

УДК 004.934

ОЦЕНКА МЯГКОЙ МАСКИ С ИСПОЛЬЗОВАНИЕМ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ПОДАВЛЕНИЯ ШУМА В РЕЧЕВЫХ СИГНАЛАХ

*Тупицин Г.С., к.т.н., Ярославский государственный университет им. П.Г. Демидова
e-mail: genichyar@genichyar.com;*

*Топников А.И., к.т.н., доцент, Ярославский государственный университет им. П.Г. Демидова
e-mail: topartgroup@gmail.com.*

SOFT MASK ESTIMATION TECHNIQUE USING RECURRENT NEURAL NETWORK FOR SPEECH ENHANCEMENT

Tupitsin G.S., Topnikov A.I.

The work is devoted to the speech quality improvement through speech signals processing by a noise reduction algorithm based on the soft mask. A recurrent neural network for the soft mask estimation was proposed. It was pretrained on speech database records, for a total duration of more than an hour. The algorithm was implemented in Python language and it allows real-time speech processing in some cases. A comparison of proposed solution with two widely used Wiener gain function-based algorithms and convolutional neural network speech denoising algorithm was performed using PESQ speech quality measure.

Key words: speech signal, denoising, soft mask, recurrent neural network.

Ключевые слова: речевой сигнал, шумоподавление, мягкая маска, рекуррентная нейронная сеть.

Введение

Алгоритмы восстановления речевого сигнала, искаженного аддитивным некоррелированным шумом, в случае, когда доступен только зашумленный сигнал, имеют широкое применение в различных приложениях цифровой обработки речевых сигналов, таких как распознавание речи, распознавание говорящего, детектирование речевой активности, улучшение качества и разборчивости речевых сигналов и др. [1]. Предложены методы подавления шума в частотной области, использующие различные функции коррекции спектра (ФКС), зависящие от апостериорного отношения сигнал/шум (ОСШ) и/или оценки априорного ОСШ. Для ФКС спектрального вычитания [2] оценка априорного ОСШ не требуется. В ином случае она может осуществляться с помощью подхода прямого принятия решения (decision-directed) [3, 4], его модификации на основе двухступенчатого алгоритма (two step noise reduction, TSNR) [5, 6], а также других методов [7, 8]. Для коррекции спектра на практике используются различные ФКС: Винера [9], минимальной среднеквадратичной ошибки кратковременного амплитудного спектра (minimum mean square error short-time spectral amplitude, MMSE-STSA) [4] и др. Помимо методов подавления шума в частотной области существуют и другие подходы, например, работающие в модуляционной области [10].

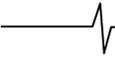
Еще одна группа методов шумоподавления в частотной области основана на применении частотно-временных бинарных [11-13] и мягких масок [14-16] в качестве ФКС. При построении бинарных масок используется предположение, что аддитивный шум искажает от-

Работа посвящена улучшению качества зашумленных речевых сигналов путем обработки с использованием алгоритма подавления шума на основе мягкой маски. Для оценки маски предложено использовать рекуррентную нейронную сеть, предварительно обученную на записях из базы речевых сигналов общей продолжительностью более часа. Разработанный алгоритм реализован на языке Python и в отдельных случаях позволяет производить обработку речевых сигналов в реальном времени. С использованием показателя качества PESQ проведено его сравнение с двумя широко используемыми алгоритмами шумоподавления, использующими функцию коррекции спектра Винера, а также с алгоритмом на основе свёрточной сети, разработанным авторами ранее.

дельные точки частотно-временного представления сигнала сильно, а другие – незначительно. Соответственно, они могут быть разделены на «зашумленные» и «не очень зашумленные». Мягкие маски конструируются по аналогичному принципу, однако для каждой точки частотно-временного представления сигнала вычисляется вероятность того, что в ней преобладает речь, а не шум.

С развитием эффективных методов машинного обучения широкое распространение стали получать алгоритмы подавления шума на основе искусственных нейронных сетей [17-19]. Данные алгоритмы предварительно обучаются на большом наборе речевых сигналов и могут использовать дополнительную априорную информацию, такую как, например, видеозапись говорящего [20].

