УДК 004.627

ПРИМЕНЕНИЕ НЕЙРОСЕТЕВОГО КВАНТОВАТЕЛЯ В АУДИОРЕЧЕВОМ КОДЕРЕ НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ И ИССЛЕДОВАНИЕ ЕГО ЭФФЕКТИВНОСТИ

В.В. АВРАМОВ, В.Ю. ГЕРАСИМОВИЧ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 29 октября 2017

Приводится описание алгоритма кодирования аудиосигналов на основе согласованной подгонки со словарем частотно-временных функций с применением аппарата искусственных нейронных сетей для квантования параметров кодера. Определение словаря функций происходит посредством пакетного дискретного вейвлет-преобразования, перцептуально оптимизированного для анализируемого фрейма входного сигнала. Описана возможность применения нейронных сетей прямого распространения для решения задачи квантования параметров кодера. Приводятся структура нейросетевого квантователя и результаты исследования его эффективности.

Ключевые слова: аудиосигнал, нейронные сети, согласованная подгонка, квантование, частотно-временное преобразование.

Введение

Сжатие звуковой информации является актуальной задачей в современном мире, поскольку активное развитие и внедрение получают такие технологии как передача аудиоинформации по коммуникационным каналам (VoIP, VoLTE), потоковое вещание мультимедиа (Streaming Media), цифровое радиовещание (DAB). В каждой из перечисленных технологий одной из наиболее важных задач является компактное представление цифрового звукового сигнала. Современные алгоритмы сжатия звука обладают высокими показателями качества реконструированного сигнала и различными вариациями скорости битового потока. Однако они дифференцированы относительно типа входного сигнала: часть из них рассчитана на работу с речевыми входными данными (вокодеры) [1], другие же предназначены для работы со аудиоинформацией и не позволяют добиться максимальной степени сжатия для речи [2]. Существуют гибридные подходы, позволяющие работать с двумя типами входного звукового сигнала [3, 4], однако в данном случае в кодере предполагается использование двух моделей и детектора входного сигнала, который определяет, на какую из них направлять сигнал. Следовательно, разработка универсального кодера, способного работать со входными сигналами с различным звуковым информационным наполнением в реальном масштабе времени является актуальной задачей.

В работе показан алгоритм сжатия на основе согласованной подгонки (СП) со словарями частотно-временных функций, построенный на базе пакетного дискретного вейвлет-преобразования (ПДВП), перцептуально оптимизированного для входного фрейма сигнала. В любом аудиокодере с потерями важным шагом процесса кодирования является алгоритм квантования данных. В работе показан вариант применения аппарата искусственных нейронных сетей (ИНС) прямого распространения в качестве квантователя. Нейросетевое квантование [5] заключается в отображении многомерного пространства векторов с вещественными, непрерывными амплитудами компонент в некоторое дискретное множество. Иными словами, нейросетевое квантование – это совместное квантование вектора параметров, которое позволяет

исключить избыточность за счет эффективного использования взаимосвязанных свойств векторных параметров, к которым относят линейные и нелинейные зависимости, форму функции плотности вероятности, а также многомерность векторной величины.

Описание аудиоречевого кодера на основе согласованной подгонки

Алгоритм сжатия универсального масштабируемого аудиокодера состоит из двух основных частей: перцептуально оптимизированного ПДВП и СП. Входной фрейм сигнала поступает в блок адаптивного ПДВП, в котором осуществляется определение наиболее оптимального дерева декомпозиции на базе психоакустического моделирования [6, 7] и двух стоимостных функций – временной и перцептуальной энтропии [8]. Дерево декомпозиции ПДВП представляет собой словарь частотно-временных функций для следующего шага работы алгоритма кодирования – параметризации входного фрейма на основе СП алгоритма, позволяющего осуществить разреженную аппроксимацию сигнала, т.е. его представление минимальным числом ненулевых компонентов [9]. В разрабатываемом аудиокодере на данном шаге выбирается наиболее перецептуально значимые для восприятия элементы (атомы), которые дают максимальное соответствие между моделируемой и исходной скалограммами [10]. Алгоритм СП повторяется до тех пор, пока не достигнуто определенное условие остановки. Таким условием может служит фиксированное количество отобранных атомов, определенный порог энергии остаточного сигнала, а также перцептуальный критерий, который говорит о том, на сколько релевантная для восприятия человеческим ухом информация находится в сигнале-остатке. Выбранные путем такого моделирования атомы необходимо заквантовать для их компактного представления.

