

УДК 621.396.669

## НЕЙРОСЕТЕВАЯ РЕАЛИЗАЦИЯ ПРЕДСКАЗАТЕЛЯ РЕЧЕВЫХ СИГНАЛОВ В КОДЕРАХ РЕЧИ

С. Н. Кириллов, зав. каф. РУС РГРТУ, д.т.н., профессор, kirillov.lab@gmail.com

Е. С. Сазонова, магистрант РГРТУ, katrinik94@mail.ru

Предложена реализация алгоритма предсказания речевых сигналов в кодерах речи на основе искусственных нейронных сетей (ИНС). Целью работы является анализ возможности реализации предсказателя речевых сигналов (РС) в кодерах речи на базе ИНС. Показаны преимущества реализации предсказателей на базе ИНС по сравнению с известными предсказателями на основе нерекурсивных КИХ-фильтров. Доказана возможность уменьшения порядка предсказания от 10% до 60% при той же ошибке, и снижение ошибки предсказания от 15% до 70% при том же порядке при реализации предсказателей на основе ИНС. Достигнуто увеличение субъективной оценки качества РС по шкале MOS на 0,1-0,45 балла, что является существенным преимуществом данных систем.

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

DOI: 10.21667/1995-4565-2016-56-2-31-37

### Введение

Предсказатели являются одной из основных частей современных кодеров источника информации в виде речевых сигналов (РС), которые широко применяются в таких системах первичного кодирования, как дифференциальная импульсно-кодовая модуляция, адаптивная дифференциальная импульсно-кодовая модуляция, различных вокодерах и липредерах [1 – 3]. В данных кодерах речи используются алгоритмы линейного предсказания, с помощью которых при анализе в передающем устройстве определяются коэффициенты предсказания, а в приемном устройстве на основе этих коэффициентов с помощью рекурсивного цифрового фильтра синтезируются исходные значения отсчетов РС или создается эквивалент голосового тракта в случае вокодеров. В связи с этим подобные системы получили наибольшее распространение в мире и являются одними из самых перспективно развивающихся [4].

Для реализации линейного кодирования чаще всего используются нерекурсивные КИХ-фильтры (рисунок 1).

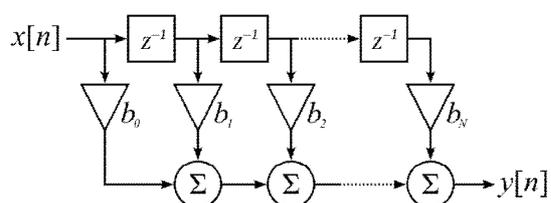


Рисунок 1 – Нерекурсивный КИХ-фильтр

Математически выходные значения предсказателя, реализованного по данной схеме, могут быть описаны выражением:

$$\tilde{y}(n) = \sum_1^N b_k * x(n - k), \quad (1)$$

а ошибка предсказания рассчитывается по формуле

$$\varepsilon(n) = x(n) - \tilde{y}(n) = x(n) - \sum_{k=1}^N b_k * x(n - k), \quad (2)$$

где  $x(n - k)$  – отсчеты РС,  $b_k$  – коэффициенты КИХ-фильтра,  $\tilde{y}$  – предсказанное значение. В выражении (1) точность предсказания определяется точностью оценки коэффициентов КИХ-фильтра  $b_k$ .

Для увеличения точности предсказания могут быть использованы искусственные нейронные сети (ИНС). ИНС представляют собой весьма перспективную вычислительную технологию, дающую новые подходы к исследованию различных динамических задач [5, 6]. Свойство толерантности, присущее для ИНС, позволяет получать решения, робастные к различным видам искажений. Способность к моделированию нелинейных процессов, работе с зашумленными данными и адаптивность дают возможности по применению ИНС при решении широкого класса задач [7]. Помимо этого, ИНС применяются в ситуациях, когда невозможно точно установить зависимость между входными и выходными данными, как, например, при задаче предсказания отсчетов РС.

Целью работы является анализ возможности реализации предсказателя РС в кодерах речи на базе ИНС.

