АНАЛИЗ ПРЕДЕЛЬНЫХ ВОЗМОЖНОСТЕЙ МЕТОДОВ ШУМОПОНИЖЕНИЯ И РЕКОНСТРУКЦИИ РЕЧЕВЫХ СИГНАЛОВ, МАСКИРУЕМЫХ РАЗЛИЧНЫМИ ТИПАМИ ПОМЕХ

Хорев А. А.[1], Дворянкин С. В.[2], Козлачков С. Б.[3], Василевская Н. В.[4]

DOI: 10.21681/2311-3456-2024-1-89-100

Цель исследования: *оценка границ применимости методов шумопонижения и реконструкции (далее – методов шумоочистки) речевых сигналов.*

Метод исследований: *артикуляционные испытания.*

Результат и практическая ценность: *авторами на основе теоретических и экспериментальных исследований проведена оценка возможностей улучшения качества речевых сигналов путем применения различных методов шумоочистки и определены границы применимости данных методов. В результате экспериментальных исследований установлено, что все современные методы шумоочистки имеют недостаточную эффективность в случае корректного применения для маскирования фонограммы речеподобной помехи. Кроме того, в ходе исследований определено граничное значение отношения «сигнал/шум», при котором методы шумоочистки становятся неэффективными.*

Вклад авторов: *Хорев А. А. провел артикуляционные испытания и произвел статистическую обработку полученных результатов. Дворянкин С. В. провел испытания средства шумопонижения «Лазурь». Козлачков С. Б. подготовил артикуляционные тексты и провел испытания средства шумопонижения «GritTec's Noise Cancellation». Василевская Н. В. подготовила аналитический обзор и провела испытания средства шумопонижения «Sound Cleaner».*

Ключевые слова: *акустическая речевая разведка, разборчивость речи, речевой сигнал, защита информации, шумоочистка, спектральное вычитание, фильтрация, линейное предсказание.*

THE ANALYSIS OF THE POTENTIAL CAPABILITIES OF METHODS OF NOISE REDUCTION AND RECONSTRUCTION OF ACOUSTIC SPEECH SIGNALS MASKED BY VARIOUS TYPES OF NOISE

Horev A. A.[5], Dvoryankin S. V.[6], Kozlachkov S. B.[7], Vasilevskaya N. V.[8]

Purpose of the study: *the analysis of the potential capabilities of improving the quality of speech signals by applying various methods of noise reduction and reconstruction of acoustic speech signals.*

Research method: *articulation tests.*

1 Хорев Анатолий Анатольевич, доктор технических наук, профессор, Национальный исследовательский университет «МИЭТ», Зеленоград, Россия. E mail: horev@miee.ru, https://orcid.org/0000-0001-9074-385X

2 Дворянкин Сергей Владимирович, доктор технических наук, профессор, Национальный исследовательский ядерный университет «МИФИ», Москва, Россия. E mail: svdvoryankin@mephi.ru, https://orcid.org/0000-0001-6908-0676

3 Козлачков Сергей Борисович, кандидат технических наук, Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия. E mail: ksb.perovo@mail.ru, https://orcid.org/0000-0002-7096-6711

4 Василевская Надежда Валерьевна, Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия. E mail: infuzoriavalenoc@yandex.ru, https://orcid.org/0000-0002-0078-8665

5 Anatoly A. Horev, Dr.Sc., Professor, National Research University of Electronic Technology, Moscow Zelenograd, Russia,. E mail: horev@miee.ru, https://orcid.org/0000-0001-9074-385X

6 Sergey V. Dvoryankin, Dr.Sc., Professor, National Research Nuclear University MEPhI, Moscow, 115409, Russia, e-mail: svdvoryankin@mephi.ru, https://orcid.org/0000-0001-6908-0676

7 Sergey B. Kozlachkov, Ph.D. in Technology, Bauman Moscow State Technical University, Moscow, Russia. E mail: ksb.perovo@mail.ru, https://orcid.org/0000-0002-7096-6711

8 Nadezhda V. Vasilevskaya, Bauman Moscow State Technical University, Moscow. Russia. E mail: infuzoriavalenoc@yandex.ru, https://orcid.org/0000-0002-0078-8665

Result and Practical value: *While making a theoretical and experimental studies the authors evaluated the possibilities of improving the quality of speech signals by applying various methods of noise reduction and reconstruction of acoustic speech signals and determined the limits of their applicability. As a result of experimental studies, it has been established that all modern noise reduction technologies are insufficiently effective when a phonogram is protected by speech-like noise. The research also determined the limiting value of the signal-to-noise ratio at which noise reduction methods become ineffective.*

