

УДК 621.396.669

ОБРАБОТКА И КОДИРОВАНИЕ РЕЧЕВЫХ СИГНАЛОВ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

*Кириллов С.Н., зав. кафедрой РУС РГРТУ, д.т.н., профессор, e-mail: kirillov.lab@mail.ru;
Сазонова Е.С., магистрант РГРТУ, e-mail: katrinik94@mail.ru.*

PROCESSING AND CODING VOICE SIGNAL USING ARTIFICIAL NEURAL NETWORKS

Kirillov S.N., Sazonova E.S.

The algorithm of noise reduction in speech signals, and the theoretical and practical aspects of design and optimization structures of neural network implementations predictors of speech signals. Managed to achieve a decrease in the level of dispersion of acoustic noise dozens of time. Justify bath implementation advantages predictors based on artificial neural networks compared with known predictors based on non-recursive FIR filters. In proved possible to reduce the prediction order of 10 % to 60 % at the same error, and the prediction error reduction from 15 % to 70 % at the same manner.

Key words: artificial neural network, perceptron, linear regression network, FIR filter, predictor, voice signals, algorithm of noise reduction.

Ключевые слова: искусственные нейронные сети, перцептрон, линейная регрессионная сеть, КИХ-фильтр, предсказатель, речевые сигналы, алгоритмы шумоподавления.

Введение

В настоящее время разработка новых алгоритмов обработки и передачи речевых сигналов (РС) в телекоммуникационных системах вызывает значительный интерес. При этом РС, с которыми приходится иметь дело на практике, всегда в той или иной степени подвержены действию акустических помех (АП). В тех случаях, когда АП имеют значительную интенсивность, их 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 АП на РС носит самостоятельный характер и является единственной целью работы. Поэтому разработка алгоритмов снижающих влияние АП на РС является весьма актуальным направлением исследований.

Применение известных алгоритмов подавления АП при цифровой обработке РС, осуществляется в случае наличия параллельного канала при включенном на входе АЦП, который позволяет представить непрерывный поток данных в цифровом виде, для его 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, на-

Предложен алгоритм подавления акустических помех (АП) в кодерах речи, а так же показана возможность реализации предсказателя речевых сигналов (РС) в системах кодирования речи на основе искусственных нейронных сетей (ИНС). Удалось достигнуть значительного снижения уровня дисперсии акустического шума на выходе системы кодирования РС, при изменении отношения сигнал-шум от 7 до 23 дБ, без снижения субъективной оценки качества речи по шкале MOS. Показаны преимущества реализации предсказателей РС на базе ИНС по сравнению с известными предсказателями на основе КИХ-фильтров. Доказана возможность уменьшения порядка предсказания от 10 % до 60 % при той же ошибке и снижения ошибки предсказания от 15 % до 70 % при том же порядке в случае реализации предсказателей на основе ИНС. Достигнуто увеличение субъективной оценки качества РС по шкале MOS до 0,5 балла, что является существенным преимуществом систем кодирования на основе ИНС.

пример, 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 к работе с зашумленными данными и адаптивность [5] дают возможность применению ИНС при решении широкого класса задач [6, 7].

Помимо обработки РС в усовершенствовании также нуждается процесс кодирования РС. Одним из основных способов первичного кодирования является алгоритм адаптивной дифференциальной импульсно-кодовой модуляции (АДИКМ). Для реализации линейного предсказателя в системах АДИКМ чаще всего используются нерекурсивные КИХ-фильтры, точность предсказания которых ограничена и определяется порядком фильтра и точностью оценки коэффициентов КИХ-фильтра b_k .

Для увеличения точности предсказания могут быть использованы ИНС, которые широко применяются в ситуациях, когда невозможно точно установить зависимость между входными и выходными данными, как например, при задаче предсказания отсчетов РС.

Таким образом, **целью работы является** разработка и исследование алгоритма подавления АП, а также анализ возможности реализации предсказателя РС в кодерах речи на базе ИНС.