Квантование на основе ИНС прямого распространения

Естественной архитектурой ИНС для реализации квантователя является сеть прямого распространения, содержащая входной, кодовый и выходной слои, причем размерность входного слоя должна быть равна размерности выходного слоя. Основной задачей обучения такой ИНС является получение на выходе вектора с минимальным отклонением от входного.

Однослойная сеть (сеть с одним скрытым слоем), описанная выше, весьма ограничена по своим вычислительным возможностям. Объем информации, запоминаемый и воспроизводимый ИНС данного типа, зависит от количества нейронов. Исходя из этого, очевидно, что одного скрытого слоя будет недостаточно, и выбор следует сделать в пользу многослойной архитектуры. Таким образом, для расширения возможностей необходимо применять многослойные нейронные сети [11].

При обучении многослойной нейронной сети (сети с числом скрытых слоев более двух), методом обратного распространения ошибки [12] возникает проблема затухающего градиента, которая приводит к низкой эффективности обучения таких ИНС. При вычислении градиента по методу обратного распространения ошибки его значение уменьшается по мере распространения от выходного слоя к входному [13]. Решением данной проблемы является разделение процесса обучения многослойной ИНС [14] на следующие две стадии.

1. Предварительное послойное обучение сети – последовательное попарное обучение соседних слоев ИНС.

2. Тонкая подстройка весовых коэффициентов – обучение всей сети по методу обратного распространения ошибки одним из градиентных методов.

Идея такого обучения заключается в том, что во время обучения не потребуется значительно менять веса многослойной сети, которые были проинициализированы послойным предобучением, как следствие проблема затухания градиента уже не будет так сильно влиять на данный процесс. Предобучение заключается в выделении пары соседних слоев многослойной нейронной сети, начиная со входного слоя, и обучения этой пары слоев. Процедура последовательно повторяется для всех слоев сети.

Установим формально задачу обучения нейросетевого квантователя как минимизацию функции потери в пространстве весовых коэффициентов. Обозначим функцию, реализуемую сетью – f, обучающее множество – $\mathbf{X}$, количество обучающих примеров – N, весовые

коэффициенты сети – **W**, квантованный и неквантованный выходы кодового слоя **B** и **H** соответственно. Тогда функцию потери в матричном виде можно записать следующим образом:

$$E\left(f\left(\mathbf{X},\mathbf{W}\right),\mathbf{X}\right)=\frac{1}{2N}\cdot\left\|f\left(\mathbf{X},\mathbf{W}\right)-\mathbf{X}\right\|^2+\frac{\alpha}{2N}\cdot\left\|\mathbf{H}-\mathbf{B}\right\|^2+\frac{\beta}{2}\cdot\left\|\frac{1}{N}\cdot\mathbf{H}\cdot\mathbf{H}^T-\mathbf{I}\right\|^2+\frac{\gamma}{2N}\cdot\left\|\mathbf{H}\cdot 1_{N\times 1}\right\|^2,$$ где **I** – единичная матрица. В данном выражении первое слагаемое представляет собой среднеквадратичную ошибку между входом сети и ее выходом и соответствует основной цели обучения (получение на выходе вектора с минимальным отклонением от входного). Второе слагаемое в выражении отвечает за минимизацию ошибки квантования выходов кодового слоя. Его присутствие обусловлено тем, что функции активации кодового слоя, как и всех остальных слоев ИНС, является гиперболический тангенс, а получить двоичные коды можно лишь при использовании пороговой функцией активации, которую нельзя использовать в сетях, обучающихся по алгоритму обратного распространения ошибки, т.к. данный алгоритм требует дифференцируемости передаточной функции. С целью увеличения емкости кодового слоя и предотвращения получения тривиальных кодов в алгоритм обучения были включены условия ортогональности выходов кодового слоя [15] и сбалансированности [16], регулируемые параметрами β и γ соответственно. Условие ортогональности позволяет получать более независимые преобразования в кодовом слое, а условие сбалансированности – коды, где «0» и «1» равновероятны.

Как показано на рис. 1, нейросетевой квантователь и деквантователь для атомов аудиокодера представлены шестислойной ИНС прямого распространения.

