспективность применения нейросетевых методов и эмбеддинговых представлений в задачах музыкального информационного поиска.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

[1-А] Каминский А.В. Сравнение подходов к построению хромаграммы для анализа аудио / Каминский А.В., Петровский Н.А. // Информационные технологии и системы 2025 : материалы междунар. науч. конф., Минск, 19 нояб. / редкол. Л.Ю. Шилин [и др.]. – Минск : БГУИР,2025. – 149–150с.

[2-А] Каминский А.В. Применение нейронных сетей в задачах поиска мелодии по напеву / Каминский А.В. // 62-ая студенческая научно-техническая конференция аспирантов, магистрантов и студентов БГУИР : материалы науч. конф., Минск, 13-17 апр. – Минск : БГУИР,2026. – 149–150с.