УДК 621.396.669

С.Н. Кириллов, Е.С. Попова

НЕЙРОСЕТЕВАЯ РЕАЛИЗАЦИЯ КОДЕРА РЕЧЕВЫХ СИГНАЛОВ АДАПТИВНОГО К УРОВНЮ АКУСТИЧЕСКИХ ШУМОВ

Рассматриваются вопросы, связанные с проектированием адаптивного к уровню акустических шумов кодера источника речевых сигналов с μ-компандированием на основе искусственной нейронной сети. Удалось достигнуть снижение уровня дисперсии акустического шума на выходе нейросетевой реализации кодера источника сообщения в десятки раз по сравнению со стандартным кодеком при увеличении отношения сигнал-шум от 7 до 23 дБ в случае различной дисперсии акустического шума, без снижения субъективной оценки качества речевого сигнала по шкале MOS.

Ключевые слова: *искусственная нейронная сеть, речевой сигнал, кодер, μ-компандирование, акустические шумы.*

Введение. Речевые сигналы (РС), с которыми приходится иметь дело на практике, всегда в той или иной степени подвержены действию акустических шумов (АШ). В тех случаях, когда шум имеет значительную интенсивность, его наличие может существенно исказить результаты обработки, анализа или распознавания речи. В целом ряде других случаев, например, при анализе зашумленных РС в криминалистических целях или восстановлении аудиозаписей в архивах, задача уменьшения негативного воздействия АШ на РС носит самостоятельный характер и является единственной целью работы. Поэтому разработка алгоритмов, снижающих влияние АШ на РС, является весьма актуальным направлением исследований.

Известны различные алгоритмы подавления АШ при обработке РС [1]. Большинство существующих алгоритмов шумоподавления реализуются в частотной области и используют различные варианты метода спектрального вычитания [2, 3]. Основным недостатком данных методов является появление в отфильтрованном РС искажений, известных как «музыкальные тона». Помимо этого может быть использован подход обработки зашумленного РС в подпространствах, который является обобщением методов спектрального взвешивания. Оценка параметров РС в данном алгоритме рассматривается как задача оптимизации с ограничениями, где искажения РС минимизируются с учётом остаточной мощности шума. При этом оценка качества отфильтрованной речи выполняется с использованием ряда объективных и субъективных показателей.

Применение данных алгоритмов шумоподавления при цифровой обработке РС осуществляется в случае наличия параллельно включенного АЦП, который позволяет представить непрерывный поток данных в цифровом виде для дальнейшей обработки на ЭВМ. Данный факт является существенным техническим и экономическим недостатком применения таких алгоритмов [4]. Кроме того, известно, что в настоящее время для микропроцессоров наступает так называемый «технологический предел», заключающийся в том, что они достигли максимального уровня повышения быстродействия.

Одним из способов решения данной проблемы может быть использование новой элементной базы, например, на основе искусственных нейронных сетей (ИНС). Нейронные сети представляют собой весьма перспективную вычислительную технологию, дающую новые подходы к исследованию различных динамических задач. Свойство толерантности, присущее для ИНС, позволяет находить решения, робастные к различным видам искажений. Способность к моделированию нелинейных процессов, работе с зашумленными данными и адаптивность [5] дают возможность применению ИНС при решении широкого класса задач [6].

Цель работы – проектирование и исследование нейросетевой реализации кодера источника сигнала при воздействии АШ.

Анализ статистических характеристик АШ. Акустический сигнал, поступающий через микрофон на вход системы цифровой обработки, практически всегда содержит в себе не только РС, но и различного рода АШ. Отрицательное

влияние АШ на РС проявляется в уменьшении разборчивости и ухудшении качественных характеристик речи. Особенно сильно данный эффект проявляется в цифровых системах обработки речи, так как приводит к дополнительным нелинейным искажениям РС.