**Архитектуры ИНС, используемые в качестве предсказателей РС**

Известно, что для задач регрессии рекомендуется использовать такие типы архитектур ИНС, как многослойный перцептрон и линейная регрессионная сеть [7, 8].

ИНС состоит из нейронов, в состав которых входят умножители, сумматор и нелинейный преобразователь, рисунок 2.

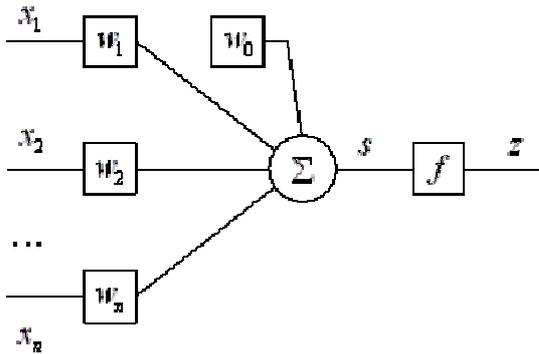


Рисунок 2 – Модель нейрона

Здесь  $s$  — результат суммирования;  $x_1, x_2, \dots, x_n$  — компоненты входного вектора (матрица входного сигнала  $X$ );  $n$  — число входов нейрона;  $w_1, w_2, \dots, w_n$  — веса (матрица-вектор  $W$ );  $w_0$  — смещение (скаляр);  $z$  — выходной сигнал нейрона;  $f$  — нелинейное преобразование (функция активации). Обычно функция активации имеет вид:  

$$z = f(s) = f(W * X + w_0). \quad (3)$$

Математически модель нейрона записывается в следующем виде:

$$s = \sum_1^n x_i * w_i + w_0 = W * X + w_0. \quad (4)$$

Данная модель имеет сходство с адаптивным линейным сумматором, что позволяет сделать вывод о возможности применения ИНС для реализации предсказателя РС.

Рассмотрим архитектуры ИНС типа многослойный перцептрон, рисунок 3, и линейная рег-

рессионная сеть, рисунок 4, рекомендованные в [8] в качестве предсказателей.

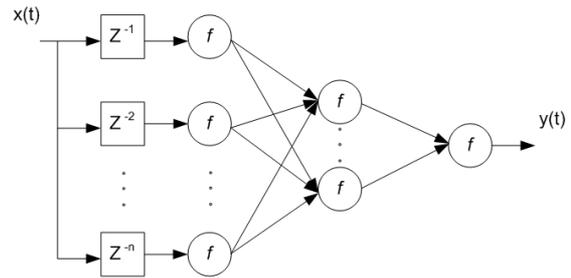


Рисунок 3 – Многослойный перцептрон

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

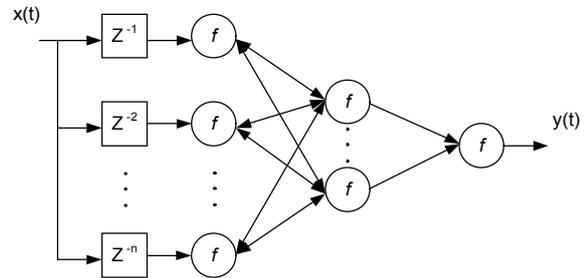


Рисунок 4 – Линейная регрессионная сеть

Математически значение выходов нейронов полностью соответствует выражению (4). 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 рисунке 5.

