

# НЕЙРОСЕТЕВОЙ КВАНТОВАТЕЛЬ ПАРАМЕТРОВ НИЗКОСКОРОСТНОГО ВОКОДЕРА С МОДЕЛЬЮ РЕЧЕОБРАЗОВАНИЯ “ТОН+ШУМ”

Серков В.В.<sup>1</sup>, Петровский А.А.<sup>2</sup>

Государственный университет информатики и радиоэлектроники,  
ул. П.Бровки 6, Минск, Беларусь, кафедра электронных вычислительных средств  
<sup>1</sup>e-mail: [sercov@gmx.net](mailto:sercov@gmx.net) <sup>2</sup>e-mail: [palex@it.org.by](mailto:palex@it.org.by)

**Реферат.** В докладе представляется метод квантования параметров для низкоскоростного вокодера на базе модели речеобразования тон+шум, использующий искусственную нейронную сеть. Описаны структура сети и методика ее тренировки. Предложенный метод позволяет эффективно учитывать психоакустические особенности слуха. По сравнению с традиционно применяемым векторным квантованием обеспечивается более полное использование статистических закономерностей речевого сигнала и, следовательно, лучшее качество восстановленной речи.

## 1. Введение

В последние годы в области обработки и передачи речевых сигналов отмечался значительный интерес к методам спектрального кодирования, позволяющим получать хорошее качество восстановленной речи при скоростях передачи ниже 6 кбит/с. В спектральных вокодерах можно выделить две относительно независимые части: анализа–синтеза и квантования. Разработанные к настоящему времени модели речеобразования обеспечивают синтез речи, практически не отличающейся от естественной. При представлении параметров этих моделей ограниченным количеством битов с целью хранения или передачи по каналам связи вносятся существенные искажения. Таким образом, для низкоскоростного кодирования речи схема квантования параметров оказывает определяющее влияние на качество восстанавливаемого сигнала.

Наиболее простой в реализации является схема, предполагающая независимое квантование отдельных параметров. Такой подход называется скалярным квантованием. Он не позволяет учитывать корреляцию между параметрами; часть потока битов расходуется на передачу избыточной информации. Более эффективна в этом смысле схема векторного квантования [0], при которой набор параметров модели в некоторый момент времени квантуется совместно как один вектор. Основной ее недостаток – высокая вычислительная стоимость поиска в кодовой книге – заставляет на практике применять компромиссные варианты. Например, в вокодере, использующем Frequency Bin Vector Quantization [2], имеется несколько независимых векторных квантователей, а часть параметров квантуется скалярно. Распределение битов под отдельные параметры фиксировано для данного режима работы и, очевидно, не может быть оптимальным для любого сигнала. Другой недостаток схемы векторного квантования – сложность учета особенностей восприятия звука человеком, изучаемых психоакустикой. В работе [2] удалось задействовать лишь одну из них, расположив векторы в кодовой книге частот в соответствии со шкалой ERB.

В настоящей работе предлагается новый способ квантования параметров для низкоскоростного вокодера, основанный на применении аппарата искусственных нейронных сетей. Он ориентирован на модель речеобразования с отдельным представлением спектров тональной и шумовой компонент [3] и позволяет обойти ограничения, присущие векторным квантователям.

## 2. Метод квантования

Искусственные нейронные сети (ИНС) успешно используются для решения трудно формализуемых задач, к которым относится квантование. Хотя в обработке речи они и применялись ранее ([4]), авторам не известны примеры реализации квантователя на ИНС.

Основная идея сжатия данных с помощью ИНС была предложена в [5] и заключается в следующем. Рассмотрим трехслойный перцептрон, у которого число элементов входного и выходного слоев одинаково, а число элементов скрытого слоя значительно меньше. Предположим, что в результате обучения на примерах сеть может воспроизводить на выходе тот же самый вектор, который подается на ее входной слой. Такой репликативный перцептрон автоматически осуществляет сжатие информации: на элементах скрытого слоя возникает представление каждого вектора, которое значительно короче, чем длина вектора, подаваемого на вход. Для передачи по линии связи поместим на передающем конце входной и скрытый слой перцептрона, а результат работы элементов скрытого слоя (короткие векторы) будем передавать по линии. На другом конце линии поместим копию скрытого слоя и выходной слой, тогда переданный короткий вектор с элементов скрытого слоя перейдет на элементы выходного слоя, где будет воспроизведен исходный вектор (декомпрессия).