При экспериментальных исследованиях изучалось влияние АШ на РС, создаваемых автотранспортом вблизи дороги, и АШ внутри автомобилей марок ВАЗ (Lada) 2112, ВАЗ (Lada) 2190, Hyundai Sonata. Запись АШ проводилась с частотой дискретизации 40 кГц. В дальнейшем для корректного наложения данных АШ на исходный РС осуществлялась фильтрация записанных аудио файлов в полосе частот 0.3-3.4 кГц и децимация до частоты дискретизации 8 кГц.

Проводился анализ спектральных и статистических характеристик данных АШ. Примеры частотных спектров АШ представлены на рисунке 1.

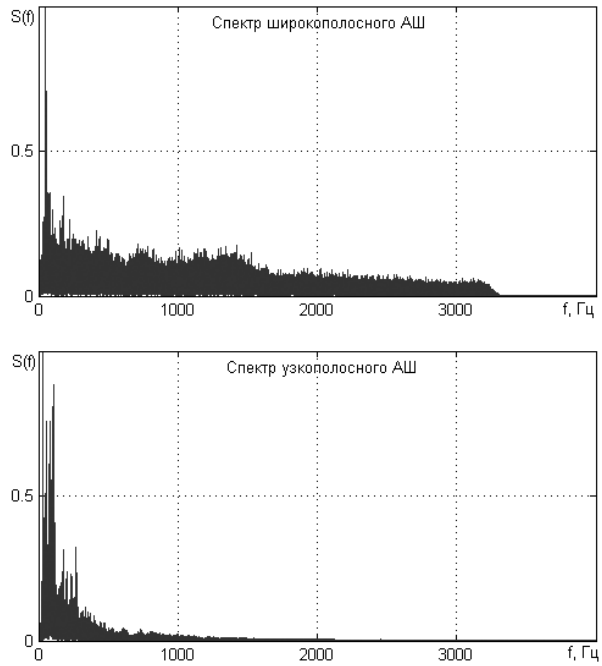

Рисунок 1 – Частотные спектры АШ, используемых в экспериментальном исследовании

Показано, что АШ, создаваемый автотранспортом около дороги, является широкополосным, так как его спектр относительно равномерно распределен в полосе частот 0.3-3.4 кГц, а АШ внутри автомобиля узкополосный, так как его спектр расположен в полосе частот 0.3-1.5 кГц.

На рисунке 2 приведены функции плотности вероятности (ФПВ) исследуемых АШ.

Проанализировав полученные результаты по критерию согласия χ^2, было показано, что данные ФПВ исследуемых АШ могут быть аппроксимированы нормальным законом распределения.