Анализ статистических характеристик АП

Акустический сигнал, поступающий через микрофон на вход системы цифровой обработки, практически всегда содержит в себе не только РС, но и различного рода АП. Отрицательное влияние АП 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 анализировалось влияние АП на РС, создаваемых автотранспортом вблизи дороги, а также АП внутри автомобилей марок ВАЗ (Lada) 2112, ВАЗ (Lada) 2190, HyundaiSonata. Запись АП проводилась с частотой дискретизации 40 кГц. В дальнейшем, для корректного наложения данных АП на исходный РС, осуществлялась фильтрация записанных аудио файлов в полосе частот 0,3-3,4 кГц, и децимация отсчетов до частоты дискретизации 8 кГц.

Проводился анализ спектральных и статистических характеристик данных АП. Примеры частотных спектров АП представлены на рис. 1.

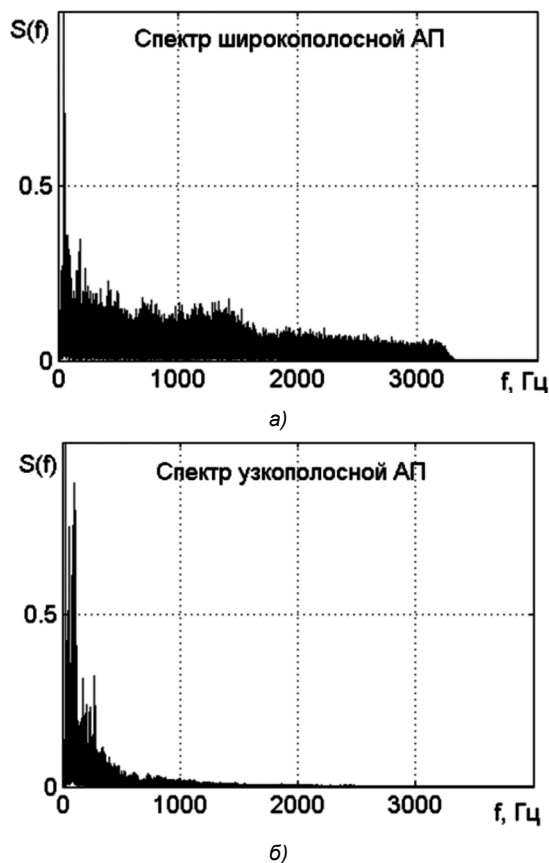


Рис. 1. Частотные спектры АП, используемых в экспериментальных исследованиях

Показано, что АП, создаваемые автотранспортом около дороги, являются широкополосными, так как их спектр относительно равномерно распределен в полосе частот 0,3-3,4 кГц, а АП внутри автомобиля – узкополосными, при этом их спектр расположен в полосе частот 0,3-1,5 кГц.

На рис. 2 приведены функции плотности вероятности (ФПВ) исследуемых АП.

Анализ полученных результатов по критерию согласия χ^2 показал, что данные ФПВ исследуемых АП могут быть аппроксимированы нормальным законом распределения.