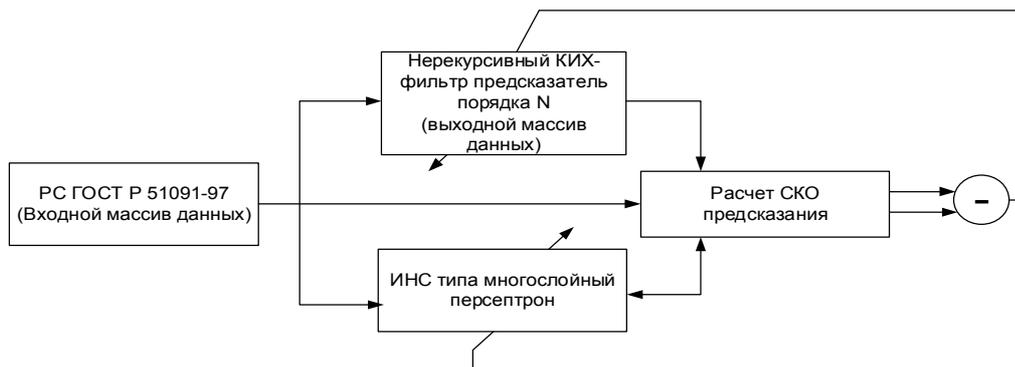


Рисунок 5 – Схема обучения ИНС типа многослойный перцептрон

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

В качестве выходного массива данных были выбраны выходные значения фильтра предсказателя соответствующего порядка. Обучение мно-

гослояного персептрона проводилось по алгоритму Левенберга-Марквардта и обратного распространения, а также квази-Ньютоновским методом обратного распространения. Зависимости среднеквадратичной ошибки предсказания  $\sigma$  от порядка фильтра  $N$  изображены на рисунке 6.



Рисунок 6 — Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для ИНС типа многослойный персептрон

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

На рисунке 7 изображена схема эксперимента.

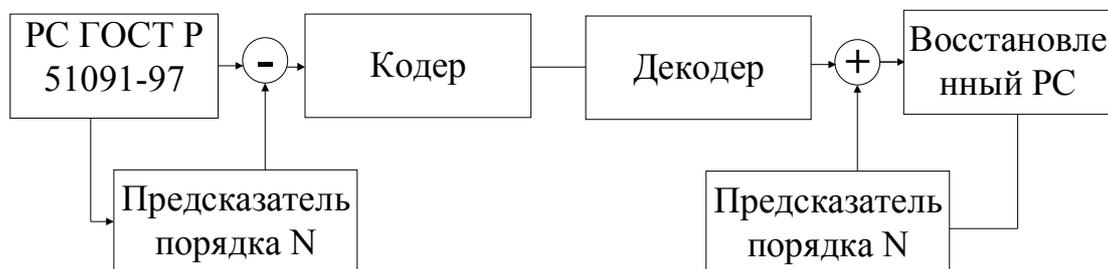


Рисунок 7 — Схема эксперимента

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

- с обучением ИНС как на передающей, так и на принимающей стороне, 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 рисунке 8 изображены зависимости среднеквадратичной ошибки предсказания  $\sigma$  от

порядка фильтра  $N$  при наличии и отсутствии обучения ИНС-предсказателя на приемной стороне.

Из анализа рисунка 8 видно, что существенных различий 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 персептрон, перед известным нерекursивным КИХ-фильтром предсказателем, так возможно уменьшить порядок предсказателя от 20% до 50% при той же ошибке предсказания.

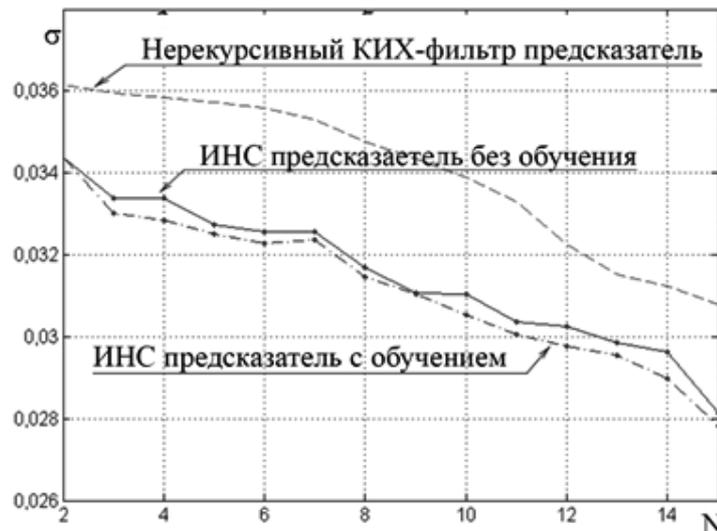


Рисунок 8 — Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для ИНС типа многослойный персептрон

#### Анализ ИНС типа линейная регрессионная сеть

В отличие от многослойного персептрона 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 рисунком 7, при этом адаптация нерекурсивного КИХ-фильтра осуществлялось по алгоритму Левенсона-Дарбина. В данном случае для реализации схемы эксперимента с помощью предсказателя на основе нерекурсивного 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 рисунке 9.