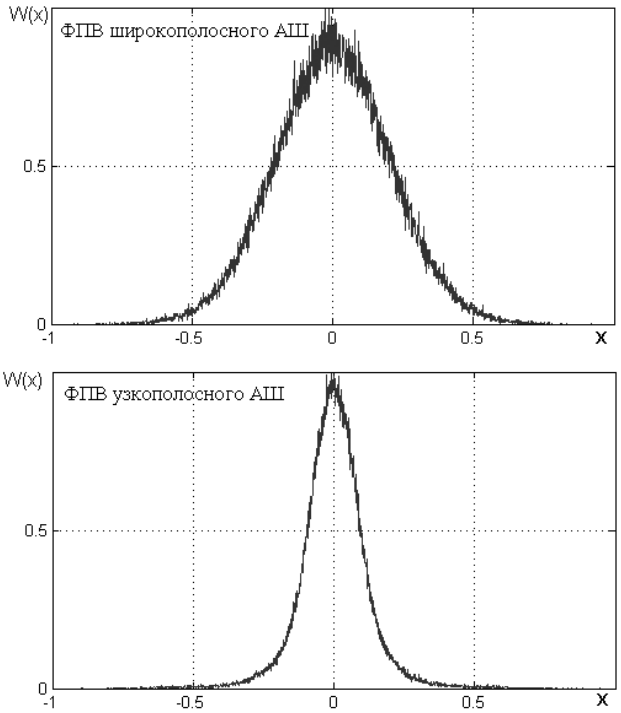

Рисунок 2 – ФПВ исследуемых АШ

Алгоритм обучения ИНС. Для решения какой-либо задачи с применением ИНС следует, прежде всего, спроектировать структуру сети, адекватную поставленной задаче. Для реализации кодера источника РС была выбрана нейронная сеть типа «многослойный персептрон». Данный тип архитектуры нейронной сети является классической многослойной сетью с полными последовательными связями нейронов. На сегодняшний день многослойный персептрон является наиболее простой в реализации сетью, а также он находит множество применений в различных прикладных задачах, например, таких как обработка речевых сигналов [7, 8], анализ изображений [9, 10], экспертных системах [11] и т.д. Основное преимущество многослойного персептрона – это возможность решать трудно формализуемые задачи или задачи, для которых алгоритмическое решение неизвестно, но для которых возможно составить набор примеров с известными решениями. При обучении ИНС за счёт своего внутреннего строения выявляет закономерности и связи входных и выходных образов. Таким образом, ИНС типа «многослойный персептрон» позволяет наиболее точно аппроксимировать выходные данные при обучении[12]. Обучение ИНС проводилось по методу Левенберга-Марквардта. Данный метод может быть представлен как комбинация методов наискорейшего спуска и Гаусса – Ньютона, которые являются примером способа быстрой оптимизации обучения [13]. Главными достоинствами данного алгоритма являются скорость обучения

и отсутствие необходимости в указании критериев остановки обучения [14]. В процессе проектирования была проведена оптимизация структуры ИНС по критерию минимума СКО обучения. При оптимизации происходило изменение количества скрытых слоев, количества нейронов в слоях и наклона сигмоидальной функции активации нейронов.

Оптимальная структура ИНС включала в себя:

- количество входов – 1;
- количество выходов – 1;
- количество скрытых слоев – 2;
- количество нейронов в первом скрытом слое – 10;
- количество нейронов во втором скрытом слое – 10;
- вид активационной функции – гиперболический тангенс.

Обоснование структуры нейросетевой реализации кодера источника РС. Сравнительный анализ известного 8-разрядного кодера с μ-компанированием по стандарту G.711 μ-Law, реализованного на ИНС, показал возможность подавления АШ в ИНС. При этом оказалось, что уровень подавления АШ существенно зависит от параметров ИНС. Поэтому возникла необходимость в дополнительном исследовании адаптивного кодера на основе ИНС с различными весовыми коэффициентами и векторами смещения, которые бы обеспечивали различный порог подавления АШ. Если за условие перестройки параметров ИНС выбрать полное подавление АШ, то будет происходить существенное искажение РС, а при больших уровнях АШ полная потеря сообщения. В связи с этим критерием перестройки параметров ИНС было предложено использовать значение СКО АШ σ, что позволяло подавить примерно 68,2 % шума, при условии что его плотность распределения вероятности хорошо аппроксимируется нормальным распределением. Таким образом, был введен порог подавления ИНС p=σ. Для определения значения порога подавления необходимо в паузах речи оценивать дисперсию АШ.

Таким образом, структурная схема перестраиваемого кодера источника РС на основе ИНС имеет вид, показанный на рисунке 3.

На вход кодера источника сообщения на основе ИНС поступали отсчеты зашумленного РС. Далее сигнал обрабатывался устройством определения пауз, которое было реализовано с помощью алгоритма Voice Activity Detection (VAD) [15]. В использованном методе реализации алгоритма VAD РС разбивался на фреймы, а затем для каждого фрейма одновременно вычислялись три характеристики РС: краткосрочная энергия [16], мера спектральной плоскостности и составляющая фрейма речи с преобладающими частотами [17]. Это позволяло наиболее точно определить паузу в РС даже при очень низких отношениях сигнал-шум (ОСШ). В паузах РС определялось СКО шума, данное значение поступало на ПЗУ, в котором хранились весовые коэффициенты и векторы смещения ИНС. В зависимости от уровня СКО шума происходило изменение порога подавления кодера. Затем РС с подавленным АШ поступал на декодер, где сравнивался с исходным не зашумленным РС.