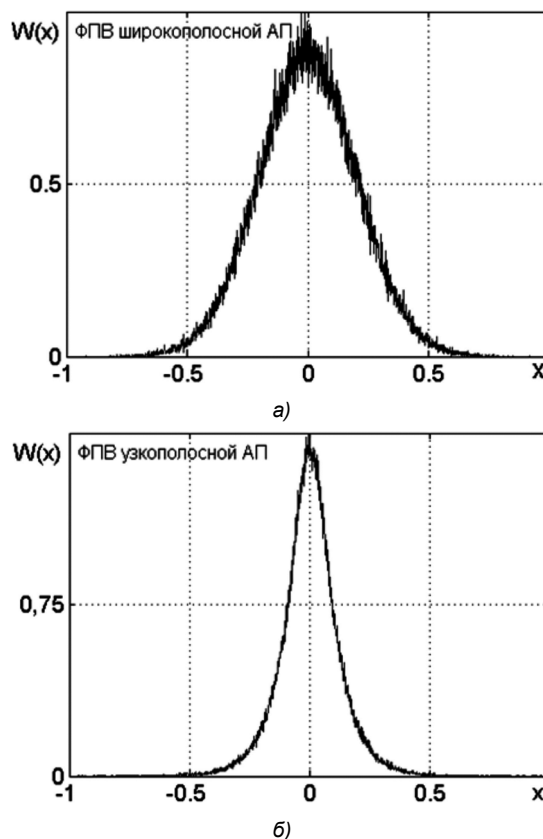


Рис. 2. ФПВ исследуемых АП

Структура проектируемой ИНС

Для решения какой-либо задачи с применением ИНС следует, прежде всего, спроектировать структуру сети, адекватную поставленной задаче. Для реализации алгоритма подавления АП была выбрана нейронная сеть типа «многослойный перцептрон». Данный тип архитектуры нейронной сети является классической многослойной сетью с полными последовательными связями нейронов. При обучении ИНС, за счёт своего внутреннего строения, выявляются закономерности и связи входных и выходных образов. Таким образом, ИНС типа «многослойный перцептрон» 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 высокая скорость обучения и отсутствие необходимости в указании критериев остановки обу-

чения [14]. В процессе проектирования была проведена оптимизация структуры ИНС по критерию минимума среднеквадратического отклонения (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 функцией – сигмоидальная.

Уровень подавления АП существенно зависит от параметров ИНС. В связи с этим критерием перестройки параметров ИНС было выбрано значение СКО АП – σ , что позволяло подавить примерно 68,2 % помех, при условии, что их плотность распределения вероятности хорошо аппроксимируется нормальным распределением. Таким образом, был введен порог подавления ИНС $p = \sigma$. Для определения значения порога подавления необходимо в паузах речи оценивать дисперсию АП.

Таким образом, структурная схема алгоритма подавления АП на основе ИНС имеет вид показанный на рис. 3.



Рис. 3. Структурная схема нейросетевой реализации алгоритма подавления АП

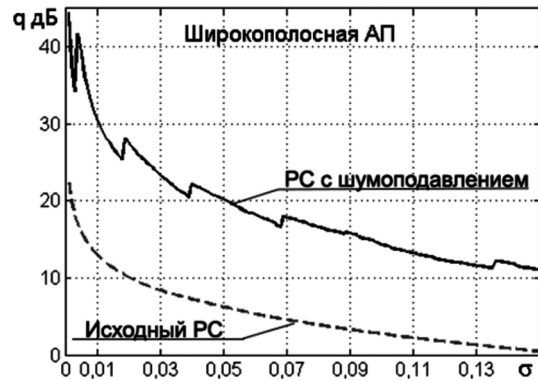
Для проведения экспериментального исследования нейросетевой реализации алгоритма подавления АП был использован РС содержащий в себе акустически взвешенные фразы представленные в ГОСТ Р 51061-97. АП накладывались на РС с различной дисперсией шума. При исследованиях было показано, что погрешность квантования обученной ИНС при отсутствии АП не превышала погрешности квантования стандартного кодера G.711. Исследовались зависимости ОСШ q зашумленного РС от уровня СКО АП σ . Для этого на исходный РС аддитивно накладывались различные АП с уровнем нормированного СКО от 0,01 до 0,12. Нормировка осуществлялась к СКО РС. После чего для сравнения зависимостей ОСШ от уровня СКО АП зашумленные данные РС пропускались через нейросетевую реализацию алгоритма шумоподавления, приведенную на рис. 3 и через стандартный кодер с μ -компаундированием без шумоподавления, результаты сравнения приведены на рис. 4.

Из анализа рис. 4 следует, что нейросетевая реализация алгоритма шумоподавления позволяет увеличить ОСШ от 7 до 23 дБ.

