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

УДК 004.627

Аврамов
Владислав Владимирович

Нейросетевой кодер аудиосигналов для систем мультимедиа

АВТОРЕФЕРАТ
на соискание степенимагистра технических наук
по специальности 1-40 80 01 «Элементы и устройства вычислительной техники и систем управления»

Научный руководитель
Петровский Александр Александрович
Профессор, доктор технических наук

Минск 2018

ВВЕДЕНИЕ

Алгоритмы кодирования звука или сжатия звука используются для получения компактных и высококачественных цифровых представлений широкополосных аудиосигналов с целью эффективной передачи или хранения. Основной задачей в кодировании звука является представление сигнала минимальным количеством бит при достижении качественнойреконструкции сигнала, то есть генерирование выходного аудиосигнала, который нельзя отличить от входного, даже чувствительным слушателем.

В настоящее время существует целый ряд алгоритмов, нацеленных на решение задачи сжатия и компрессии аудио. Однако даже при использовании современных методов сжатия по-прежнему актуальными являются исследования, результаты которых позволяют как повысить эффективность известных методов компрессии цифровых аудио-сигналов, так и получить новые алгоритмы кодирования, которые могут быть успешно использованы в различных мультимедийных программных и аппаратных комплексах.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

В настоящее время основным требованием, предъявляемым к мультимедийным системам, является доставка абоненту медиа-контента высокого качества. Следовательно, для современных мультимедиа-систем аудиокодер должен быть инвариантным к акустическому наполнению сигнала, а также обладать возможностью масштабирования скорость битового потока, которая определяется исходя из загруженности канала передачи данных.

Для решения проблемы сжатия и компрессии цифровых аудиосигналов существует целый ряд алгоритмов и разработанных на их основе кодеров. В зависимости от выделяемых параметров входного аудиосигнала кодеры можно разделить на два класса: вокодеры и аудиокодеры. Вокодеры предназначены для сжатия речевого сигнала и позволяют достичь низких скоростей передачи данных при сохранении высокого качества выходного речевого сигнала. Алгоритмы, на основе которых функционируют аудиокодеры, нацелены на работу с таким типом входных аудиосигналов, как, например, музыка или звуки природы. На практике, наличие двух совершенно разных подходов к кодированию аудиосигналов, вызывает затруднения при выборе определенного аудиокодера, или вокодера, для решения конкретной задачи. Это обусловлено тем, что необходимо заранее знать, какой тип входного сигнала будет

преобладать, а также необходимо изучить все представленные алгоритмы для того, чтобы выбрать наиболее эффективный. Таким образом, решением вышеизложенной задачи, является построение инвариантного к акустическому наполнению аудиокодера, способного максимально эффективно работать со всеми известными типами аудиосигналов.

Цель и задачи исследования

Целью данного исследования является разработка масштабируемого нейросетевго кодера аудиосигналов для систем мультимедиа.

В соответствии с поставленной целью, в работе сформулированы и решены следующие задачи:

1. Провести анализ современных алгоритмов кодирования аудиосигналов.
2. Провести анализ применения преобразований на основе многослойных искусственных нейронных сетей для задачи кодирования аудиосигналов.
3. Выбрать архитектуру нейронной сети для решения поставленной задачи и разработать алгоритм ее обучения.
4. Экспериментально оценить разработанную схему кодирования аудиосигналов.

Объектом исследования является кодер аудиосигналов для систем мультимедиа.

Предметом исследования выступают алгоритмы кодирования аудиосигналов.

Область исследованияисодержание диссертационной работы соответствует образовательному стандарту высшего образования второй ступени (магистратуры) специальности 1-40 80 01 «Элементы и устройства вычислительной техники и систем управления».

Научная новизнадиссертационной работы заключается в разработке и верификации алгоритма кодирования аудиосигналов, использующего все преимущества современных схем кодирования аудиосигналов, а именно использование частотно-временных декоррелирующих преобразований и психоакустической модели слуха человека для формирования пространства признаков, с последующей обработкой данного пространства на основе аппарата нейронных сетей для получения компактного его отображения.