В работе [21] предложено использовать свёрточную нейронную сеть для оценки мягкой маски, при этом целевой переменной при обучении нейронной сети служит бинарная маска, определенная по порогу 0 дБ. В настоящем исследовании предлагается использовать рекуррентную нейронную сеть вместо свёрточной, а в качестве целевой переменной использовать мягкую маску, рассчитанную, исходя из предположения о рэлеевском распределении амплитудного спектра шума в каждой



частотной полосе. Как и оригинальный алгоритм, усовершенствованная версия учитывает априорную информацию о спектральной плотности мощности шума в каждой частотной полосе.

Целью работы является повышение качества речевых сигналов в условиях воздействия аддитивных шумов различных типов с помощью нейросетевого алгоритма шумоподавления.

Мягкие маски

Рассмотрим речевой сигнал $x(t)$, искаженный аддитивным шумом $n(t)$. Наблюдаемый зашумленный речевой сигнал $y(t)$ можно выразить следующим образом:

$$y(t) = x(t) + n(t)$$

Разбив зашумленный сигнал на перекрывающиеся окна и выполнив преобразование Фурье для каждого окна, получим:

$$Y_{k,w} = X_{k,w} + N_{k,w},$$

где k – номер частотной полосы, w – номер окна; $Y_{k,w}$, $X_{k,w}$, $N_{k,w}$ – Фурье-образы зашумленного сигнала, незашумленного сигнала и шума соответственно.

Задача подавления шума в частотной области сводится к поиску фильтра $G_{k,w}$ для зашумленного сигнала (иначе – ФКС), который позволял бы получить оптимальную (по некоторому критерию) оценку незашумленного сигнала [22]. Применить его (путем поэлементного умножения) необходимо к амплитудному спектру зашумленного сигнала $R_{k,w}$ для получения оценки незашумленного $\hat{A}_{k,w}$:

$$\hat{A}_{k,w} = G_{k,w} R_{k,w},$$

При этом оптимальной оценкой фазового спектра незашумленного сигнала по критерию минимума среднеквадратичной ошибки является непосредственно фазовый спектр зашумленного сигнала [4].

Бинарные и мягкие маски также могут использоваться в качестве функции коррекции спектра. ФКС бинарной маски определяется следующим образом:

$$B_{k,w} = \begin{cases} 1, & H_0 \\ 0, & H_1 \end{cases}$$

где H_0 – гипотеза, что в некоторой частотно-временной точке присутствует преимущественно речь; H_1 – гипотеза, что в некоторой частотно-временной точке присутствует преимущественно шум. Для выделения частотно-временных точек, содержащих речь, предложено использовать следующее неравенство [11, 23]:

$$\zeta_{k,w}^{local} > \tau,$$

$\zeta_{k,w}^{local}$ – локальное априорное отношение сигнал-шум, равное отношению спектральных плотностей мощности незашумленного сигнала и шума для окна w и номера частотной полосы k ; τ – порог, обычно принимается равным 1 (что соответствует 0 дБ) [24, 25].

Так как локальное априорное отношение сигнал-шум неизвестно, бинарная маска может быть оценена только приблизительно. При этом значение бинарной маски

для каждого окна и частотной полосы можно рассматривать как случайную величину, имеющую распределение Бернулли с параметром p , равным вероятности справедливости гипотезы H_0 . Используя подход, описанный в [14, 26] можно записать:

$$\begin{aligned} \hat{A}_{k,w} &= E\{B_{k,w}\}R_{k,w} = \\ &= [E\{B_{k,w}H_0\}P(H_0) + E\{B_{k,w}H_1\}P(H_1)]R_{k,w}. \end{aligned}$$

Учитывая, что $E\{B_{k,w}H_0\} \equiv 1$, а $E\{B_{k,w}H_1\} \equiv 0$, выражение (1) можно привести к виду:

$$\hat{A}_{k,w} = S_{k,w} R_{k,w},$$

$$S_{k,w} = P(H_0) = P(\zeta_{k,w}^{local} > \tau).$$

Для $S_{k,w}$ предложен термин «мягкая маска» [14, 27].

Ее значения в каждой точке частотно-временного представления сигнала соответствуют вероятности того, что этой точке преобладает речь, а не шум.

Существуют различные алгоритмы для оценки мягкой маски. Так в [16, 27] предложена методика оценки мягкой маски на основе фильтра Винера и предположения о рэлеевском распределении амплитудного спектра шума. В [21] для оценки мягкой маски используется свёрточная нейронная сеть.