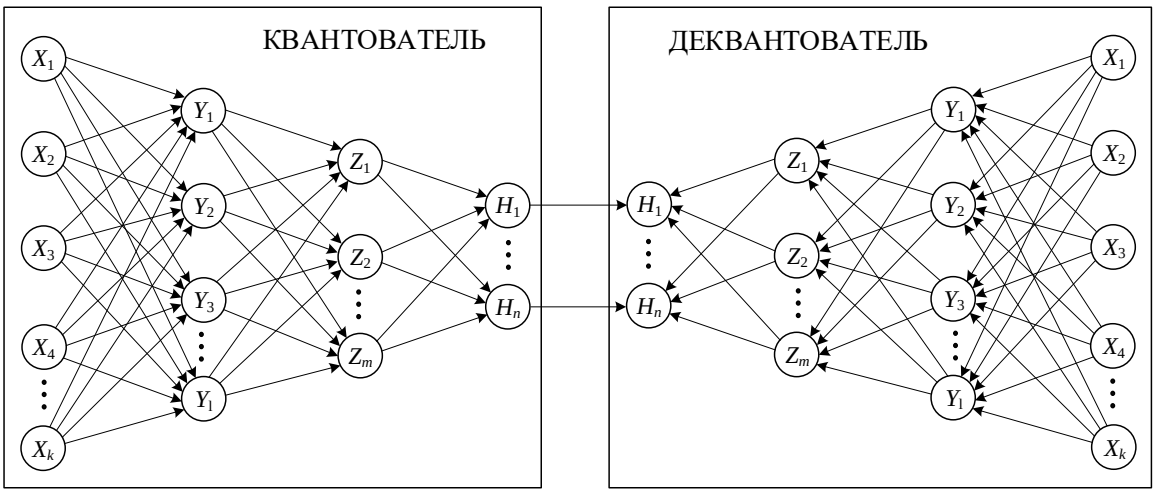

Рис. 1. Структура нейросетевого квантователя

В процессе обучения многослойная ИНС рассматривается как единая сеть, однако для функционирования в составе аудиокодера ее необходимо разделить на две части: квантователь, находящийся на стороне кодера, и деквантователь – на стороне декодера. Центральный слой рассматривается как кодовый в квантователе и как входной в деквантователе. Данный слой позволяет получить двоичное отображение входной информации (атомов). Кодовый вектор формируется в процессе прямого распространения от входного слоя нейронов квантователя к выходному. Реконструированный сигнал формируется симметричной нейронной сетью в процессе прямого прохода от входного слоя нейронов деквантователя к выходному. Поскольку деквантователь представляет собой симметричную ИНС, которая обучается вместе с квантователем, то он может реконструировать входной вектор с точностью, зависящей от обучения сети и ее внутренней архитектуры. Данный подход позволяет добиться большей глубины компрессии относительно скалярного квантования, при этом, с низкой степенью искажения выходного реконструированного аудиосигнала в случае успешного эффективного обучения ИНС.

Экспериментальные результаты исследований

Обучение нейросетевого квантователя проводилось для 200 атомов. Структура слоев ИНС квантователя аудиокодера в процессе обучения приведена в табл. 1.

Таблица 1. **Конфигурации нейросетевого квантователя аудиокодера**

Слой	Размер	Функция активации
1	200 (вход)	-
2	160	Гиперболический тангенс
3	128	Гиперболический тангенс
4	100 (код)	Гиперболический тангенс
5	128	Гиперболический тангенс
6	160	Гиперболический тангенс
7	200 (выход)	Линейная

Входной тестовой последовательностью служили одноканальные образцы звуковых сигналов (табл. 2) с частотой дискретизации 44.1 кГц и разрядностью отсчетов 16 бит. В табл. 2 показано усредненное для фреймов каждого образца среднеквадратическое отклонение (СКО).

Таблица 2. **Тестовые образцы и ошибка их реконструкции**

№	Наименование образца	Описание	Ошибка реконструкции (СКО)
1	es01	Вокал (SuzanVega)	$2,57 \cdot 10^{-5}$
2	es02	Речь на немецком языке	$1,22 \cdot 10^{-4}$
3	es03	Речь на английском языке	$1,69 \cdot 10^{-4}$
4	sc01	Соло на трубе и оркестр	$5,27 \cdot 10^{-5}$
5	sc03	Современная поп-музыка	$1,32 \cdot 10^{-4}$
6	si01	Клавесин	$1,76 \cdot 10^{-5}$
7	si02	Кастаньеты	$1,45 \cdot 10^{-5}$