Положения, выносимые на защиту

1. Структура нейросетевогокодера аудиосигналов, основу которого составляет параметрический перцептуальный аудиоречевой кодер, построенный на базе алгоритма разреженной аппроксимации сигнала, и многослойный нейросетевой квантователь.

2.Подход к получению дискретных кодов, основанный на примененииступенчатой функции активации в центральном кодовом слое,позволяющий эффективно использовать отведенное кодовое пространство.

3.Подход к представлению данных на основе их сортировки, сочетающий в себе преимущества как скалярного квантования, в частности его инвариантность к взаимосвязям в данных, а также совместного квантования вектора параметров, которое позволяет использовать взаимосвязанные параметры векторных величин.

Апробация результатов диссертации

Основные положения и результаты диссертационной работы докладывались и обсуждались на следующих конференциях:54-я научная конференция аспирантов, магистрантов и студентов БГУИР (Минск, 2018); международная научная конференция International Symposium on Neural Networks (Минск, 2018); международная научная конференция International Conference on Multimedia and Network Information Systems (Wroclaw, 2018).

Опубликованность результатов исследования

По результатам исследований, представленных в диссертации,опубликовано 4 печатные работы, в том числе 3 статьи и 1 тезис в сборниках иматериалах научных конференций.

Структура и объем диссертации

Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, четырёх глав и заключения, библиографического списка и приложений. Общий объем диссертации − 86 страниц. Работа содержит 4 таблицы, 59рисунков. Библиографический список включает 34 наименования, графический материал включает 18 слайдов презентации.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние проблемы кодирования аудиосигналов, определены основные направления исследований, а также дается обоснование актуальности темы диссертационной работы.

В **общей характеристике работы** сформулированы ее цель и задачи, показана связь с научными программами и проектами, даны сведения об объекте исследования и обоснован его выбор, представлены положения, выносимые на защиту, приведены сведения о личном вкладе соискателя, апробации результатов диссертации и их опубликованность, а также, структура и объем диссертации.

В **первой главе** рассматриваются требуемые параметры современной схемы кодирования аудиосигналов, приводиться описание существующих алгоритмов перцептуального кодирования аудиосигналов и аудикодеров на их основе, рассматриваются алгоритмы и подходы к квантованию и кодированию параметров.

Основная цель аудиокодирования состоит в том, чтобы максимизировать воспринимаемое качество при определенной скорости передачи битов или минимизировать скорость передачи битов для определенного качества восприятия.

К требуемым характеристикам современного алгоритма кодирования аудиосигнала относят:

1 Низкая скорость передачи и масштабируемость.

2 Высокое качество реконструкции.

3 Инвариантность к акустическому наполнению.

4 Низкая вычислительная сложность.

5 Надежность при наличии ошибок канала.

6 Низкая задержка кодирования.

Применение нейронных сетей и преобразований на их основе для обработки аудио в настоящее время, как правило, сосредоточено на задачах распознаванияили улучшении сигнала (удаление шума, изменение отдельных характеристик речевых сигналов и т.п.). Применительно к задаче кодирования аудио существуют примеры применения нейронных сетей для задачи кодирования речи.В основе этих систем, лежит частотное декоррелирующее преобразование и кодирование результата преобразования с использованием нейронной сети.

Временная область сигнала не дает никакой информации о его частотных свойствах и, как следствие, обладает низкой информативностью. Поэтому

рациональным подходом к построению нейросетевого аудиокодера является использование всех преимуществ современных схем кодирования аудиосигналов, а именно использование частотно-временных декоррелирующих преобразований и психоакустической модели слуха человека для формирования пространства признаков, с последующей обработкой данного пространства на основе аппарата нейронных сетей для получения компактного его отображения, то есть для задачи квантования.

Векторное квантование заключается в отображении в многомерном пространстве векторов с вещественными, непрерывными амплитудами компонент к некоторому дискретному множеству. Процесс векторного квантования позволяет исключить избыточность за счет эффективного использованиявзаимосвязанных свойств векторных параметров, к которым относят линейные и нелинейные зависимости, форму функции плотности вероятности, а такжемногомерность векторной величины. Проблема достижения глобальной оптимальности в дизайне системы векторного квантования по-прежнему представляет собой сложную проблему. Кроме того, реализация оптимального векторного квантователя требует больших затрат памяти, и вычислительных затрат для поиска по кодовым книгам.