Нейросетевой алгоритм оценки мягкой маски

Предлагаемая рекуррентная нейронная сеть имеет 2 входа. На вход № 1 подается текущее обрабатываемое окно с амплитудным спектром сигнала, а также 31 предыдущее. На вход № 2 – апостериорное отношение сигнал/шум текущего окна, которое рассчитывается по следующей формуле:

$$\gamma_{k,w} = R_{k,w}^2 / D_k^2,$$

где D_k^2 – спектральная плотность мощности шума.

После входа № 1 следует следующая цепочка из двух последовательно идущих управляемых рекуррентных блоков (Gated Recurrent Unit, GRU) [28] со 129 нейронами и функцией активации на основе гиперболического тангенса.

После входа № 2 следует плотный слой со 129 нейронами и функцией активации на основе гиперболического тангенса. Выходы описанных цепочек скрытых слоев объединяются в одну путем конкатенации. Далее следует цепочка из 5 последовательно идущих плотных слоев с 258 нейронами и функцией активации ELU (Exponential Linear Unit) [29]. На выходе нейронной сети расположен плотный слой со 129 нейронами и логистической функцией активации. Логистическая функция активации представлена формулой:

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

где x – выход с нейрона предыдущего слоя. Ее выбор связан непосредственно с тем, что возможные значения мягкой маски находятся на интервале [0, 1].

В качестве целевой переменной используется мягкая маска, рассчитанная исходя из предположения о рэлеевском распределении амплитудного спектра шума:

$$S_{k,w} = P(D_{k,w} \leq A_{k,w}) =$$

$$= F^{P_{злей}}(A_{k,w} D_k) = 1 - \exp[-A_{k,w}^2 / (2D_k^2)].$$

На этапе обучения нейронной сети используется логистическая функция потерь, обеспечивая минимизацию перекрестной энтропии.

Для обучения нейронной сети сигналы искажались аддитивным белым гауссовским шумом (АБГШ), а также шумами «Speech babble» и «Vehicle interior noise» из библиотеки NOISEX-92 [30] для достижения ОСШ от 0 до 20 дБ. Тип шума и ОСШ для каждого сигнала выбирался случайным образом. После искажения сигналы подавались на входы нейронной сети для ее обучения. Затем данная процедура повторялась, причем тип шума и ОСШ вновь выбирались случайно для каждого сигнала, что обеспечило очень большой набор вариаций сигнал – тип шума – ОСШ. Произведено 1000 эпох обучения (1 эпоха – полный набор сигналов обучающей выборки).

Реализация и тестовое окружение

Предлагаемый нейросетевой алгоритм подавления шума реализован с помощью языка Python и библиотек SciPy [31] и Keras [32].

В работе речевые сигналы предварительно разбивались на перекрывающиеся окна длиной 32 мс, что при частоте дискретизации 8 кГц соответствует 256 отсчетам. Величина перекрытия окон составляла 50 % от длины окна. При получении спектра сигнала использовалось окно Хеннинга.

Для исследования использована база речевых сигналов, составленная на основе записей речевого корпуса CHAINS школы компьютерных наук и информатики университетского колледжа Дублина [33]. Данный корпус содержит записи 36 дикторов в различных стилях произношения, из которых в работе используется лишь стиль одиночного чтения (маркируется в корпусе как «Solo reading»). Записи выбранного стиля записаны в профессиональной студии звукозаписи на микрофон Neumann U87. Запись произведена в одну сессию. Число дикторов в корпусе – 36 (20 мужчин и 16 женщин).

Для обучения нейронной сети использованы записи 30 дикторов (по 4 записи на диктора), общей продолжительностью 1 час 17 минут 29 секунд. Тестирование производилось на записях оставшихся 6 дикторов (по 4 записи на диктора), общей продолжительностью 9 минут 12 секунд.

Исследование качества работы алгоритма

Проведено сравнение предложенного алгоритма шумоподавления с алгоритмом на основе свёрточной нейронной сети, а также двумя широко используемыми

алгоритмами в задаче повышения качества зашумленных речевых сигналов. Для удобства обозначим участвующие в сравнении алгоритмы следующим образом:

Алгоритм № 1 – предложенный алгоритм, основанный на оценке мягкой маски с использованием рекуррентной нейронной сети.

Алгоритм № 2 – алгоритм, основанный на оценке мягкой маски с использованием свёрточной нейронной сети.

Алгоритм № 3 – двухступенчатый алгоритм шумоподавления ($\alpha = 0,98$) на основе ФКС Винера.

