

$$T_{об} = N \cdot t_i \cdot \log_2 N . \quad (8)$$

Возможны последовательная и параллельная реализация данного метода. При параллельной реализации сравнение чисел происходит параллельно. При последовательной – используется память для хранения промежуточных максимумов.

Вывод: данный метод удобно использовать при поиске максимума, для полного ранжирования этот метод требует достаточно больших аппаратных затрат, показывая при этом долгое время выполнения.

Заключение. Результаты сравнения рассмотренных реализаций алгоритмов сортировок, приведенные в таблице, показывают, что минимальное время достигается при последовательной реализации метода подсчета при количестве схем сравнения, равным количеству элементов в неупорядоченной последовательности.

Таблица

Время ранжирования различных аппаратных реализаций

название метода		количество сравниваемых элементов			
		4	8	16	32
Метод подсчета	Последовательный с 1 схемой	$12t_i$	$56t_i$	$240t_i$	$992t_i$
	Последовательный 2 схемами	$6t_i$	$28t_i$	$120t_i$	$446t_i$
	Последовательный n-схемами	$3t_i$	$7t_i$	$15t_i$	$31t_i$
Метод пузырька	Последовательный с 1 схемой	$6t_i$	$28t_i$	$120t_i$	$446t_i$
	Последовательный 2 схемами	$4t_i$	$14t_i$	$60t_i$	$223t_i$
	Параллельный	$6t_i$	$28t_i$	$120t_i$	$446t_i$
	Модернизированный	$5t_i$	$13t_i$	$29t_i$	$61t_i$
Пирамидальная сортировка	Последовательный с 1 схемой	$12t_i$	$48t_i$	$240t_i$	$992t_i$
	Последовательный n схемами	$8t_i$	$24t_i$	$64t_i$	$160t_i$
	Параллельный	$6t_i$	$20t_i$	$56t_i$	$155t_i$

Ю.В. Чернухин, А.В. Лунев

НЕЙРОСЕТЕВОЙ ПОДХОД К ВЕКТОРНОМУ КВАНТОВАНИЮ КОЭФИЦИЕНТОВ ОТРАЖЕНИЯ ПРИ ВОКОДИРОВАНИИ МЕТОДОМ ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ

Введение. В настоящее время эффективное сжатие речевых данных (вокодирование) является весьма важной задачей в области современных систем телекоммуникаций. В большинстве существующих алгоритмов вокодирования одним из этапов является векторное квантование первичных признаков при анализе речи [1]. Эффективность квантования таких признаков непосредственно сказывается как на качестве воспроизводимой речи, так и на скорости ее передачи. Оба эти параметра являются основополагающими для вокодеров.

Традиционно векторное квантование выполняется на основе кодовой книги [1], которая формируется на этапе проектирования. Наиболее распространен-

ным алгоритмом построения такой книги является алгоритм LBG (Linde-Buzo-Gray) [2]. Однако в работе [3] показано, что данный алгоритм часто сходится к локальному оптимуму. Альтернативным решением для формирования кодовой книги является самоорганизующийся обучающий алгоритм Кохонена [4], позволяющий избежать проблему локального оптимума и, в отличие от алгоритма LBG, позволяет обновлять кодовую книгу в реальном масштабе времени. В работе [5] представлен сравнительный анализ, который подтверждает эффективность алгоритма Кохонена по отношению к алгоритму LBG. Иными словами, в общем случае для решения задачи векторного квантования целесообразно использование самоорганизующихся нейросетей Кохонена [4].

В работе [6] векторное квантование первичных признаков с помощью нейросети Кохонена было нами экспериментально исследовано на основе канального вокодера. Однако, в отличие от полной схемы вокодера, в данном эксперименте не использовалось распознавание вокализованных и невокализованных звуков, и не определялся основной тон речевого сигнала. В декодирующей части вокодера в качестве сигнала возбуждения использовался только источник белого шума, поэтому звучание синтезированной речи осуществлялось в виде шепота. По этой причине приведенная в [6] схема хотя и подтверждает принципиальную возможность высокой степени сжатия речи с помощью нейросети Кохонена, не может служить в качестве полноценного вокодера речи.