Проанализировав рисунок 9, можно сделать вывод, что адаптивный ИНС-предсказатель обеспечивает более высокие результаты по сравнению с известным предсказателем на основе нерекурсивного КИХ-фильтра при этом не требуя организации дополнительного низкоскоростного канала для передачи коэффициентов предсказания, что является его существенным преимуществом.

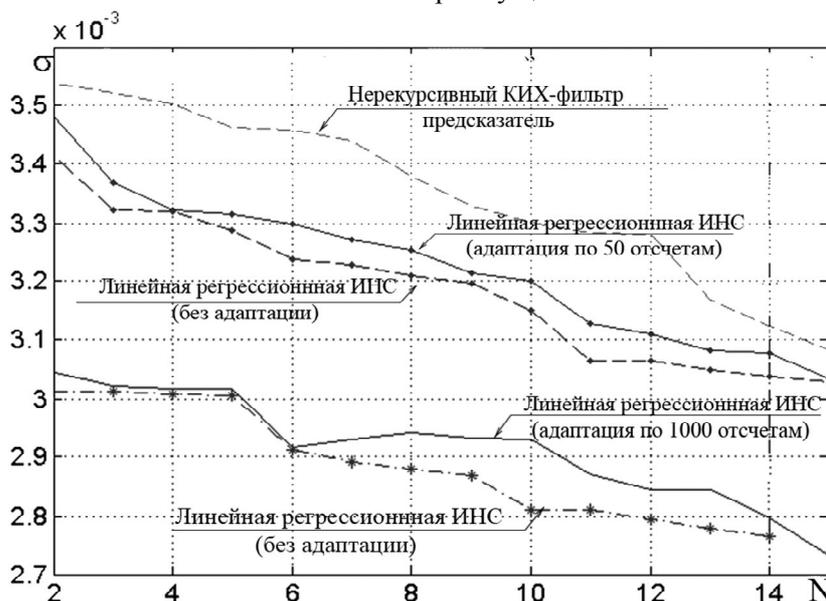


Рисунок 9 — Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для линейной регрессионной ИНС

На рисунке 10 приведены зависимости среднеквадратичной ошибки предсказания от

порядка предсказателя для 50, 150, 500 и 1000 отсчетов адаптации РС.

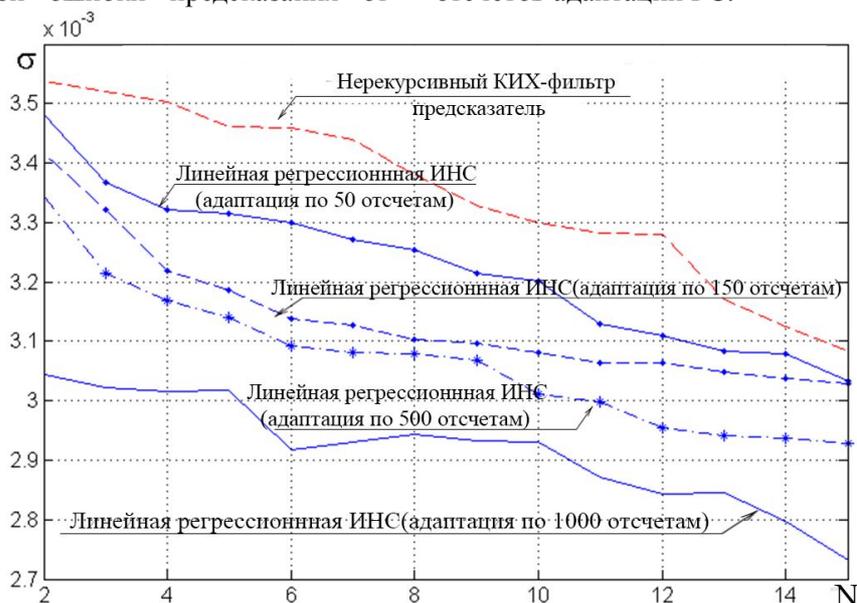


Рисунок 10 — Зависимости среднеквадратичной ошибки предсказания от порядка предсказателя для линейной регрессионной ИНС

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 нейросетевых реализаций предсказателей в кодерах речи

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

Для получения субъективной оценки качества РС по методу MOS [10] 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 и с помощью нейросетевой реализации фильтра-предсказателя. Десять аудиторов производили субъективную оценку прослушиваемых РС по 5-бальной шкале [11].