Другим подходом к совместному отображению в многомерном пространствевекторов с вещественными, непрерывными амплитудами компонент к некоторому дискретному множеству является нейросетевое квантование. Нейросетевое квантование, как и векторное квантование, позволяет эффективно использовать взаимосвязанные свойства векторных параметров, такие как линейные и нелинейные зависимости, форму функции плотности вероятности, а такжемногомерность векторной величины, что позволяет исключить избыточность входных данных. В отличие от классических вариантов векторного квантования, нейросетевой квантователь«генерирует» каждый выходной вектор на основе входного вектора и вычислений, заключающихся в умножениях входных векторов на матрицы весовых коэффициентов слоев нейросетевого квантователя.

Во **второй главе** приведена структура параметрического перцептуального аудиокодера, и структура нейросетевого квантователя.

Параметрический перцептуальный аудиоречевой кодер построен на основе алгоритма разреженной аппроксимации (согласованная подгонка). Параметрическая модель данного аудиокодера позволяет уменьшить количество информации необходимое для описания кодируемого сигнала. Для более эффективной параметризации используется психоакустическая модель,

которая позволяет исключить избыточную для восприятия человеком информацию. Для проведения многомасштабного анализа и точного определения временной локализации частот используется частотно-временные преобразования сигнала.

Перцептуально значимые вейвлет-коэффициенты определенные в алгоритме согласованной подгонки и их позиции в дереве ПДВП из блока разреженной аппроксимации сигнала поступают в блок квантования и кодирования.

Установлено, что естественной архитектурой нейронной сети для решения задачи квантования, является структура сети, представляющей собой сеть прямого распространения, содержащая входной, кодовый и выходной слои, при чем размерность входного слоя должна быть равна размерности выходного слоя. Основной задачей обучения нейронной сети данной архитектуры является получение на выходе сигнала наиболее близкого к входному.

Многослойные нейронные сети, образованные каскадами слоев, имеют, как правило, большую вычислительную мощность. Увеличение количества слоев сети обеспечивает прирост вычислительных возможностей только в случае, когда функция активации между слоями является нелинейной.

В **третьей главе**описан разработанный алгоритм обучения нейронной сети.

Обучение нейросетевого квантователя аудиокодера рассматривается как неконтролируемый алгоритм обучения (без учителя), в процессе которого устанавливаются целевые выходные значения равными входам, а коды в центральном слое формируются в центральном слое автоматически в процессе обучения.

При обучении многослойной нейронной сети, сети с числом скрытых слоев более двух, методом обратного распространения ошибки возникает проблема затухающего градиента.При вычислении градиента по методу обратного распространения ошибки его значение уменьшается по мере распространения от выходного слоя к входному. Эта проблема приводит к низкой эффективности обучения таких нейронных сетей. Наиболее эффективно обучать многослойные сети с использованием алгоритма послойного обучения.Согласно данному алгоритму, процедура обучения многослойного нейросетевого квантователя аудиокодера была разделена на следующие две фазы: послойное предварительное обучение сети – это последовательное попарное обучение соседних слоёв нейронной сети; тонкая подстройка весовых

коэффициентов – обучение по методу обратного распространения ошибки одним из градиентных методов.

Рассмотренные алгоритмы корректировки параметров и методы регуляризации позволяют ускорить процесс обучения нейронной сети по алгоритму обратного распространения ошибки, а также получить лучшие функции представления данных во внутренних (скрытых) слоях.

Выходы кодового слоя квантователя должны быть дискретны. Данное ограничение ведет к NP-сложной проблеме оптимизации. На основе анализа современных подходов к решению данной проблемы, было установлено, что к наилучшим результатам приводит использование ступенчатой функции активации.

В процессе поиска оптимального представления входных данных был предложен уникальный подход к представлению данных на основе их сортировки, сочетающий в себе преимущества как скалярного квантования, в частности его инвариантность к взаимосвязям в данных, а также совместного квантования вектора параметров, которое позволяет использовать взаимосвязанные параметры векторных величин.