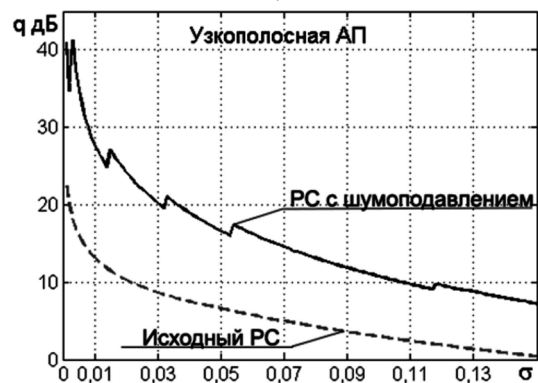
Соответствующие выбросы на рис. 4 связаны с дискретной перестройкой порога подавления АП. Можно отметить, что степень подавление АП не существенно

зависит от типа АП и его спектральных характеристик.

Для оценки дисперсии $D_{\text{дкш}}$ декодированного шума АП с различным уровнем дисперсии (от 0 до 0,01) проходили через схему, представленную на рис. 3. В данном случае АП не накладывались на РС для более точного определения его параметров. В результате экспериментальных исследований было отмечено снижение уровня дисперсии шума $D_{\text{ш}}$ на выходе проектируемого устройства по сравнению со стандартным кодером источника сообщения в 10 – 100 раз.



а)



б)

Рис. 4. Зависимости ОСШ декодированного сигнала при нейросетевой реализации кодера РС по сравнению со стандартным кодером

Для получения субъективной оценки качества речи по методу MOS [18] была произведена запись РС шестью дикторами, которые начитывали акустически взвешенные фразы, прописанные в ГОСТ Р 51061-97. После этого каждая АП аддитивно накладывалась 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 auditors производили субъективную оценку прослушиваемых РС по 5-ти бальной 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 auditors складывались, а затем находилось среднее значение субъективной оценки по методу MOS.

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

Таблица 1. Среднее значение субъективных оценок для широкополосных АП

Среднее значение субъективных оценок для широкополосных АП	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 узкополосных АП

Среднее значение субъективных оценок для узкополосных АП	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.

Архитектуры ИНС, используемые в качестве предсказателей РС, и их анализ

Известно [8], что для задач регрессии рекомендуется использовать такие типы архитектур ИНС, как многослойный перцептрон и линейная регрессионная сеть.

Многослойный перцептрон является классической многослойной сетью с полными последовательными связями нейронов с сигмоидальной функцией активации.

Линейная регрессионная сеть является частным случаем перцептрона, где в качестве активационной функции используется линейная функция с ограничениями, определяемая минимальными и максимальными значениями массива выходных данных. Для обеспечения динамики ИНС между скрытыми слоями добавлены обратные связи, позволяющие более точно устанавливать зависимости между входными и выходными данными [9].

Анализ ИНС типа многослойный перцептрон

Схема обучения ИНС типа многослойный перцептрон приведена на рис. 5.

В качестве входного массива данных был использован акустически взвешенные РС с частотой дискретизации 8 кГц, соответствующие ГОСТ Р 51061-97.

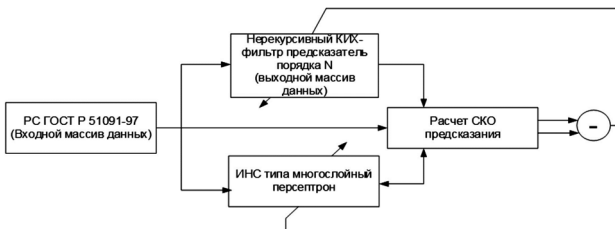


Рис. 5. Схема обучения ИНС типа многослойный перцептрон

В качестве выходного массива данных были выбраны выходные значения фильтра предсказателя соответствующего порядка. Обучение многослойного перцептрона проводилось по алгоритму Левенберга-Марквардта и обратного распространения ошибок, а также квази-Ньютоновским методом обратного распространения. Зависимости среднеквадратичной ошибки предска-

зания σ от порядка фильтра N изображены на рис. 6.