Принцип расчета скорости битового потока при использовании скалярного квантования в аудиокодере подробно показан в [17]. Так, для варианта с использованием 200 атомов для реконструкции расчетная степень сжатия составляет порядка 19 раз. При этом, в каждом фрейме передаются 200 атомов, значения которых сепарировано квантуются, и с учетом распределения бит на отсчет получается, что для их представления необходимо примерно 380 бит. В описываемом способе квантования на основе ИНС все 200 атомов отображаются в единый вектор, состоящий из 100 двоичных значений. Следовательно, скорость битового потока для данного варианта с учетом необходимости кодирования дополнительной информации ИНС будет в 3,45 раз ниже, нежели при скалярном квантовании, т.е. степень сжатия приблизительно будет составлять 65,5 раз.

На рис. 2 отражены оригинальный и закодированный сигналы. Как видно из данных спектрограмм, в реконструированном сигнале почти полностью отсутствует информация на частотах, составляющих 15 кГц и более. Это связано с тем, что на этапе выбора атомов элементы данной частотной полосы показали минимальную информативность для восприятия сигнала человеческим ухом и не были включены в выходной набор атомов. Остальная часть информации была передана при реконструкции достаточно подробно, с учетом высокой степени сжатия.

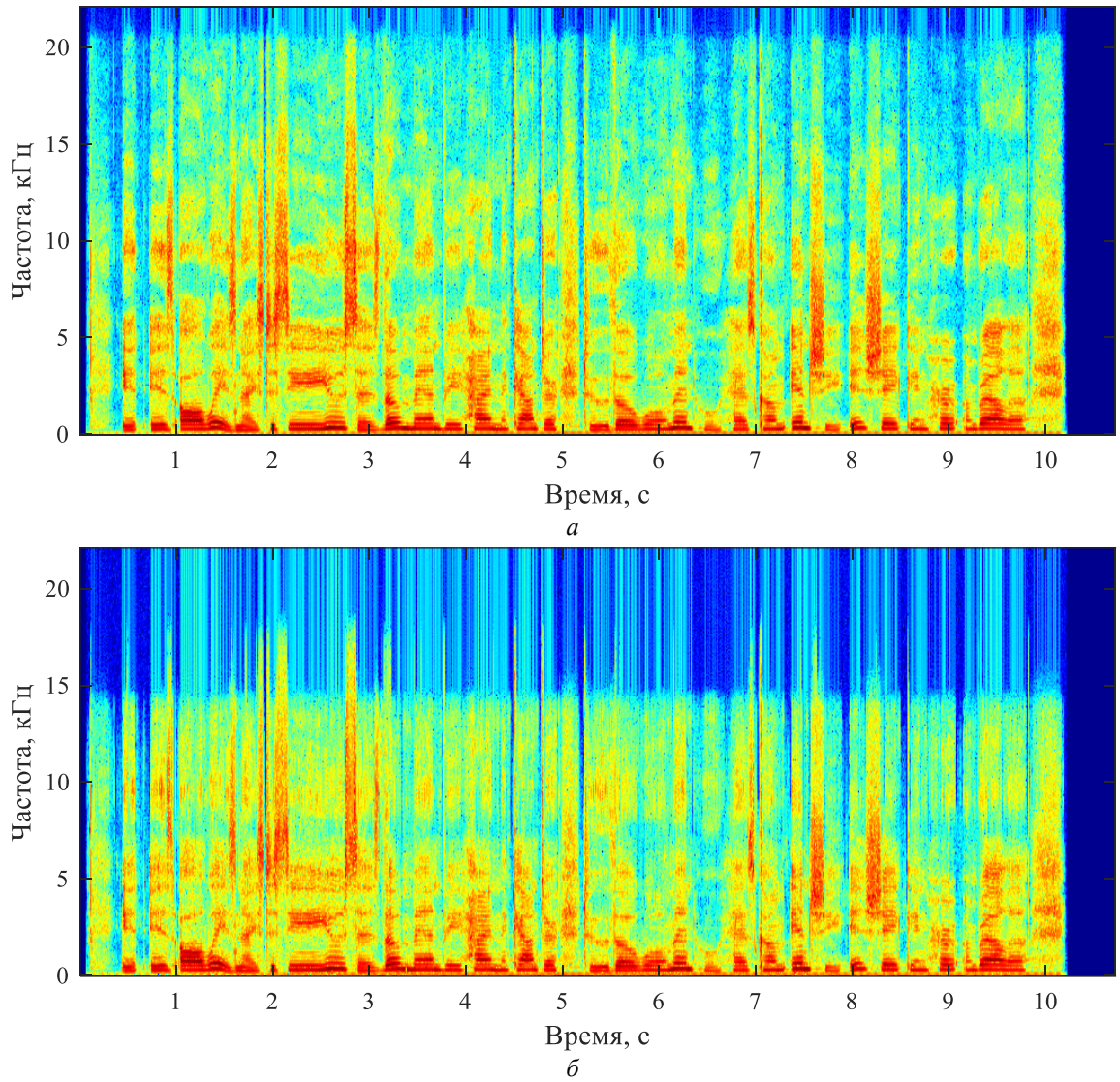

Рис. 2. Результаты работы нейросетевого квантователя универсального аудиокодера:
а – спектрограмма оригинального сигнала; *б* – спектрограмма реконструированного сигнала

Заключение

В работе описан универсальный масштабируемый аудиокодер на основе алгоритма СП с перцептуально оптимизированными словарями частотно-временных функций на базе ПДВП. Показан алгоритм квантования атомов на основе ИНС прямого распространения и проведен анализ результатов его работы. При использовании данного подхода достигнуты высокие результаты качества реконструированного сигнала при высокой степени сжатия. При этом, данный тип квантования, основанный на алгоритме машинного обучения, будет напрямую зависеть от объема информации в обучающей выборке и длительности процесса обучения, т.е. качество результатов будет возрастать при увеличении данных параметров.

APPLICATION OF THE NEURAL NETWORK QUANTIZER IN THE AUDIOSPEECH CODER BASED ON THE SPARSE APPROXIMATION AND STUDYING OF ITS EFFECTIVENESS