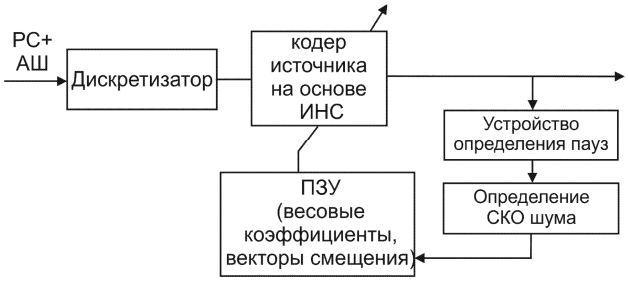

Рисунок 3 – Структурная схема нейросетевой реализации кодера РС

Анализ результатов, полученных в процессе компьютерного моделирования. Для проведения экспериментального исследования нейросетевой реализации кодера источника сообщения был использован РС, содержащий в себе акустически взвешенные фразы, представленные в ГОСТ Р 51061-97. АШ накладывались на РС с различной дисперсией шума. При исследованиях было показано, что погрешность квантования обученной ИНС при отсутствии АШ не превышала погрешности квантования стандартного кодера G.711. Исследовались зависимости ОСШ q зашумленного РС от уровня СКО АШ σ. Для этого на исходный РС аддитивно накладывались различные АШ с уровнем нормированного СКО от 0.01 до 0.12. Нормировка осуществлялась к СКО РС. После чего для сравнения зависимостей ОСШ от уровня СКО АШ данные зашумленные РС пропускались через нейросетевую реализацию кодера источника сообщения, приведенную на рисунке 3, и через стандартный кодер, результаты сравнения приведены на рисунке 4.

Из анализа рисунка 4 следует, что нейросетевая реализация кодера позволяет увеличить ОСШ от 7 до 23 дБ.

Соответствующие выбросы на рисунке 4 связаны с дискретной перестройкой порога подавления АШ. Можно отметить, что степень подавления АШ не существенно зависит от типа АШ и его спектральных характеристик.

Для оценки дисперсии декодированного шума $D_{дкш}$, АШ с различным уровнем дисперсии: от 0 до 0,01 проходил через схему, представленную на рисунке 3. в данном случае АШ

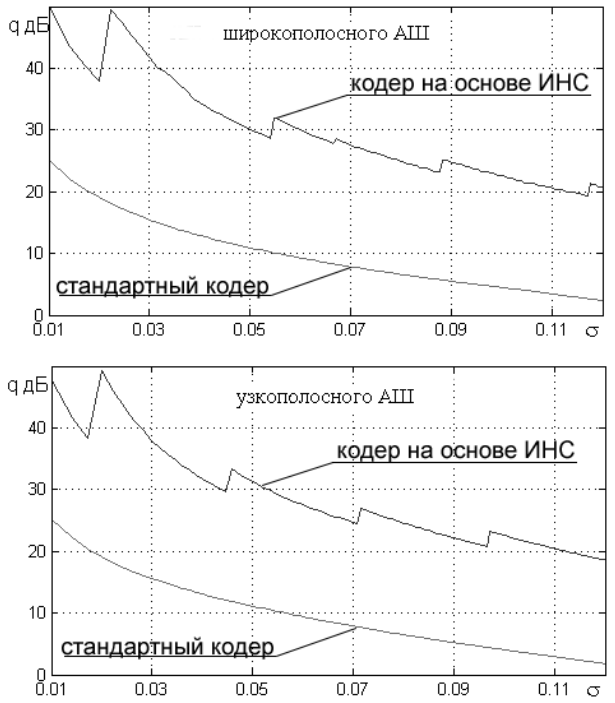

Рисунок 4 – Зависимости ОСШ декодированного сигнала при нейросетевой реализации кодера РС по сравнению со стандартным кодером

не накладывался на РС для более точного определения его параметров. В результате экспериментальных исследований было отмечено снижение уровня дисперсии шума $D_ш$ на выходе проектируемого устройства по сравнению со стандартным кодером источника сообщения в 10 – 100 раз.