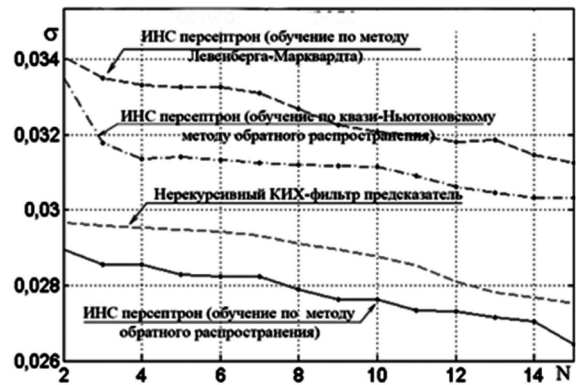


Рис. 6. Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для ИНС типа многослойный перцептрон

Анализ полученных зависимостей показал, что лучшие результаты достигаются при обучении по методу обратного распространения.

На рис. 7 изображена схема эксперимента.

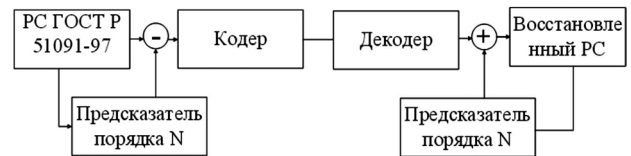


Рис. 7. Схема эксперимента

Эксперимент соответствует неадаптивному методу кодирования РС ДИКМ без передачи коэффициентов предсказания. Рассматривались два варианта реализации предсказателя на приемной стороне:

- с обучением ИНС как на передающей, так и на принимающей стороне, при этом значения векторов смещения и весов ИНС-предсказателей на приёмной и передающей сторонах различны;
- с обучением ИНС только на передающей стороне, значения векторов смещения и весов ИНС-предсказателя на приёмной стороне полностью соответствуют значениям векторов смещения и весов ИНС-предсказателя на передающей.

На рис. 8 изображены зависимости среднеквадратичной ошибки предсказания σ от порядка фильтра N при наличии и отсутствии обучения ИНС-предсказателя на приемной стороне.

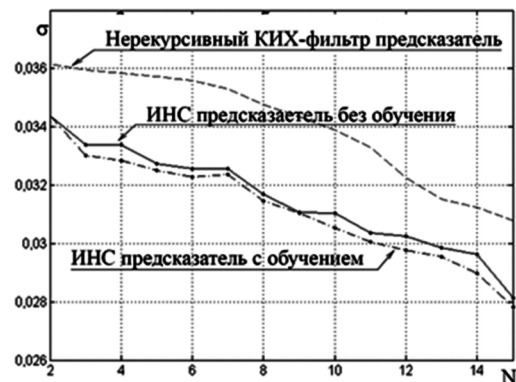


Рис. 8. Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для ИНС типа многослойный перцептрон

Из анализа рис. 8 следует, что существенных разли-

чий между схемами с обучением и без обучения ИНС-предсказателя на приемной стороне не наблюдается. Таким образом, можно использовать схему без обучения на приемной стороне, чтобы не усложнять нейросетевую реализацию предсказателя. Также очевидно преимущество предсказателя, спроектированного на базе ИНС типа многослойный перцептрон, перед известным нерекурсивным КИХ-фильтром предсказателем, что позволяет уменьшить порядок предсказателя от 20 % до 50 % при той же ошибке предсказания.

Анализ ИНС типа линейная регрессионная сеть

В отличие от многослойного перцептрона линейная регрессионная сеть не нуждается в обучении, так как самоорганизуется и адаптируется в зависимости от исходных входных и выходных массивов данных. Таким образом, данная сеть может быть использована для реализации адаптивного предсказателя. Эксперимент проводился в соответствии с рис. 7, при этом адаптация КИХ-фильтра осуществлялась по алгоритму Левенсона-Дарбина. В данном случае для реализации схемы эксперимента с помощью предсказателя на основе н КИХ-фильтра необходимо создать низкоскоростной канал для передачи коэффициентов предсказания. В связи с этим было рассмотрено две реализации организации предсказателя на базе линейно-регрессионной ИНС:

- с передачей коэффициентов весов и векторов смещения ИНС после адаптации на передающей стороне;
- с адаптацией ИНС на приемной стороне.

Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для линейной регрессионной ИНС приведены на рис. 9. Из анализа рис. 9 можно сделать вывод, что адаптивный ИНС-предсказатель обеспечивает более высокие результаты по сравнению с известным предсказателем на основе КИХ-фильтра, при этом не требуя организации дополнительного низкоскоростного канала для передачи коэффициентов предсказания, что является его существенным преимуществом.

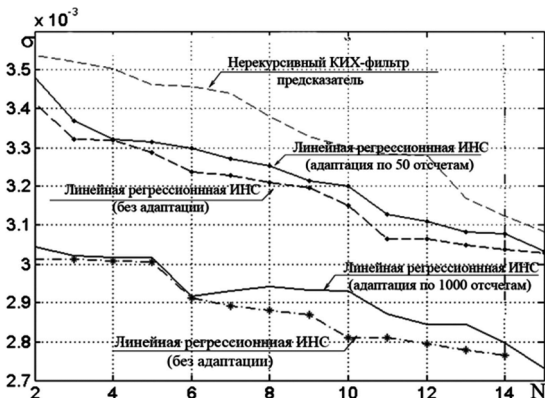


Рис. 9. Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для линейной регрессионной ИНС

На рис. 10 приведены зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для 50, 150, 500 и 1000 отсчетов адаптации РС.

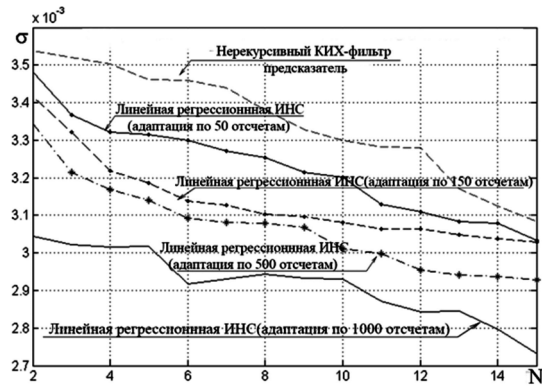


Рис. 10. Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для линейной регрессионной ИНС

Из анализа зависимостей следует, что ошибка предсказания линейной регрессионной сети существенно зависит от количества отсчетов, по которым происходит адаптация. Наилучшие результаты достигнуты при адаптации по 1000 отсчетам, что при частоте дискретизации 8 кГц составляет 125 мс. Таким образом, можно сделать вывод, что линейная регрессионная сеть может быть применена в качестве адаптивного предсказателя. Помимо этого, ИНС данной архитектуры при условии адаптации на приёмной стороне позволяет снизить ошибку предсказания от 15 % до 70 % при том же порядке или уменьшить порядок предсказания на 10 %–60 % при той же ошибке.

Анализ показателей качества речи при нейросетевых реализаций предсказателей в кодерах РС

Далеко не всегда качество восстановленного РС может быть охарактеризовано значениями СКО отсчетов, в связи с этим необходимо проводить анализ субъективной оценки качества речи.

Для получения субъективной оценки качества РС по методу MOS [18] были произведены записи речи шести дикторов, которые начитывали акустически взвешенные фразы, прописанные в ГОСТ Р 51061-97. Полученные РС кодировались стандартным кодером источника сообщения с предсказателем на основе КИХ-фильтра, а также с помощью нейросетевой реализации фильтра-предсказателя. Десять аудиторов производили субъективную оценку прослушиваемых РС по 5 бальной шкале [19].