Authors' contribution: *Horev A. A. conducted articulation tests and performed statistical processing of the results. Dvoryankin S. V. conducted tests of the Lazur noise reduction product. Kozlachkov S. B. prepared articulation texts and tested the noise reduction product «GritTec's Noise Cancellation». Vasilevskaya N. V. prepared an analytical review and tested the noise reduction product «Sound Cleaner».*

Keywords: *acoustic speech intelligence, intelligibility of speech, speech signal, information protection, noise reduction, spectral subtraction, filtration, linear prediction.*

Введение

В настоящее время широкое распространение получили методы шумопонижения и реконструкции акустических речевых сигналов (РС), принятых в условиях помех, которые применяются в различных системах распознавания и обработки речи бытового назначения и, очевидно, используются злоумышленником при перехвате речевых сигналов. К сожалению, границы применимости и предельные достижимые показатели улучшения качества акустических РС для подавляющего большинства методов и алгоритмов шумопонижения (далее – методов шумоочистки) их разработчиками не указываются, в связи с чем провести оценку потенциала злоумышленника не представляется возможным. При этом необходимо учитывать, что оценка предельных возможностей и ограничений методов шумоочистки является ключевым фактором задания и обоснования использования значений показателей защищенности акустической речевой информации.

Ввиду отсутствия достоверных сведений о предельных возможностях методов шумоочистки, используемых злоумышленником, представляется трудновыполнимой задача разработки методики оценки защищенности акустической речевой информации, и, следовательно, создания системы защиты помещений от утечки акустической речевой информации по техническим каналам, обеспечивающей требуемую эффективность защиты от всех возможных видов помех.

Авторами настоящей работы предпринята попытка провести анализ предельных возможностей применяемых в настоящее время методов шумоочистки, определить границы их использования и обосновать наиболее эффективный тип помехи, которая была бы достаточно устойчива по отношению к подавляющему большинству современных методов шумоочистки.

Основные применяемые в настоящее время функциональные классы и методы шумоочистки представлены на рисунке 1.

Рис. 1. Классы методов шумопонижения и реконструкции речи

Методы классов фильтрации, компенсации и коррекции на рис. 1 условно можно отнести к методам шумопонижения, а методы четвертого класса – к методам реконструкции РС по его остаточным следам (в частности трекам речевых вокализмов), определяемым на зашумленных спектрограммах.

Рассмотрим некоторые из этих методов подробнее с учетом их вклада в улучшение отношения сигнал/шум при использовании процедур шумоочистки.

Методы, основанные на оценке спектральных характеристик помех (спектральное вычитание)

Поскольку человеческий слух крайне слабо чувствителен к фазе акустического сигнала, методы спектрального вычитания направлены исключительно на восстановление амплитуды спектра исходного РС. При этом корректировка амплитуды спектра очищенного РС осуществляется с помощью вычитания средней амплитуды шума из мгновенного спектра амплитуды зашумленного сегмента (кадра) РС[9]:

$$|\hat{S}(k)| = |Y(k)| - |\acute{N}(k)|, \quad (1)$$

где – $|\acute{N}(k)|$ – средняя амплитуда спектра шума. Величина $|\acute{N}(k)|$ вычисляется на основе предположения о локальной стационарности шума.

При этом очевидно, что при превышении мгновенного спектра неочищенной речи $|Y(k)|$ средней амплитудой спектра шума $|\acute{N}(k)|$ амплитуда спектра $|\hat{S}(k)|$ может принимать отрицательные значения. Для исключения подобных ситуаций применяется метод «выпрямления» значений $|\hat{S}(k)|$ (ограничения значений $|\hat{S}(k)|$ некоторой минимальной величиной (Noise Floor)).

Очевидно, что предельно достижимая эффективность данного метода ограничивается условием $|Y(k)| > |\acute{N}(k)|$. То есть спектральные компоненты РС, амплитуда которых на кадре обработки сопоставима с амплитудой спектра шума, уже не могут быть выделены. Это условие приблизительно соответствует наблюдаемости «следов» РС на спектрограмме.

Согласно результатам исследования возможностей нескольких методов спектрального вычитания[10] их применение для фонограмм, зашумленных белым шумом и помехой типа «речевой хор», при сегментарном отношении «сигнал\шум» (далее *SNR*) первоначальной записи 0 дБ позволяет обеспечить приращение указанного отношения на 5,5 дБ и 3 дБ соответственно.