Оценки аудиторов складывались, а затем

находилось среднее значение субъективной оценки по методу MOS.

На рисунке 11 и рисунке 12 изображены зависимости субъективной оценки по шкале MOS от порядка предсказателя.

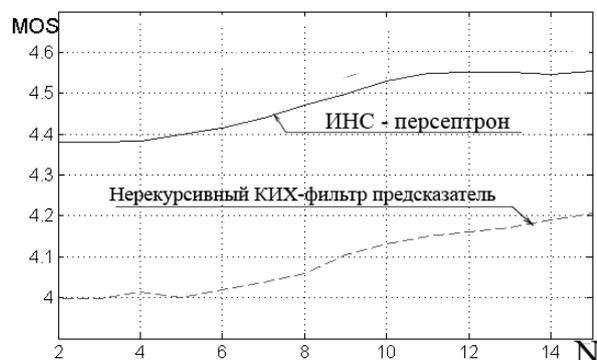


Рисунок 11 — Зависимости субъективной оценки по шкале MOS от порядка предсказателя

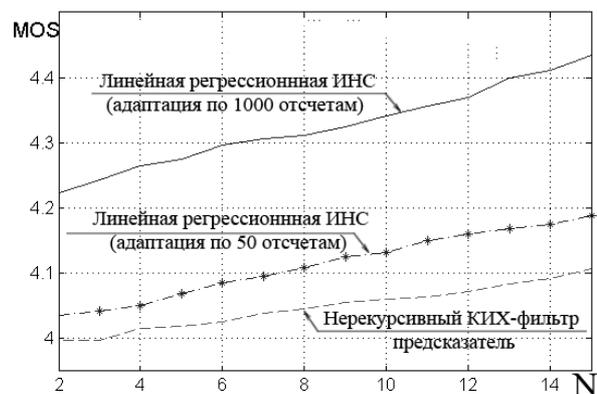


Рисунок 12 — Зависимости субъективной оценки по шкале MOS от порядка предсказателя

Из анализа зависимостей видно, что пред-

сказатель на базе ИНС позволяет увеличить субъективную оценку качества по шкале MOS на 0,1-0,45 балла.

### Заключение

Проведенные исследования показали преимущество нейросетевых предсказателей перед известным нерекурсивным КИХ-фильтром предсказателем. Так, 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. Также отмечено увеличение субъективной оценки качества РС по шкале MOS на 0,1 – 0,45 балла, что является существенным преимуществом данных систем.

### Библиографический список

1. Кириллов С. Н., Стукалов Д. Н. Цифровые системы обработки речевых сигналов: учеб. пособие // Рязан. гос. радиотехн. акад. Рязань, 1995.
2. ITU-T Recommendation G.726. 40, 32, 24, 16 kbit/s ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION (ADPCM)
3. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов: пер. с англ.; под ред. Назарова М.В., Прохорова Ю.Н. М.: Издательство «Радио и связь», 1981.

4. Кагановский Ю. Д. Применение модели линейного предсказания для анализа стохастических сигналов // Технические науки: традиции и инновации: материалы междунар. науч. конф. (г. Челябинск, январь 2012 г.) — Челябинск: Два комсомольца, 2012. — С. 12-14.

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. Кириллов С. Н., Попова Е. С. Нейросетевая реализация кодера речевых сигналов адаптивного к уровню акустических шумов // Вестник Рязанского государственного радиотехнического университета. 2015. № 54-1. С. 40-44.