Для получения субъективной оценки качества сигнала по методу MOS [18] была произведена запись РС шестью дикторами, которые начитывали акустически взвешенные фразы, прописанные в ГОСТ Р 51061-97. После этого каждый АШ аддитивно накладывался на РС с различным ОСШ – от 0 до 40 дБ. Полученные РС кодировались стандартным кодером источника сообщения и с помощью нейросетевой реализации перестраиваемого кодера. Десять аудиторов производили субъективную оценку прослушиваемых РС по пятибалльной шкале [19]:

5 – понимание речи без малейшего напряжения внимания;

4 – понимание речи без затруднений;

3 – понимание речи с напряжением внимания без переспросов и повторений;

2 – понимание речи с некоторым напряжением внимания, редкими переспросами и повторениями;

1 – понимание речи с большим напряжением внимания, частыми переспросами и повторениями.

Оценки аудиторов складывались, а затем находилось среднее значение субъективной оценки по методу MOS.

В таблице 1 и таблице 2 приведены средние значения субъективных оценок по шкале MOS в зависимости от ОСШ.

Таблица 1 – Среднее значение субъективных оценок для широкополосного АШ

Среднее значение субъективных оценок для широкополосного АШ	ОСШ, дБ				
	0	10	20	30	40
Стандартный кодер	1,1	2,3	3,2	4,2	4,8
Кодер на основе ИНС	1,2	2,6	3,8	4,3	4,8

Таблица 2 – Среднее значение субъективных оценок для широкополосного АШ

Среднее значение субъективных оценок для широкополосного АШ	ОСШ, дБ				
	0	10	20	30	40
Стандартный кодер	1,1	2,1	3,1	3,9	4,6
Кодер на основе ИНС	1,1	2,3	3,4	4,2	4,6

Таким образом, удалось достигнуть увеличения субъективной оценки качества РС на 0.2-0.5 балла по шкале MOS, рисунок 5.

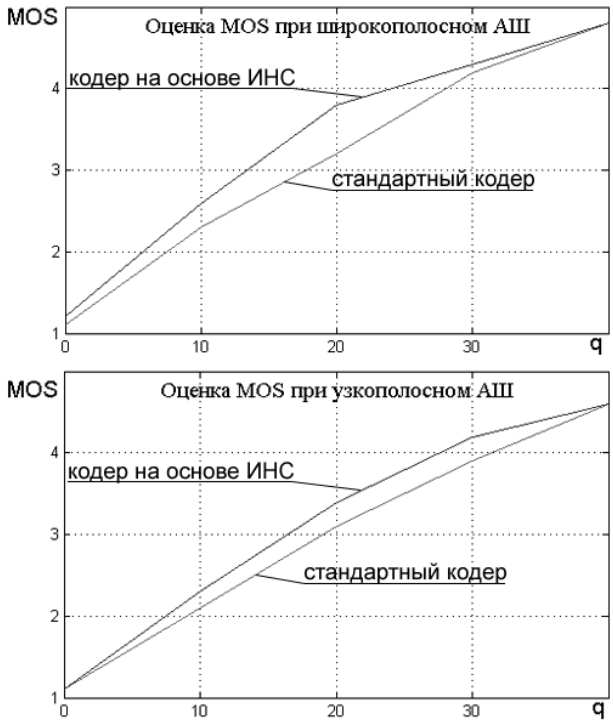

Рисунок 5 – Зависимости MOS от отношения сигнал-шум

Заключение. Проведенные исследования показали эффективность применения нейросетевой реализации кодера источника РС при действии АШ. Удалось достигнуть снижения уровня дисперсии шума на выходе кодера в десятки раз по сравнению со стандартным кодером при увеличении отношения сигнал-шум от 7 до 23 дБ в случае различной дисперсии шума. При этом не снижалась субъективная оценка качества РС по шкале MOS. Следует отметить, что полученные результаты существенно не зависят от типа шума и его спектральных характеристик, а также от исходного РС, что, в свою очередь, исключает эффект запоминания, свойственный для ИНС.