Схожие результаты были получены в работе[11], где исследователям удалось обеспечить приращение сегментарного *SNR* на 8,5 дБ и 6,5 дБ соответственно для записей с исходным сегментарным SNR минус 10 дБ.

В работе[12] приращение указанного показателя относительно первоначального значения минус 5 дБ для фонограммы, зашумленной белым шумом, составило 14 дБ.

Самой распространенной проблемой методов спектрального вычитания является появление помех вида «музыкальный шум», т.е. возникновение в спектре очищенного РС изолированных максимумов, звучащих после преобразования сигнала во временное пространство как случайные тона. При этом многими исследователями отмечается, что «музыкальный шум» зачастую снижает восприятие РС сильнее, чем исходный стационарный шум. Поэтому значительные усилия направлены на разработку способов нивелирования влияния «музыкального шума» на разборчивость речи [1–3]. Одним из них является добавление к очищенному РС с остаточными музыкальными вставками низкоуровневого фонового шума, нивелирующего музыкальный эффект без потери речевой разборчивости (РР).

Большинство современных методов анализа звуков речи основаны на спектральной модели стационарного сигнала. Недостатком такой модели является отсутствие вероятностных характеристик для основных шумовых составляющих в произносимых согласных (консонантных фонем).

В рамках реализации алгоритмов спектрального вычитания определяются акустические признаки только вокализованных фонем (аллофонов), в консонантных фонемах анализируется только их длительность.

Эти ограничения вызваны следующими свойствами преобразования Фурье при обработке нестационарных сигналов: исходный сигнал заменяется на периодический с периодом, равным длительности анализируемого участка; преобразование Фурье не обеспечивает необходимую точность при изменении параметров процесса во времени (нестационарности), поскольку дает усредненные коэффициенты для всего исследуемого сигнала. Для выполнения анализа нестационарного процесса необходимо использовать базисные функции, имеющие способность выявлять в анализируемом сигнале как частотные, так и его временные характеристики. Другими словами, сами функции должны обладать свойствами частотно-временной локализации [4, 5]. В этой связи стоит обратить внимание на методы вейвлет-обработки.

9 Boll S., Suppression of acoustic Noise in Speech Using Spectral Subtraction, IEEE Tr. On ASSP, vol.27, N4, pp.113–120, 1979.

10 Asaduzzaman M., A Spectral Domain Speech Enhancement Method Based on Noise Compensations in Both Magnitude and Phase Spectra, http://lib.buet.ac.bd:8080/xmlui/bitstream/handle/123456789/3452/Full%20Thesis.pdf?sequence=1&isAllowed=y.

11 T. Gerkmann, C. H. Richard, Unbiased MMSE-based noise power estimation with low complexity and low tracking delay, IEEE Transactions on Audio Speech and Language Processing, vol. 20, no. 4, pp. 1383–1393, 2012.

12 I. Cohen and B. Berdugo, Speech enhancement for non-stationary noise environments, ELSEVIER Signal Process., vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

Методы, основанные на оценке спектральных характеристик помех (методы вейвлет-обработки)

Большинство алгоритмов шумопонижения реализуется в пространстве частот с использованием кратковременного преобразования Фурье (*Short-time Fourier transform, STFT*), которое позволяет анализировать нестационарные сигналы. *STFT* реализует компромисс между временным и частотным разрешением. Однако *STFT* формирует для всех частот одинаковое разрешение по времени, что не вполне согласуется со сложной структурой фонем (аллофонов) речи. Некоторые алгоритмы шумопонижения разработаны с использованием вейвлет-преобразований, дающих более гибкое частотно-временное представление РС [6, 7].

Одним из популярных алгоритмов вейвлет шумопонижения является алгоритм вейвлет сжатия. Алгоритм вейвлет сжатия основан на сравнении вейвлет коэффициентов с заданным порогом. Оцениваемый порог задает границу между коэффициентами, соответствующими шуму и коэффициентами, соответствующими РС. Однако разделить коэффициенты, соответствующие шуму и сигналу с использованием порога не всегда возможно, особенно для консонантных фонем.

Для зашумленной речи энергия вокализованных звуков сопоставима с энергией шума. Использование одинакового порога для всех коэффициентов преобразования приводит не только к подавлению шума, но и самих вокализованных фонем речи[13]. Это приводит к плохому качеству РС после обработки.

Более удачной идеей является комбинирование вейвлет банка фильтров с фильтрацией (например, фильтрацией Винера) в пространстве вейвлет коэффициентов[14] или применения адаптивной пороговой фильтрации коэффициентов дискретного обучаемого вейвлет-преобразования [8].