V.V. AVRAMOV, V.Y. HERASIMOVICH

Abstract

Description of the audio signal encoding algorithm based on the matching pursuit with the time-frequency dictionary and the artificial neural networks for quantization is given. Determination of the dictionary occurs through the perceptually optimized for the input signal frame wavelet packet transform. Utilization of the feed forward neural networks for the quantization goal is described. Structure of the quantizer and its performance evaluation is given.

Keywords: audiosignal, neural networks, matching pursuit, quantization, time-frequency transforms.

Список литературы

1. *Spanias A.* // Proceedings of the IEEE, Vol. 82. 1994. P. 1541–1582.
2. *Painter T., Spanias A.* // Proceedings of the international conference on digital signal processing. DSP 97. 1997. P. 179–205.
3. *Valin J.-M. et al.* // AES 135th Convention. 2013. P. 8942.
4. *Vos K. et al.* // AES 135th Convention. 2013. P. 8941.
5. *Серков В.В., Петровский А.А.* // Труды 4-й Международной конференции Цифровая обработка сигналов и ее применение DSPA'2002. 2002. С. 426–428.
6. *Петровский Ал.А.* // Речевые технологии. 2008. №4. С. 61–71.
7. *Герасимович В.Ю., Петровский Ал.А.* // Информатика. 2017. №4(56). С. 89–103.
8. *Petrovsky Al., Krah D., Petrovsky A.* // AES 114th Convention. 2003. P. 5778.
9. *Mallat S., Zhang Z.* // IEEE Transactions on Signal Processing. 1993. Vol. 41. №12. P. 3397–3415.
10. *Petrovsky Al., Herasimovich V., Petrovsky A.* // Signal Processing: Algorithms, Architectures, Arrangement, and Applications (SPA). 2016. P. 225–229.
11. *Уоссермен Ф.* Нейрокомпьютерная техника: Теория и практика. М., 1992.
12. *Rummelhart D.E., Hinton G.E., Williams R.J.* // Computational models of cognition and perception. 1986. Vol. 1, Chap. 8.
13. *Хайкин С.* Нейронные сети: полный курс. М., 2006.
14. *Bengio Y., Lamblin P., Popovici D.* // NIPS. 2006. P.153–160.
15. *Thanh-Toan Do, Anh-Dzung Doan, Ngai-Man Cheung* // The 14th European Conference on Computer Vision. 2016. P. 219–234.
16. *Erin V., Liong J., Lu G. et al.*// IEEE Conference on Computer Vision and Pattern Recognition. 2015. P. 2475–2483.
17. *Petrovsky Al., Herasimovich V., Petrovsky A.* // AES 138th Convention. 2015. P. 9264.