Библиографический список

1. *Gibak Kim, Phillips C Loizou.* Why do speech-enhancement algorithms not improve speech intelligibility // Processing of ICASSP-2010. Vol. 1. P. 397–400.
2. *Phillips C Loizou.* Speech enhancement theory and practice: 1st ed. Boca Raton, FL.: CRC,2007. Releases Taylor & Francis.
3. *J. Benesty, J. Chen, Y. Huang, I.* Cohen Noise Reduction in Speech Processing // Springer-Verlag, 2009.
4. *Злобин В.К., Григоренко Д.В., Ручкин В.Н., Романчук В.А.*. Кластеризация и восстанавливаемость нейропроцессорных систем обработки данных// Известия ТулГУ. Технические науки. 2013. Вып. 9. Ч.2 С. 125-133.
5. *Данилин С.Н., Макаров М.В., Щаников С.А.* Проектирование технических средств с нейросетевой архитектурой при искажении шумами входной информации. 24-я Международная Крымская конференция "СВЧ-техника и телекоммуникационные технологии": материалы конф.: в 2 т. - Севастополь, 2014.
6. *Осовский С.*Нейронные сети для обработки информации / пер. с польского И.Д. Рудинского. - М.: Финансы и статистика, 2002.
7. *Hinton G., Deng L., Yu D., Dahl G., Mohamed A., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T. and Kingsbury B.* Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE Signal Processing Magazine, Vol. 29, No. 6, 2012. P. 82 – 97.
8. *Болодурина, В.Н. Решетников, М.Г. Таспаева.* Применение и адаптация нейросетевых технологий в задаче идентификации динамических объектов. // Программные продукты, системы и алгоритмы. № 1. 2012.
9. *Ciresan D., Meier U., Masci J and Schmidhuber J.* Multi-column Deep Neural Network for Traffic Sign Classification. Neural Networks, Vol. 34, August 2012, P. 333 – 338И.П.
10. *Карасев О.Е.* Применение теории нечётких множеств для обработки видеоинформации в телекоммуникационных системах. VI Всероссийские научные Зворыкинские чтения: сб. тез. докл. Всероссийской межвузовской научной конференции. Муром. - Муром: Изд. - полиграфический центр МИ ВлГУ, 2014.- 791 с.
11. *Еремин Д. М., Гарцеев И. Б.* Искусственные нейронные сети в интеллектуальных системах управления. — М.: МИРЭА, 2004. — 75 с.
12. *Ясницкий Л. Н.* Введение в искусственный интеллект. — М.: Издат. центр «Академия», 2005. — 176 с. — ISBN 5-7695-1958-4.
13. *Martinetz M., Berkovich S., Schulten K.* “Neural-gas” network for vector quantization and its application to time scries prediction H Trans. Neural Networks, 1993.- Vol. 4.
14. *Gill P. Murray W., Wright M.* Practical Optimization. - N.Y.: Academic Press, 1987.
15. *J. Sohn, N. S. Kim, and W. Sung.* A statistical model-based voice activity detection. IEEE Signal Processing Lett., 6 (1): 1–3, 1999.
16. *Ephraim, Y. & Malah, D.* Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator IEEE Trans Acoustics Speech and Signal Processing, 32(6):1109-1121, Dec 1984
17. *Rainer Martin.* Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech and Audio Processing, 9(5):504-512, July 2001.
18. Рекомендация МСЭ-Т Р.80/Р.800.
19. ГОСТ Р 51061-97 Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений. – Введ. 01.01.98. – М. : Госстандарт России.