В данной работе мы проводим более детальное исследование эффективности векторного квантования с помощью нейросети Кохонена в составе полнофункционального вокодера. За основу взят вокодер DoD FS 1015 LPC-10 [7], являющийся первым стандартизованным вокодером, работающим на скорости передачи 2400 бит/с. Применяв нейросеть Кохонена для квантования первичных признаков, мы рассчитываем существенно снизить скорость передачи бит без значительного снижения качества воспроизводимой вокодером речи.

Квантование коэффициентов отражения в вокодере LPC-10. В вокодере LPC-10 в качестве формирования первичных признаков на этапе анализа речевого сигнала используется кодирование с линейным предсказанием (КЛП). Для вокодеров, использующих КЛП, основной трудностью снижения скорости передачи битов является квантование параметров линейного предсказания [8]. Очень важно квантовать эти параметры, используя как можно меньше бит, но при этом, незначительно ухудшая качество речи. В вокодере LPC-10 параметрами линейного предсказания являются так называемые коэффициенты отражения, представляющие собой побочные величины, формируемые в результате вычисления коэффициентов линейного предсказания. Коэффициенты отражения более удобны, чем коэффициенты линейного предсказания для передачи по линии связи, так как в силу своих статистических характеристик в меньшей степени могут приводить к потере устойчивости фильтра линейного предсказания при квантовании [9]. В вокодере LPC-10 вычисляются десять коэффициентов отражения, первые два из которых дополнительно преобразуются функцией логарифмического отношения площадей LAR (Log-Area Ratio). Затем все десять полученных коэффициентов скалярно квантуются.

Однако скалярное квантование параметров линейного предсказания является субоптимальным в отношении коэффициента спектральных искажений в том смысле, что коэффициент спектральных искажений не может быть снижен до минимума, если каждый параметр квантуется отдельно. Таким образом, векторное квантование параметров линейного предсказания значительно эффективнее скалярного [1]. Поэтому с целью снижения скорости передачи вокодера LPC-10 ниже 2400 бит/с без значительного снижения качества воспроизводимой речи, в данной

работе мы заменили в алгоритме LPC-10 скалярное квантование коэффициентов отражения векторным квантованием этих коэффициентов с помощью нейросети Кохонена. Преобразование первых двух коэффициентов функцией логарифмического отношения площадей (LAR) в данном случае не выполняется.

Расчет скоростей передачи бит при различных схемах квантования. В стандартном вокодере LPC-10 [7] для вокализованных речевых сигналов на формирование результата скалярного квантования всех десяти коэффициентов отражения выделено 41 бит информации, а для невокализованных сигналов – 20 бит для квантования коэффициентов отражения и 21 бит для защиты от ошибок в канале связи. Для квантования основного тона речевого сигнала и информации о вокализации выделено 7 бит, для квантования энергии речевого сигнала – 5 бит, для синхронизации – 1 бит. Все эти 54 бита образуют кадр, который передается по каналу связи каждые 22,5 мс. Таким образом, скорость передачи бит составляет $54 \text{ бит} / 22,5 \text{ мс} = 2400 \text{ бит/с}$. В случае, если не передавать биты защиты и допустить, что данные передаются при переменной скорости, то для передачи вокализованного кадра потребуется 54 бита, а для невокализованного кадра – 33 бита. Таким образом, средняя скорость передачи бит составит $((54 \text{ бит} + 33 \text{ бит}) / 2) / 22,5 \text{ мс} = 1933 \text{ бит/с}$.

При векторном квантовании коэффициентов отражения требуется меньше бит информации. В данной работе построены две конфигурации нейросети Кохонена. Одна состоит из 4096 нейронов, а другая из 1024 нейронов. Для формирования результата квантования при использовании нейросети с 4096 нейронами необходимо $\log_2 4096 = 12$ бит. Следовательно, экономия для вокализованных кадров почти в 3,5 раза больше по сравнению со скалярным квантованием (12 бит вместо 41 бита). А для невокализованных кадров – в 1,5 раза больше (12 бит вместо 20 бит). Поскольку для передачи кадра в сумме потребуется 25 бит, то постоянная скорость передачи бит составит $25 \text{ бит} / 22,5 \text{ мс} = 1111 \text{ бит/с}$.