Алгоритм № 4 – алгоритм на основе подхода прямого принятия решения ($\alpha = 0,98$) и ФКС Винера.

В ходе исследования тестовые сигналы искажались АБГШ, а также шумами «Speech babble» и «Vehicle interior noise» из библиотеки NOISEX-92 для достижения ОСШ 0, 5, 10, 15, 20 дБ. Затем производилась обработка сигналов с помощью указанных алгоритмов шумоподавления.

Эффективность работы алгоритмов оценивалась с помощью показателя качества PESQ (Perceptual Evaluation of Speech Quality) [34]. Оценка качества производилась для каждого восстановленного сигнала, затем результаты усреднялись. Результаты проведенного сравнения представлены в табл. 1.

Видно, что предложенный алгоритм обеспечивает лучшее качество обработанного сигнала по показателю качества PESQ для всех используемых типов шума по сравнению с тремя другими алгоритмами.

Так по сравнению с алгоритмом на основе свёрточной нейронной сети (алгоритм № 2) преимущество для АБГШ в зависимости от ОСШ составляет 0,06-0,11 (в среднем 0,08) единиц шкалы MOS (Mean Opinion Score); для шума SB – 0,04-0,18 (в среднем 0,13); для шума VIN – 0,03-0,18 (в среднем 0,08). По сравнению с двухступенчатым алгоритмом шумоподавления на основе ФКС Винера (алгоритм № 3) преимущество составляет: для АБГШ – 0,39-0,44 (в среднем 0,42); для шума SB – 0,36-0,74 (в среднем 0,58); для шума VIN – 0,04-0,29 (в среднем 0,14). По сравнению с алгоритмом на основе подхода прямого принятия решения и ФКС Винера (алгоритм № 4) составляет: для АБГШ – 0,2-0,34 (в среднем 0,25); для шума SB – 0,25-0,47 (в среднем 0,39); для шума VIN – 0,03-0,17 (в среднем 0,09).

Отметим, что в случае с шумом VIN по сравнению с АБГШ и SB алгоритмы обеспечивают близкие значения показателя качества PESQ. Это можно объяснить особенностью шума, энергия которого сосредоточена в области низких частот. Таким образом, значительная часть диапазона частот, соответствующего речевым сигналам,

Таблица 1. Сравнение алгоритмов шумоподавления на основе показателя качества PESQ

Тип шума	АБГШ				Speech babble				Vehicle interior noise			
	1	2	3	4	1	2	3	4	1	2	3	4
ОСШ, дБ												
0	2,40	2,29	1,98	2,06	2,11	2,07	1,37	1,66	3,82	3,64	3,53	3,65
5	2,70	2,61	2,29	2,42	2,58	2,45	1,89	2,11	4,05	3,93	3,86	3,94
10	3,00	2,92	2,56	2,77	2,97	2,83	2,39	2,57	4,22	4,16	4,10	4,14
15	3,31	3,24	2,87	3,11	3,31	3,13	2,80	2,95	4,33	4,30	4,27	4,29
20	3,60	3,54	3,21	3,40	3,61	3,47	3,25	3,36	4,41	4,38	4,37	4,38



в данном случае практически не искажается, и на итоговую оценку влияет в том числе и способность алгоритма обрабатывать практически неискаженные спектральные составляющие речи.

Заключение

В работе предложена усовершенствованная методика оценки мягкой маски с использованием рекуррентной нейронной сети и алгоритм подавления шума в речевых сигналах на ее основе. По сравнению с предыдущей версией алгоритма, использующей свёрточную нейронную сеть, изменена структура сети и пересмотрен алгоритм обучения – в предлагаемом варианте в качестве целевой переменной используется мягкая маска, рассчитанная, исходя из предположения о рэле-евском распределения амплитудного спектра шума в каждой частотной полосе.

С использованием показателя качества PESQ проведено сравнение разработанного алгоритма с алгоритмом шумоподавления на основе свёрточной нейронной сети, двухступенчатым алгоритмом на основе ФКС Винера, алгоритмом на основе подхода прямого принятия решения и ФКС Винера. В ходе исследования тестовые сигналы искажались АБГШ, а также шумами «Speech babble» и «Vehicle interior noise» из библиотеки NOISEX-92 для достижения ОСШ 0, 5, 10, 15, 20 дБ. В результате предлагаемый алгоритм показал преимущество для всех использованных сочетаний типа шума и ОСШ.