Алгоритм фильтрации при помощи вейвлет-преобразования позволяет эффективно удалять высокочастотный шум, даже превышающий по величине исследуемый сигнал (следует отметить, что маскирующий эффект создают преимущественно низкочастотные сигналы), в то время как преобразование Фурье теряет информацию об особенностях низкочастотной части сигнала, что приводит к искажению временной формы полезного сигнала.

Вейвлет-преобразование отличается наиболее сложной и гибкой структурой представления сигналов в пространстве «масштаб-время». Это дает возможность более полного и тонкого вейвлет-анализа РС, по сравнению с другими известными видами анализа. Более того вейвлет-преобразование позволяет более достоверно отобразить кратковременные консонантные фонемы. При этом особенности сигналов «привязаны» к временной шкале.

Основной проблемой фильтрации при помощи вейвлет-преобразования является выбор вида материнского вейвлета для проведения анализа[15]. Очевидно, что вид вейвлета должен повторять форму исходного сигнала. В качестве материнских вейвлетов при получении частотно-временного представления РС (сонограмм) чаще всего выбирают следующие вейвлеты: Морле, Шеннона, «мексиканская шляпа», вейвлет Дебеши.

Вейвлет Морле относится к «грубым» вейвлетам и представляет сигнал с меньшей точностью, чем вейвлет Шеннона. Применение вейвлета Морле целесообразно при анализе сигналов с частотой дискретизации близкой к 8 кГц, либо сигналов, подвергнутых компрессии.

Применение вейвлета «мексиканская шляпа» целесообразно при анализе кратковременных участков сигнала, так как обеспечивается возможность «рассмотреть» каждый период сигнала в отдельности.

Вейвлет Дебеши применяют для отыскания межфонемных границ в случае, когда форма речевого тракта при переходе от аллофона к аллофону изменяется относительно медленно.

Ввиду того, что заранее невозможно предугадать форму РС и невозможно определить, на каком масштабе нужно искать интересующую нас информацию, выбор «материнского вейвлета» представляет нетривиальную задачу.

В зависимости от исходного *SNR* РС, применяемого (в рамках алгоритма) фильтра, а также вида материнского вейвлета алгоритмы вейвлет-обработки сигналов дают различный выигрыш в *SNR*. Приращение *SNR* для фонограмм с исходным *SNR* порядка 6–10 дБ составляет в среднем около 8 дБ[16].

В исследовании отмечается, что корректное сопоставление результатов работы всех методов вейвлет-обработки фактически невозможно, поскольку каждый из алгоритмов шумопонижения по-своему деформирует исходную запись и формирует в обработанном варианте свои артефакты, которые совершенно по-разному воспринимаются разными слушателями.

В рамках исследования рассчитывалось *SNR* для исходной записи голоса, зашумленной белым шумом, и прошедшей обработку по анализируемому алгоритму. Для анализируемых фонограмм методы,

13 H. Tasmaz, and E. Ercalebi, Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments, Digital Signal Process., vol.18. N.5. pp.797–812, 2008.

14 M. K. Hasan, S. Saluhuddin, and M. R. Khan, Reducing signal-bias from mad estimated noise level for dct speech enhancement, Signal Process., vol.84. N.1. pp.151–162, 2004.

15 Горшков Ю. Г. Обработка речевых сигналов на основе вейвлетов // T-comm: Телекоммуникации и транспорт. – 2015. – № 2. – С. 46–53.

16 Wieland B. Speech Signal noise reduction with wavelets, https://www.uni-ulm.de/fileadmin/website_uni_ulm/uzwr/wieland/wieland-diplomarbeit-speech-signal-noise-reduction-with-wavelets.pdf.

основанные на БПФ и стационарном вейвлет-преобразовании (*SWT – Stationary Wavelet Transform*), продемонстрировали меньшую эффективность, нежели методы быстрого вейвлет-преобразования (*FWT – Fast Wavelet Transform*).

Однако в исследовании[17] отмечается, что при малых исходных значениях *SNR* фонограмм (10 дБ и менее) все указанные алгоритмы значительно снижают уровень консонант (например, в одной из фонограмм после применения алгоритма оказался полностью вырезанным звук «т»).

В исследовании [9] также отмечается, что классические подходы к улучшению качества речи, основанные на задании порога в области вейвлет-преобразования, могут вносить определенные искажения в исходный РС. Особенно это касается глухих консонант. Поэтому зачастую указанные методы комбинируют с другими, такими как: спектральное вычитание, Винеровская фильтрация и т.д.