В случае конфигурации нейросети, состоящей из 1024 нейронов, выходной результат представлен 12 битами. Постоянная скорость передачи при этом – 1022 бит/с.

Конфигурация и обучение нейросети Кохонена. Как было уже сказано, было построено две конфигурации нейросети Кохонена. Отличие этих конфигураций только в количестве нейронов сети – 4096 и 1024 нейрона. Входное пространство – десятимерное. Топология нейросети – двумерная решетка.

Функционирование нейросети выполняется следующим образом [4].

Нейрон-“победитель” определяется по правилу:

$$d(x, w_c) = \min_i \{d(x, w_i)\}, \quad i = 1, 2, \dots, N, \quad (1)$$

где x – входной вектор нейросети, w_c – весовой вектор нейрона-“победителя”, w_i – весовой вектор i -го нейрона, $d(x, w_i)$ – функция расстояния (в данном случае Эвклидово расстояние), N – число нейронов в сети.

Процесс обучения выполняется по алгоритму:

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)], \quad (2)$$

$$h_{ci} = h_0 \exp\left(-\|r_i - r_c\|^2 / \sigma^2\right), \quad (3)$$

где $w_i(t+1)$ – новое значение весового вектора i -го нейрона, $w_i(t)$ – предыдущее значение весового вектора i -го нейрона, $h_0 = h_0(t)$ – скорость обучения, $\sigma = \sigma(t)$ – размер окрестности, $\|r_i - r_c\|$ – Эвклидово расстояние между нейроном-“победителем” и i -м нейроном согласно их топологическому расположению в двумерной решетке.

Так как диапазон входных значений колеблется для каждого из элементов входного вектора x от -1 до $+1$, начальная инициализация весов нейронов установлена в среднее значение диапазона, а именно в 0 . Обучение нейросети выполнялось в два этапа.

На первом этапе выполняется упорядочивание нейронов. Количество шагов обучения было выбрано равным 50000 . Коэффициент скорости обучения h_0 равномерно снижается от $0,9$ на первом шаге до $0,5$ на последнем шаге. Размер окрестности σ вначале этапа устанавливается равной величине максимального расстояния между нейронами сети и снижается к концу этапа до 0 , после чего величина h_{ci} устанавливается равной h_0 на время выполнения второго этапа.

На втором этапе выполняется тонкая подстройка весов нейронов. Количество шагов обучения было выбрано равным 10 млн. Коэффициент скорости обучения h_0 равномерно снижается от $0,5$ на первом шаге до 0 на последнем шаге.

Подготовка аудио данных. Речевые данные для обучения и тестирования нейросети Кохонена были записаны с телевизионного канала BBC World. Речевой материал включает в себя только английский язык, речь мужского и женского рода и содержит как чистую речь (например, постоянный репортер, интервью внутри помещения), так и речь с высоким уровнем фонового шума (например, уличные репортажи, шум транспорта, ветер, фоновые голоса, фоновая музыка и т.д.).

Все аудио данные были разделены на две части (обучающее и тестовое множество). Таким образом, обучающие данные составили по длительности около 500 секунд. А тестовые данные – примерно 540 секунд. Все данные были оцифрованы с частотой дискретизации 8 кГц и разрешением 16 разрядов в режиме моно.

Эксперименты и результаты. В целях тестирования нового вокодера с векторным квантованием коэффициентов отражения и сравнения его с исходным вокодером LPC-10, выполнена объективная (математическая) оценка качества воспроизводимой вокодерами речи. Использовались следующие объективные оценки: перцепционная оценка качества речи PESQ (Perceptual Evaluation of Speech Quality), коэффициент искажений Итакура-Саито (IS), логарифмическое отношение правдоподобия LLR (Log Likelihood Ratio), мера различия взвешенного спектрального наклона WSS (Weighted Spectral Slope), а также комбинированная мера различия (4), представляющая собой линейную комбинацию мер LLR, PESQ и WSS [10].

$$C_{sig} = 3,093 - 1,029 \cdot LLR + 0,603 \cdot PESQ - 0,009 \cdot WSS, \quad (4)$$

Результаты объективных оценок качества воспроизводимой вокодерами речи при различных способах квантования коэффициентов отражения приведены в табл. 1.