В работе [6] выполнен более строгий математический анализ. Показано, что в процессе обучения перцептрона с тремя скрытыми слоями можно построить систему <естественных> координат, которая соответствует обучающей выборке. Доказана единственность этой системы. <Естественные> координаты обладают рядом важных свойств: признаки объектов в таких координатах являются попарно независимыми; в результате перехода к системе <естественных> координат происходит не только значительное сокращение размерности задачи, но и фильтрация шума, который мог присутствовать в исходных данных.

При использовании репликативного перцептрона в качестве квантователя вокодера на него можно возложить и задачу предсказания сигнала. При этом принципиальное отличие от схем, традиционно использующих LPC, состоит в том, что квантованию подлежит не спектр сигнала-остатка, а спектральные параметры исходного сигнала с учетом их значений на предыдущих фреймах. При тренировке ИНС «сама» выявит статистические взаимосвязи между соседними фреймами и в рабочем режиме будет осуществлять нелинейное предсказание.

Психоакустическая оптимизация в предлагаемом методе заключается в двух аспектах. Во-первых, выполняется предварительное преобразование спектров тональной и шумовой компонент в координаты сон-барк [7] и на квантователь подаются огибающие спектров в этих координатах. Такой подход обеспечивает перцептуальную линейность схемы квантования: при ограничении количества разрядов для представления параметров погрешности равной величины, привнесенные на разных частотах и амплитудах, на слух будут восприниматься равнозначными. Во-вторых, в процессе обучения ИНС вектор ошибки, получаемый в результате предъявлении сети каждого фрейма, взвешивается с помощью порогов маскирования [7]. Обученная сеть будет хранить информацию о правилах маскирования в своих коэффициентах и автоматически использовать ее при квантовании, что делает ненужным расчет порогов при работе вокодера.

### 3. Структура сети квантователя

Структурная схема ИНС квантователя показана на рис. 1. Она представляет собой пятислойный перцептрон. На вход сети подаются параметры очередного фрейма с выхода вокодера (частота основного тона и огибающие спектров тональной и шумовой компонент речевого сигнала) и нескольких предыдущих фреймов с выходного слоя самой сети. Для ускорения процесса обучения все элементы входного вектора предварительно нормируются к диапазону -1 ... 1. Соответственно, в декодере должна выполняться обратная операция.

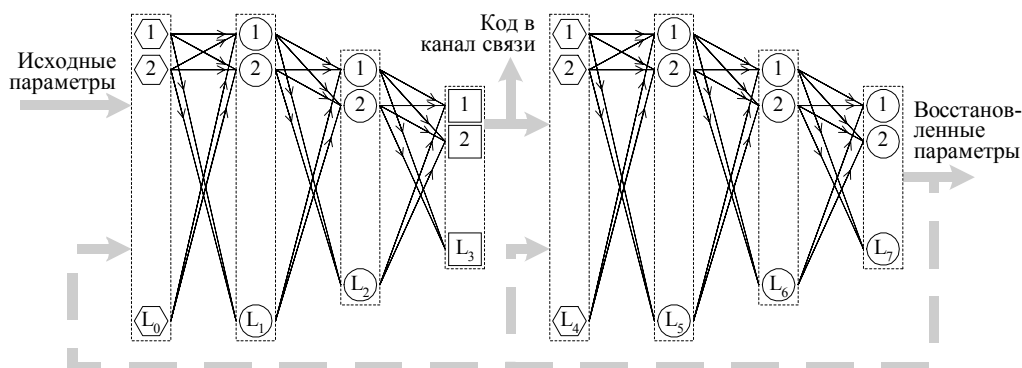


Рис.1. Структурная схема ИНС квантователя

Входной слой сети используется только для хранения и не выполняет никакой обработки. Выходные значения нейронов остальных слоев рассчитываются следующим образом:

$$y_{n,i} = f_n \left( \sum_{j=0}^{L_{n-1}} y_{n-1,j} w_{n,i,j} \right), \quad n = 1 \dots N, \quad i = 1 \dots L_n, \quad y_{n-1,0} = 1,$$