В исследовании [9] предложен метод шумопонижения, основанный на оценке минимального среднеквадратического отклонения значений сигнала, прошедшего процедуру шумопонижения, от исходного сигнала в пространстве вейвлетов, который демонстрирует заметно большую эффективность нежели классические методы спектрального вычитания *MMSE (Minimum Mean Square Error)* и *MMSE-SMPO (Minimum Mean Square Error Soft Masking Based on Posteriori SNR Uncertainty)*. Приращение *SNR* для фонограммы, зашумленной белым шумом, с исходным *SNR* минус 5 дБ составило 13,4 дБ[18].

Простые методы фильтрации

Как правило, методы фильтрации, основанные на оценке спектральных характеристик помех, эффективны только при попытке устранения стационарной помехи, спектральный состав которой заранее известен. Такие фильтры предназначены для компенсации в РС достаточно узкополосных квазистационарных помех. Практическими примерами таких помех могут служить промышленные помехи сети электропитания, трансформаторные шумы, сосредоточенные по спектру шумы механизмов и т.п.

Очевидно, что такой алгоритм наиболее работоспособен в условиях стационарной периодической помехи и является оптимальным для фильтрации гармонической помехи, снижение уровня которой в случае применения указанного метода шумопонижения возможно на величину до 25–35 дБ[19].

В то же время даже в случае заранее известного спектрального состава помехи при применении фильтрации происходит значительное искажение полезного сигнала, связанное с вычитанием спектра в полосе частот, занимаемой помехой. В случае помехи с узкополосным спектром, близкой к гармонической, сужение полосы режекции вызывает появление боковых лепестков (явление Гиббса). Правильный выбор вида оконной функции может снизить уровень боковых лепестков, но ценой ухудшения разрешающей способности по частоте и не связанным с этим искажением исходного сигнала.

Наиболее часто в процедурах шумопонижения используют окно Хемминга или усеченное окно Гаусса с минимальным уровнем боковых лепестков в частотной области.

Методы коррекции и сглаживания спектра РС

Методы используют свойство периодичности вокализованной речи, в которой звонкие звуки можно представить сигналами с периодом, кратным частоте основного тона голоса (ЧОТ). Это означает, что их энергетический спектр сосредоточен в определенных полосах частот, в то время как уровень энергии спектра, связанный с воздействием искажающих факторов, в общем случае, определен по всему диапазону частот. Существующие способы реализации этого метода можно разделить на два различных вида.

Первый – это фильтрация исходного искаженного сигнала гребенчатым фильтром. Качество очистки РС зависит от точности определения ЧОТ. Поскольку она постоянно меняется, при обработке разных участков речи требуется постоянная адаптивная подстройка фильтра, что не всегда просто реализуется на практике. Например, такая фильтрация совершенно неприемлема в случае воздействия на сигнал суммы гармонических и (или) речеподобных помех.

При некорректном задании АЧХ гребенчатого фильтра (ГФ) и (или) при любом способе округления ЧОТ максимумы АЧХ ГФ уже на второй и последующих гармониках достаточно сильно отстоят от максимумов спектрального распределения исходного сигнала. Это означает, что на этих гармониках полезный сигнал не выделяется, а подавляется. Более того, выделяются спектральные составляющие помех в окрестности гармонических составляющих.

В реальном сигнале ЧОТ изменяется во времени и, следовательно, простая модель цифрового ГФ становится неэффективной[20] .

Второй способ реализации основан на совмещении процедуры оценки и фильтрации. В этом

17 N. A. Whitmal, J. C. Rutledge, J. Cohen Wavelet-based noise reduction, Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on 5:3003-3006 vol.5, 1995 DOI:10.1109/ICASSP.1995.479477.

18 Y. Lu and P. C. Loizou, «Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty» IEEE Transactions on Audio Speech and Language Processing, vol. 19, no. 5, pp. 1123–1137, 2011.

19 Дворянкин С.В. Цифровая шумоочистка аудиоинформации. – М.: ИП Радиософт. – 2011. – 208 с.

20 Чесноков М., Цифровой гребенчатый фильтр с линией задержки продолжительностью в дробное число отсчетов Научно-технические ведомости СПбГПУ 5' (181) 2013 Информатика. Телекоммуникации. Управление. – С. 9–15.

случае используется вариант адаптивного фильтра, рассмотренный в работе[21]. При этом выходной сигнал снимается с выхода компенсатора, а задержка выбирается исходя из времени корреляции РС. Полученный алгоритм фильтрации РС работает в условиях широкополосного некоррелированного шума. Частотная характеристика такого фильтра представляет собой характеристику гребенчатого фильтра.