В **четвертой главе** представлены результатыэкспериментальных исследований.

На основании анализа зависимости ошибки от количества эпох обучения, было установлено, что послойное предварительное обучение доказало свою эффективность, и обеспечивает хорошее приближение к решению задачи оптимизации и снижение ошибки.

Анализ ошибки реконструкции для тестовых образцов показал, что предложенный подход к представлению данных позволяет снизить ошибку реконструкции на 3 порядка.

Сравнительный анализ ошибки реконструкции для нейросетевого квантователяи векторного квантователяпоказывает, что нейросетевой квантовательпревосходит векторный квантователь на большинстве тестовых образцов при вдвое меньшем количестве вычислительных затрат.

Результаты объективной оценки качества реконструируемого аудиосигнала показали, что предложенная схема аудиокодирования близка к скалярному квантованию по качеству реконструкции, однако кратно превосходит ее в степени сжатия сигнала.

В **приложениях** представлен графический материал иисходное описание разработанной системы на языке Matlab

ЗАКЛЮЧЕНИЕ

В настоящей работе было проведено исследование применения многослойных искусственных нейронных сетей и преобразований на их основе применительно к задаче задачи кодирования аудиосигналов.

В результате был разработан масштабируемый нейросетевой кодер аудиосигналов, выбрана и описана архитектура нейронной сети для решения поставленной задачи и разработан алгоритм ее обучения.

В основе разработанного нейросетевого аудиокодера лежит параметрический перцептуальный аудиоречевой кодер, представляющий собой синтез всех преимуществ современных схем кодирования аудиосигналов, среди которых применение частотно-временных декоррелирующих преобразований и психоакустической модели слуха человека, используемых для формирования пространства признаков, с последующей обработкой данного пространства на основе аппарата нейронных сетей для получения компактного его отображения, то есть для задачи квантования.

В ходе проведения экспериментальных исследований были определены оптимальные параметры нейросетевого квантователя. Экспериментальные результаты подтвердили актуальность настоящего исследования. Разработанный нейросетевой квантователь в составе аудиокодера обеспечивает высокую степень компрессии, без значительных потерь качества сигнала.

Полученная схема кодирования инвариантна к акустическому наполнению сигнала и является масштабируемой. Результаты объективной оценки качества реконструируемого аудиосигнала показали, что данная схема близка к скалярному квантованию по качеству реконструкции, однако кратно превосходит ее в степени сжатия сигнала. Кроме того, экспериментальные результаты показали, что разработанный подход обеспечивает реконструкцию с качеством, сопоставимым или лучшим по сравнению с алгоритмами векторного квантования, и при этом требует меньше вычислительных затрат.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1] Аврамов, В. В. Применение нейросетевого квантователя в аудиоречевом кодере на основе разреженной аппроксимации и исследование его эффективности / В. В. Аврамов, В. Ю. Герасимович // Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных : материалы международного научно-технического семинара. – Минск : БГУИР, 2017. – С. 77-82.

[2] Аврамов, В. В. Алгоритм обучения многослойного нейросетевогоквантователя аудиокодера/ В. В. Аврамов // Компьютерные системы и сети: материалы 54-й научной конференции аспирантов, магистрантов и студентов. – Минск: БГУИР, 2018. – С. 222-223.

[3] Avramov, V. SoundSignalInvariantDAENeuralNetwork-BasedQuantizerArchitectureofAudio/Speech Coder Using the Matching Pursuit Algorithm / V. Avramov, V. Herasimovich, A. Petrovsky // Advances in Neural Networks –Proceedings of the 15th International Symposium on Neural Networks. – Minsk, 2018. – P. 511-520.

[4] Herasimovich, V. Audio/Speech Coding Based on the Perceptual Sparse Representation of the Signal with DAE Neural Network Quantizer and Near-End Listening Enhancement/ V. Herasimovich, Al. Petrovsky, V. Avramov, A. Petrovsky // Multimedia and Network Information Systems–Proceedings of the 10th International Conference MISSI 2018. – Wroclaw, 2018.