7. Осовский С. Нейронные сети для обработки информации / пер. с польского Рудинского И.Д.. – М.: Финансы и статистика, 2002.

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

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

10. Рекомендация МСЭ-Т Р.80/Р.800

11. ГОСТ Р 51061-97 Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений. – Введ. 01.01.98. – М.: Госстандарт России.

UDC 621.396.669

## NEURAL NETWORKS PREDICTOR SPEECH SIGNALS

**S. N. Kirillov**, PhD (technical sciences), full professor, Head of the Department, RSREU, Ryazan; kirillov.lab@gmail.com

**E. S. Sazonova**, master, katrinik94@mail.ru

*The theoretical and practical aspects of design and optimization structures of neural network implementations predictors of speech signals are offered. The aim of the work is to justify bath implementation advantages predictors based on artificial neural networks compared with known predictors based on non-recursive FIR filters. The possibility to reduce the prediction order from 10% to 60% at the same error, and the prediction error reduction from 15% to 70% at the same manner is proved. The increase of subjective evaluation of speech signal quality on 0,1-0,45 points according to MOS scale is achieved which can be viewed as a considerable advantage of given systems.*

**Keywords:** artificial neural network, perceptron, linear regression network, FIR filter, predictor, speech signals.

**DOI:** 10.21667/1995-4565-2016-56-2-31-37

## References

1. **Kirillov S. N., Stukalov D. N.** Cifrovye sistemy obrabotki rechevyh signalov (Digital processing of speech signals): uchebnoe posobie. Ryazan, gos. radio-tehn. Akad, 1995 (in Russian).
2. ITU-T Recommendation G.726. 40, 32, 24, 16 kbit/s ADAPTIVE DIFFERENTIAL PULSE CODE MODULATION (ADPCM).
3. **Rabiner L. R., Shafer R. V.** Cifrovaja obrabotka rechevyh signalov (Digital processing of speech signals). Per. s angl., pod redakciej Nazarova M. V., Prohorova Ju. N. Moscow: Radio i svjaz', 1981 (in Russian).
4. **Kaganovskij Ju. D.** Primenenie modeli linejnogo predskazanija dlja analiza stohasticheskikh signalov. Tehnicheskie nauki: tradicii i innovacii: materialy mezhdunar. nauch. konf (g. Cheljabinsk, janvar' 2012 g.). Cheljabinsk: Dva komsomol'ca, 2012. pp. 12-14 (in Russian).
5. **Danilin S. N., Makarov M. V., Shhanikov S. A.** Proektirovanie tehniceskikh sredstv s nejrosetevoj arhitekturoj pri iskazhenii shumami vhodnoj informacii. 24-ja Mezhdunarodnaja Krymskaja konferencija "SVCh-tehnika i telekommunikacionnye tehnologii": materialy konf. : v 2 t. Sevastopol', 2014 (in Russian).
6. **Kirillov S. N., Popova E. S.** Nejrosetevaja realizacija koderu rechevyh signalov adaptivnogo k urovnju akusticheskikh шумов. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2015, no. 54-1, pp. 40 – 44 (in Russian).
7. **Osovskij S.** Neironnye seti dlja obrabotki informacii (Neural network for information processing). Per. s pol'skogo Rudinskogo I. D. Moscow: Finansy i statistika, 2002 (in Russian).
8. **Strogonov A.** Ispol'zovanie neironnyh setej dlja prognozirovaniya degradacii vyhodnyh parametrov TTL IS v sisteme MATLAB. Simulink. *Komponenty i tehnologii*, 2006, no. 1 (in Russian).
9. **D'jakonov V. P., Abramenkova I. V., Kruglov V. V.** MATLAB 5.3.1 s paketami rasshirenij (MATLAB 5.3.1 with extension packs). Moscow: Nolidzh, 2001 (in Russian).
10. Rekomendacija MSJe-T R.80/R.800 (in Russian).
11. GOST R 51061-97 Sistemy nizkoskorostnoj peredachi rechi po cifrovym kanalim. Parametry kachestva rechi i metody izmerenij. V ved. 01.01.98. Moscow: Gosstandart Rossii (in Russian).