Определенно, сложная структура РС и нестационарность процесса речеобразования значительно снижают эффективность метода и приводят лишь к незначительному повышению разборчивости.

В работе[17] приведен пример цифровой реализации гребенчатого фильтра с интерполяцией по двум отсчетам. В исследовании продемонстрирована возможность улучшения *SNR* при тестировании ГФ на обработке сигнала в аддитивной смести с белым шумом на 4–10 дБ. Однако отмечается, что величина квазистационарных интервалов в вокализованной речи и время переходных процессов фильтра не допускают применения коэффициентов обратной связи больше 0,6, в связи с чем среднее улучшение *SNR* при применении ГФ ограничивается лишь 6 дБ.

Также существуют методы шумопонижения РС, основанные на периодичности вокализованных участков речи, смысл которых заключается в фильтрации верхних частот с последующим клиппированием. Повышение разборчивости осуществляется за счет выделения и усиления частотных компонент РС, несущих основную информацию о его формантной структуре.

Поскольку помеха уменьшает модуляционную глубину исходного РС, то повысить разборчивость речи можно путем её искусственного увеличения в определенном диапазоне частот. Экспериментальное исследование показало, что некоторое повышение разборчивости можно получить увеличением модуляционной глубины РС до искажения, однако такое улучшение разборчивости незначительно и на восприятие РС практически не влияет.

Методы временной обработки, основанные на модели РС

Большая часть описанных выше методов улучшения качества РС направлены на подавление шума. Эти методы нарушают спектральный баланс РС, что сопровождается его искажением.

В работах [10, 11, 12] предложен метод улучшения качества РС, использующий характеристики источника речи, в частности выходной сигнал фильтра линейного предсказания (ФЛП). Основа подхода к повышению качества РС заключается в выявлении в зашумленной речи участков с большими SNR и сохранении этих участков неизменными в отличие от участков РС с малыми значениями SNR. Отсчеты выхода ФЛП умножаются на весовую функцию, и модифицированный остаточный сигнал подается на вход полюсного фильтра, на выходе которого синтезируется улучшенный РС.

В работе[22] предложен алгоритм, отличие которого от базового алгоритма заключается в том, что весовая функция для остатков ФЛП конструируется с использованием критерия оптимизации. Очищенный РС синтезируется на выходе изменяющего во времени характеристики полюсного фильтра, на вход которого поступают остатки ФЛП.

В работе[23] предложено использовать огибающую преобразования Гильберта для реконструирования взвешенного сигнала остатков ФЛП. Огибающая преобразования Гильберта является хорошим индикатором вокализованных звуков. Поэтому применение к остаткам ФЛП преобразования Гильберта в качестве весовой функции приводит к контрастированию периодической структуры тональной речи.

В исследовании[24] проведена оценка эффективности метода линейного предсказания, предназначенного для подавления стационарных помех. Для фонограмм с различными исходными SNR (минус 10...0 дБ) в среднем приращение значения *SNR* составило около 10 дБ.

В работе[25] предложен алгоритм на основе медианного фильтра и фильтров краткосрочного (STP) и долгосрочного (LTP) линейного предсказания. Медианный фильтр в данном алгоритме предназначен для удаления нестационарных шумов в РС. Поскольку медианный фильтр сохраняет только медленно меняющиеся компоненты входного сигнала, он может исказить характеристику быстро меняющейся области речи.

Следовательно, необходим дополнительный этап предварительной обработки для сохранения характеристик речи перед применением медианного фильтра. На этапе предварительной обработки используется фильтр краткосрочного линейного предсказания (STP) и фильтр долгосрочного предсказания (LTP).

Приращение значения сегментарного SNR для исходных фонограмм с исходными *SNR* -4 дБ и -5 дБ составило 10 дБ. Однако следует отметить, что в источнике отсутствуют сведения о типе помех, которыми «зашумлялись» исходные фонограммы.

Всем методам линейного предсказания, построенным на предположении о линейности передаточной функции голосового тракта и возможности представления РС в любой произвольный момент времени

21 Дворянкин С. В. Цифровая шумоочистка аудиоинформации. – М.: ИП Радиософт. – 2011. – 208 с.

22 W. Jin, and M. S. Scordils, Speech enhancement by residual domain constrained optimization, Speech Communication, vol.48, pp.1349–1364, 2006.

23 B. Yegnanarayana et al., Speech enhancement using excitation source information, Proc. Int. conf. ICASSP, pp.1541–1544, 2002.