где  $n$  – номер слоя, содержащего данный нейрон;  $i$  – номер нейрона в слое;  $L_k$  – количество нейронов в  $k$ -м слое,  $w_{n,i,j}$  – весовые коэффициенты;  $f_n$  – функция активации нейронов слоя. Для всех слоев, кроме среднего в качестве функции активации выбран биполярный сигмоид:

$$f_n(x) = \frac{2}{1 + \exp(-x)} - 1,$$

а для среднего слоя, формирующего пакет битов для передачи, – биполярная ступенька:

$$f_n(x) = \begin{cases} 1 & \text{при } x \geq 0 \\ -1 & \text{при } x < 0. \end{cases}$$

#### **4. Тренировка ИНС квантователя**

В качестве основного рабочего материала были приняты тренировочные и тестовые данные, разработанные Массачусетским технологическим институтом, Стэндфордским исследовательским институтом и фирмой Texas Instruments (база данных TIMIT). Перед обучением весовые коэффициенты сети инициализировались малыми случайными значениями. Тренировка сети выполнялась по методу обратного распространения ошибки [8] с пошаговой коррекцией и с моментами. Величина шага обучения постепенно уменьшалась для обеспечения стабилизации сети.

#### **5. Заключение**

Предлагаемый метод позволяет использовать при квантовании параметров вокодера искусственные нейронные сети, хорошо зарекомендовавшие себя на задачах подобного класса. Продемонстрирована возможность эффективного учета психоакустических закономерностей. По сравнению с традиционно применяемым векторным квантованием обеспечивается более полное использование статистических закономерностей речевого сигнала. То, что ИНС одновременно с квантованием выполняет нелинейное предсказание параметров и расчет порогов маскирования, снижает вычислительную сложность остальной части вокодера.

Предварительные результаты показали перспективность предлагаемого подхода для передачи речи со скоростью 2,4 кбит/с. В настоящее время ведутся работы по определению оптимального количества нейронов в слоях сети и оценке качества квантования.

#### **Литература**

1. R.M.Gray, "Vector quantization", // *IEEE ASSP Mag.* 1(2), pp.4-29, 1984.
2. Ç.Ö.Etemoğlu, V.Cuperman, A.Gersho "Speech coding with an analysis-by-synthesis sinusoidal model", // *proc. of the IEEE Int. Conf. Acoust. Speech Signal Proc.*, ICASSP'2000, vol. III, Istanbul, pp. 1371-1374, 2000.
3. A.Petrovsky, V.Sercov, "Low bit-rate AbS spectral coding based on the harmonic analysis of speech agreed upon with time-varying pitch frequency and psychoacoustical optimization", // *proc. of the Nordic Signal Proc. Symp.*, NORSIG'2000, Sweden, pp.45-48, 2000.
4. W.W.Chang, C.T.Wang, "A masking-threshold-adapted weighting filter for excitation search", // *IEEE Trans. On Speech and Audio Proc.*, vol. 4, № 2, pp. 124-132, 1996.
5. G.W.Cottrell, P.Munro, D.Zipser, "Image compression by back propagation: An example of extensional programming" // *Proc. 9th Annual Confer, Cognitive Soc.*, pp. 461-473, 1987.
6. R.Hecht-Nielsen, "Replicator neural networks for universal optimal source coding", *Science*, v. 269, pp.1860-1863, 1995.
7. В.В.Серков, А.А.Петровский, "Использование закономерностей психоакустики при низкоскоростном кодировании речи", // Доклады 3-й межд. конф. «Цифровая обработка сигналов и ее применения», DSPA'2000, Москва, Россия, т.2, стр. 241-244, 2000.
8. D.E.Rumelhart, G.E.Hinton, R.J.Williams., "Learning internal representations by error propagation" // *In Parallel distributed processing*, vol. 1, pp. 318-62. Cambridge, MA: MIT Press, 1986 .