Результаты исследования демонстрируют преимущество предлагаемого алгоритма в задаче повышения качества зашумленных речевых сигналов и свидетельствуют о его перспективности.

Литература

1. Benesty J. Speech Enhancement: A Signal Subspace Perspective / J. Benesty, J. Jensen, M.G. Christensen, J. Chen. – Elsevier, 2014.
2. Boll S. Suppression of acoustic noise in speech using spectral subtraction / S. Boll // IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1979. – Vol. 27. – № 2. – P. 113–120.
3. Scalart P. Speech enhancement based on a priori signal to noise estimation / P. Scalart, J.V. Filho // 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. – IEEE, 1996. – Vol. 2. – P. 629–632.
4. Ephraim Y. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator / Y. Ephraim, D. Malah // IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1984. – Vol. 32. – № 6. – P. 1109–1121.
5. Plapous C. A two-step noise reduction technique / C. Plapous, C. Marro, L. Mauuary, P. Scalart // 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. – 2004. – Vol. 1. – P. 289–92.
6. Тупицин Г.С. Модификация двухступенчатого алгоритма шумоподавления для улучшения качества идентификации диктора в условиях шумов / Г.С. Тупицин, А.И. Топников, А.Л. Приоров // Информационные системы и технологии. – 2015. – № 6. – P. 39–47.

7. Lu Y. A geometric approach to spectral subtraction / Y. Lu, P.C. Loizou // Speech Communication. – 2008. – Vol. 50. – № 6. – P. 453–466.

8. Plapous C. Improved Signal-to-Noise Ratio Estimation for Speech Enhancement / C. Plapous, C. Marro, P. Scalart // IEEE Transactions on Audio, Speech and Language Processing. – 2006. – Vol. 14. – № 6. – P. 2098–2108.

9. Lim J. Enhancement and bandwidth compression of noisy speech / J. Lim, A. Oppenheim // Proceedings of the IEEE. – 1979. – Vol. 67. – № 12. – P. 1586–1604.

10. Азаров И.С. Алгоритм очистки речевого сигнала от сложных помех путем фильтрации в модуляционной области / И.С. Азаров, М.И. Вашкевич, Д.С. Лихачев, А.А. Петровский // Цифровая обработка сигналов. – 2013. – P. 25–31.

11. Wang D. On Ideal Binary Mask As the Computational Goal of Auditory Scene Analysis / D. Wang // Speech Separation by Humans and Machines. – Boston: Kluwer Academic Publishers, 2005. – P. 181–197.

12. Тупицин Г.С. Предобработка зашумленных речевых сигналов с помощью бинарных масок в задаче идентификации диктора / Г.С. Тупицин, А.И. Топников, А.Л. Приоров // Научные технологии. – 2015. – № 11. – P. 56–61.

13. Тупицин Г.С. Модификация алгоритма оценки бинарной маски в задаче подавления шума для системы идентификации диктора / Г.С. Тупицин, С.А. Кравцов, А.И. Топников, А.Л. Приоров // Проектирование и технология электронных средств. – 2015. – № 3. – P. 32–37.

14. Lu Y. Estimators of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty / Y. Lu, P.C. Loizou // IEEE Transactions on Audio, Speech, and Language Processing. – 2011. – Vol. 19. – № 5. – P. 1123–1137.

15. Тупицин Г.С. Методика оценки мягкой маски для задачи предобработки зашумленных речевых сигналов в системах идентификации диктора / Г.С. Тупицин, А.И. Топников, А.Л. Приоров // Успехи современной радиоэлектроники. – 2016. – № 6. – P. 73–80.

16. Tupitsin G. Two-step noise reduction based on soft mask for robust speaker identification / G. Tupitsin, A. Topnikov, A. Priorov // 2016 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT). – IEEE, 2016. – P. 351–356.

17. Xu Y. A Regression Approach to Speech Enhancement Based on Deep Neural Networks / Y. Xu, J. Du, L. Dai, C. Lee // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2015. – Vol. 23. – № 1. – P. 7–19.

18. Zhao H. Convolutional-Recurrent Neural Networks for Speech Enhancement / H. Zhao, S. Zarar, I. Tashev, C. Lee // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. – 2018. – P. 2401–2405.