24 A. Kawamura, K. Fujii, Y. Itoh and Y. Fukui, A new noise reduction method using linear prediction error filter and adaptive digital filter, 2002 IEEE International Symposium on Circuits and Systems (ISCAS), Phoenix-Scottsdale, AZ, USA, 2002, pp. III-III, doi: 10.1109/ISCAS.2002.1010267.

25 Choi M.S., Kang H.G. Transient noise reduction in speech signal with a modified long-term predictor. EURASIP J. Adv. Signal Process. 2011, 141 (2011). https://doi.org/10.1186/1687-6180-2011-141

в виде линейной комбинации своих значений в предыдущие моменты, присущ один серьезный недостаток. В случае обработки сильно зашумленных РС точность вычисления коэффициентов линейного предсказания уменьшается. Это, в свою очередь, может еще сильнее ухудшить разборчивость сигнала на выходе системы линейного предсказания.

Методы адаптивного подавления помех

Методы адаптивного подавления обрабатывают параллельно искаженный сигнал и некий опорный сигнал [13]. При этом в качестве опорного сигнала используется сигнал, получаемый от датчиков, располагаемых в точках, где уровень речевого сигнала мал, то есть опорный сигнал максимально коррелирован с сигналом помехи. Следует отметить, что вычитание спектрограмм сигналов смеси и помех возможно как в режиме реального времени, так и в режиме отложенного анализа.

На первом этапе работы алгоритма создается оценка компонента, коррелированного с опорным сигналом, после чего он вычитается из смеси сигнала с шумом.

Метод адаптивного подавления реализован в двух типах систем, которые различаются способом получения опорного сигнала. В одноканальных системах (первый тип) опорный сигнал формируется из зашумленного с помощью различных преобразований последнего. В двухканальных системах используются два слабо коррелированных между собой источника смеси сигнала и шума.

Такой метод реализуют в режиме стереозаписи с двух микрофонов, которые находятся в разных точках пространства и по-разному ориентированы на источник звука.

Двухканальные системы сложны в реализации, однако, при корректных условиях размещения микрофонов и последующей правильной генерации опорного сигнала, двухканальные системы обеспечивают восстановление разборчивости даже крайне зашумленных сигналов.

Очевидно, что возможности данного метода ограничены возможностью получения опорного сигнала, максимально соответствующего помехе в искаженном сигнале, или – в случае вычитания помехи – образца помехи. В ряде случаев их удается получить позже, но тогда возникает достаточно сложная задача синхронизации сигналов основного и опорного каналов.

Методы реконструкции гармонической структуры спектральных описаний речи, искаженной шумами и помехами

Как известно, смысловая часть информации содержится в частотной огибающей РС, а основой конструкций РС являются вокализованные участки речи. Как правило, при наличии помех высшие гармоники, как наименее мощные, скрываются под шумом, а несколько мощных первых гармоник в низкочастотной области спектра с наибольшей амплитудой проявляются на фоне шумов. По этим оставшимся следам гармоник возможно нахождение значения частоты основного тона, восстановление гармонической структуры и, в итоге, звучания искаженного звукового сигнала.

В работе [14] показано, что имея в распоряжении только часть информации о гармонической структуре спектрограммы защищаемого речевого сигнала, другую её часть можно восстановить или реконструировать на основе свойств самого РС. Например, используя свойство кратности гармоник основного тона по оси частот спектрограммы. Также в указанной работе рассмотрен вариант шумоочистки, основанный на обнаружении «следов» помехи на изображении спектрограммы, синтезе помехи по найденным «следам» и её вычитании из исходной смеси.

В исследовании[26] представлены результаты параболической коррекции линий гармоник по вершинам парабол спектральных разверток. После проведенной коррекции достроенные высшие гармоники стали непрерывными и совпали с исходными зашумленными. В этой же работе отмечается, что при зашумлении РС белым шумом треки гармоник и восстановленная по ним гармоническая структура находится корректно даже при *SNR* до минус 12 дБ.

Несомненным достоинством методов реконструкции РС является то, что они работают даже в случае сильно зашумленного сигнала при условии, что треки нескольких первых гармоник различимы на фоне шумов. Однако в случае отсутствия треков первых гармоник или отсутствия возможности различения треков первых гармоник и других квазигармонических «следов» речеподобной помехи на сонограмме сигнала указанный метод не может быть успешно реализован.

Экспериментальные исследования эффективности средств шумопонижения

Для экспериментальных исследований авторами были выбраны типовые программные средства шумоочистки «Sound Cleaner», «GritTec's Noise Cancellation» и «Лазурь – М», в которых реализованы основные алгоритмы шумопонижения, рассмотренные ранее. В качестве тестовых РС были выбраны аудиозаписи таблиц слов (ГОСТ 16600-72) и фраз из (ГОСТ Р 50840-95).