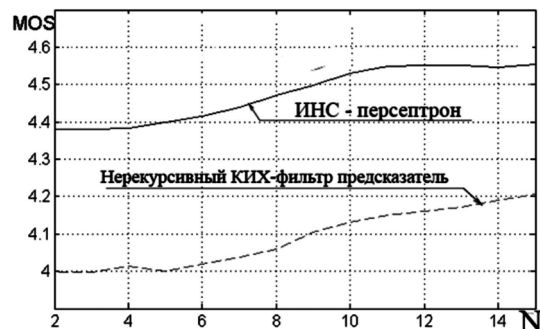
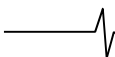


Рис. 11. Зависимости оценки качества речи по шкале MOS от порядка предсказателя

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



MOS. На рис. 11 и рис. 12 изображены зависимости субъективной оценки качества речи по шкале MOS от порядка предсказателя.

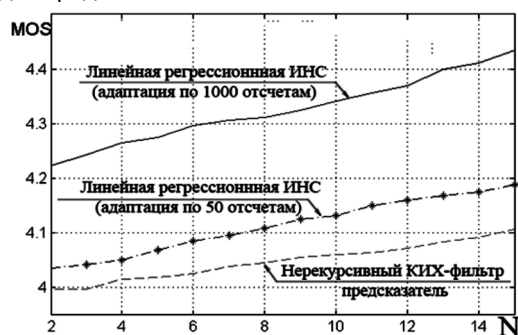


Рис. 12. Зависимости оценки качества речи по шкале MOS от порядка предсказателя

Из анализа зависимостей видно, что предсказатель на базе ИНС позволяет увеличить оценку качества речи по шкале MOS до 0,5 балла.

Заключение

Проведенные исследования показали эффективность применения нейросетевой реализации алгоритма подавления АП. Удалось достигнуть снижения уровня дисперсии помех на выходе кодера в десятки раз, по сравнению со стандартным кодером; при этом увеличилось отношение сигнал-помеха на 7 - 23 дБ. Однако субъективная оценка качества речи по шкале MOS не снижалась. Отмечено преимущество нейросетевых предсказателей перед известным предсказателем на основе КИХ-фильтра. При реализации неадаптивного предсказателя на базе перцептрона возможно уменьшить порядок предсказателя на 20% - 50% при той же ошибке предсказания, а применение в качестве адаптивного предсказателя ИНС линейно-регрессионной архитектуры позволяет снизить ошибку предсказания от 15 % до 70 % при том же порядке или уменьшить порядок предсказания от 10 % до 60 % при той же ошибке предсказания. Также отмечено увеличение субъективной оценки качества речи по шкале MOS до 0,5 балла, что является существенным преимуществом данных систем.

Литература

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. Benesty J., Chen J., Huang Y., Cohen I. Noise Reduction in Speech Processing // Springer-Verlag, 2009.
4. Злобин В.К., Григоренко Д.В., Ручкин В.Н., Романчук В.А. Кластеризация и восстанавливаемость нейро-процессорных 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. Проек-

тирование 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. Кириллов С.Н., Сазонова Е.С. Нейросетевая реализация предсказателя речевых сигналов в кодерах речи // Вестник РГРТУ № 56 – 2016 – 31-37 с.

7. Кириллов С.Н., Попова Е.С. Нейросетевая реализация кодера речевых сигналов адаптивного к уровню акустических шумов. Вестник РГРТУ, 2015 № 54, – 40-44 с.

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 г.

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, pp. 333 – 338 И.П.

10. Карасев О.Е. Применение теории нечётких множеств для обработки видеoinформации 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 series prediction H Trans. Neural Networks, 1993. – Vol. 4.

14. Gill P. Murray W., Wright M. Practical Optimization. – N.Y.: Academic Press, 1987.

15. Sohn J., Kim N.S., and Sung W. 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

20. Строгонов А. Использование нейронных сетей для прогнозирования деградации выходных параметров ТТЛ ИС в системе MATLAB/Simulink // Компоненты и технологии. №1. 2006.

21. Дьяконов В.П., Абраменкова И.В., Круглов В.В. MATLAB 5.3.1 с пакетами расширений. М.: Нолидж. 2001.