19. Kolbæk M. Monaural Speech Enhancement using Deep Neural Networks by Maximizing a Short-Time Objective Intelligibility Measure / M. Kolbæk, Z. Tan, J. Jensen. – 2018.

20. Hou J. Audio-Visual Speech Enhancement Based on Multimodal Deep Convolutional Neural Network [Электронный ресурс] / J. Hou, S. Wang, Y. Lai, Y. Tsao, H. Chang, H. Wang. – Режим доступа: <https://arxiv.org/abs/1703.10893>.

21. Тупицин Г.С. Методика оценки мягкой маски с использованием сверточных нейронных сетей для подавления шума в речевых сигналах / Г.С. Тупицин, А.И. Топников // Сборник докладов 7-й Всероссийской конференции «Радиоэлектронные средства получения, обработки и визуализации информации». – Москва, 2017. – P. 137-141.

22. Cohen I. Noise Reduction in Speech Processing: Springer Topics in Signal Processing. Vol. 2 / I. Cohen, Y. Huang, J. Chen, J. Benesty. – Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

23. Wang D. Time-Frequency Masking for Speech Separation and Its Potential for Hearing Aid Design / D. Wang // Trends in Amplification. – 2008. – Vol. 12. – № 4. – P. 332-353.

24. Hu Y. Techniques for estimating the ideal binary mask / Y. Hu, P. Loizou // Proc. 11th Int. Workshop Acoust. Echo Noise Control. – 2008. – P. 154-157.

25. Jensen J. Spectral Magnitude Minimum Mean-Square Error Estimation Using Binary and Continuous Gain Functions / J. Jensen, R.C. Hendriks // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20. – № 1. – P. 92-102.

26. McAulay R. Speech enhancement using a soft-decision noise suppression filter / R. McAulay, M. Malpass // IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1980. – Vol. 28. – № 2. – P. 137-145.

27. Тупицин Г.С. Предобработка речевых сигналов в системах автоматической идентификации диктора: диссертация ... кандидата технических наук: 05.12.04 / Г.С. Тупицин. – Владимир: Владимирский государствен-

ный университет им. А.Г. и Н.Г. Столетовых, 2015.

28. Cho K. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation / K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – Stroudsburg, PA, USA: Association for Computational Linguistics, 2014. – P. 1724-1734.

29. Clevert, D.-A. Fast and Accurate Deep Network Learning by Exponential Linear Units [Электронный ресурс] / D.-A. Clevert, T. Unterthiner, S. Hochreiter. – Режим доступа: <https://arxiv.org/abs/1511.07289>.

30. Varga A. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems / A. Varga, H.J.M. Steeneken // Speech Communication. – 1993. – Vol. 12. – № 3. – P. 247-251.

31. Jones E. SciPy: Open source scientific tools for Python [Электронный ресурс] / E. Jones, T. Oliphant, P. Peterson, others. – Режим доступа: <http://www.scipy.org>.

32. Chollet F. Keras [Электронный ресурс] / F. Chollet, others. – Режим доступа: <https://keras.io>.

33. Cummins F. The CHAINS Speech Corpus: Characterizing Individual Speakers / F. Cummins, M. Grimaldi, T. Leonard, J. Simko // Proc of SPECOM. – 2006. – P. 1-6.

34. International Telecommunication Union. P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs / International Telecommunication Union // ITU-T Recommendation. – 2001. – Vol. 862. – P. 862.

НОВЫЕ КНИГИ

Витязев В.В.

Многоскоростная обработка сигналов – М.: Изд-во «Горячая линия-Телеком», 2017 г. – 336 с.: ил.

Рассмотрена эволюция теории и технологий многоскоростной обработки сигналов в период с начала 70-х гг. прошлого столетия до наших дней с позиции вклада, который внесли в их развитие работы российских ученых и специалистов в области цифровых информационных технологий реального времени. Описаны методы и алгоритмы многоступенчатой и многокаскадной реализаций цифровых узкополосных фильтров и банков цифровых фильтров на основе эффектов прореживания по времени (децимация во временной области) и по частоте (децимация в частотной области).

Приведена методика оптимизации параметров многоступенчатых и многокаскадных структур цифровых полосовых фильтров. Построение оптимальных структур и расчет параметров фильтров частотной селекции иллюстрируется многочисленными примерами.

Для специалистов, научных работников, преподавателей вузов, аспирантов; будет полезна студентам информационных и инфокоммуникационных направлений подготовки.