Таблица 1

Результаты объективной оценки

Квантование коэффициентов отражения	PESQ	IS	LLR	WSS	C_{sig}
Скалярное (стандарт LPC-10)	2,65	3,24	1,06	52,77	3,13
Векторное, 12 бит	2,24	3,59	1,25	59,33	2,62
Векторное, 10 бит	2,20	3,78	1,28	60,25	2,56

Для мер IS, LLR и WSS значения, которые дальше от нуля, означают меньшее качество воспроизводимой речи. А для мер PESQ и C_{sig} меньшему качеству воспроизводимой речи соответствуют значения, которые ближе к единице.

Сравнение скоростей передачи бит для стандартного вокодера LPC-10 и вокодера с 10-ти битовым и 12-ти битовым векторным квантованием приведены в табл. 2.

Таблица 2

Сравнение скорости передачи бит

Квантование коэффициентов отражения	Тип скорости передачи	Наличие битов защиты	Кол-во бит в кадре	Кол-во нейронов	Скорость передачи, бит/с
Скалярное (стандарт LPC-10)	Постоянная	Да	54	–	2400
	Переменная	Нет	54 и 33	–	1933
Векторное, 12 бит	Постоянная	Нет	25	4096	1111
Векторное, 10 бит	Постоянная	Нет	23	1024	1022

Таким образом, при незначительном снижении качества воспроизводимой вокодером речи (в среднем около -10% для комбинированной меры различия), скорость передачи битов снижена больше чем 2 раза: до 1111 кбит/с при 12 битовом квантовании и при 10 битовом – до 1022 кбит/с.

Заключение. В данной работе подтверждена эффективность использования нейросетевого подхода для решения задачи векторного квантования коэффициентов отражения на примере стандарта DoD FS 1015 LPC-10 при вокодировании методом линейного предсказания. При незначительном снижении качества воспроизводимой вокодером речи, была существенно снижена скорость передачи бит, что в свою очередь позволяет использовать такой вокодер для каналов связи с меньшей пропускной способностью. Произведена объективная оценка качества воспроизводимой речи несколькими наиболее широко используемыми в речевых технологиях способами. Выполнено сравнение стандартного вокодера LPC-10 и вокодера на основе модели LPC-10, квантование коэффициентов отражения в котором производится с помощью нейросети Кохонена. Сравнение выполнено на основе качества воспроизводимой речи и на основе скорости передачи бит.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *W.C. Chu*. Speech coding algorithms: foundation and evolution of standardized coders. John Wiley & Sons, Hoboken, NJ, 2003.
2. *Y. Linde, A. Buzo and R.M. Gray*. An algorithm for vector quantizer design. IEEE Transactions on Communications, 1980, vol. 28, pp. 84-94.
3. *D.J. Vaisey, A. Gersho*. Simulated annealing and codebook design. Proc. ICASSP-88, April 1987, pp. 1176-1179.
4. *T. Kohonen*. The self-organizing map. Proc. IEEE, special issue on Neural Networks I, September 1990, pp. 1464-1480.
5. *C. Pope, L. Atlas, C. Nelson*. A comparison between neural network and conventional vector quantization codebook algorithms. Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C., Canada, June 1-2, 1989, pp. 521-524.
6. *Чернухин Ю.В., Лунев А.В.* Применение нейросети Кохонена для сжатия речевых данных в телекоммуникационных системах. Доклад. В кн. Информационные технологии, системный анализ и управление // Материалы третьей Всероссийской научной конференции. Сборник трудов. – Таганрог: Изд-во ТРТУ, 2005. – С. 16-20.
7. Federal Standard 1015, Telecommunications: Analog to digital conversion of radio voice by 2400 bit/second linear predictive coding, National Communication System - Office Technology and Standards, Nov. 1984.
8. *K.K. Paliwal and B.S. Atal*. Efficient vector quantization of LPC parameter at 24 bits/frame. IEEE Trans. Speech Audio Process., Jan. 1993, vol.1, no.1, pp.3-14.
9. *A.S. Spanias*. Speech Coding: A tutorial review. Portions published in Proceedings of the IEEE, Oct. 1994.
10. *Y. Hu and P. Loizou*. Evaluation of objective measures for speech enhancement. Proceedings of INTERSPEECH-2006, Philadelphia, PA, September 2006.