Запись исходных тестовых РС проводилась в служебном помещении при уровне шума не более 35–40 дБ бригадой дикторов, не имевших дефектов

26 Дворянкин С. В., Алюшин В. М. Метод реконструкции гармонической структуры спектральных описаний искаженной шумами и помехами речи. М.: Известия института инженерной физики. – 2013. – т. 2. – № 28. – 57–62 с.

речи. Чтение слов и фраз дикторами осуществлялось ровным голосом, четко, но без подчеркивания отдельных звуков с постоянным уровнем речи. Дикторы выдерживали постоянный ритм речи на протяжении чтения всей таблицы.

Для формирования зашумленных тестовых РС использовалась ПЭВМ и специальное программное обеспечение Adobe Audition 1.5.

Для зашумления исходных тестовых РС использовались три наиболее эффективные помехи: шум со спектром, близким к нормированному спектру речи, помеха типа «речевой хор», а также «уличный» шум. Аудиофайлы формировались для *SNR* от –20 дБ до +10 дБ с шагом в 2 дБ.

Усредненные спектры тестовых сигналов в смеси с соответствующими помехами приведены на рисунках 2–4.

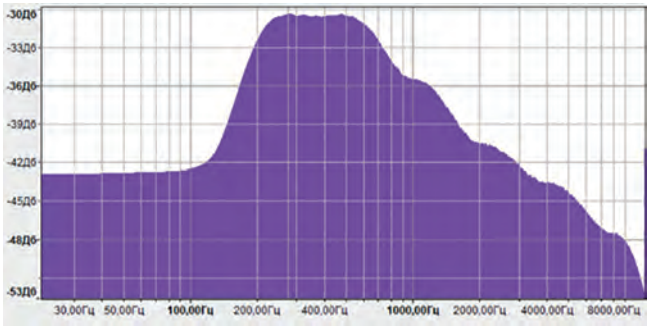

Рис. 2. Усредненная спектрограмма РС, зашумленного шумом со спектром, близким к нормированному спектру речи

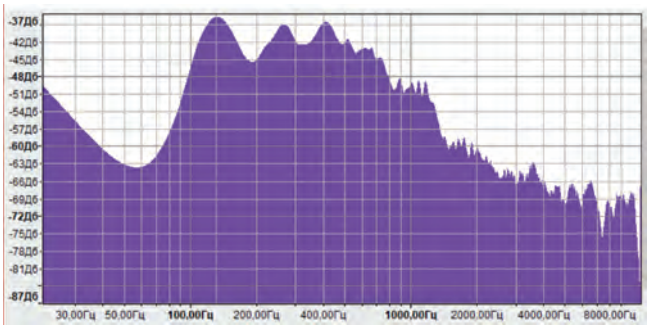

Рис. 3. Усредненная спектрограмма РС, зашумленного помехой типа «речевой хор»

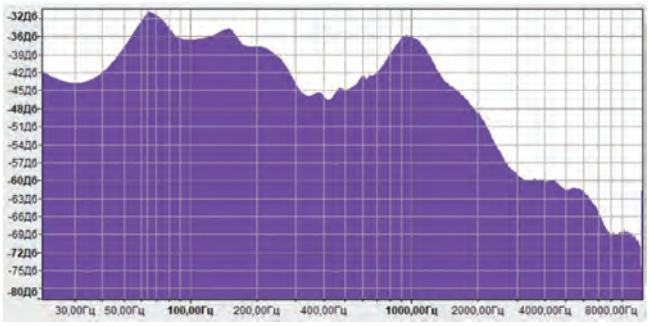

Рис. 4. Усредненная спектрограмма РС, полученного в условиях «уличного» шума

Из приведенных спектрограмм видно, что в диапазоне частот выше 1 кГц все помехи имеют спектр, близкий к «коричневому» шуму. При этом помехи типа «речевой хор» и «уличный шум» имеют локальные максимумы в диапазоне от 50 Гц до 1 кГц. Наличие указанных максимумов обусловлено, во-первых, наличием в указанном частотном диапазоне максимума нормированного спектра человеческой речи, и, во-вторых, фоновыми шумами (в большинстве своём индустриальными), которые имеют частотно зависимый характер и, как правило, уменьшаются по мере роста частоты.

Оценка разборчивости РС после шумопонижения проводилась артикуляционным методом путем оценки словесной разборчивости речи (отношения количества правильно понятых слов к их общему количеству в таблице).

Некоторые результаты исследования представлены на рис. 5.

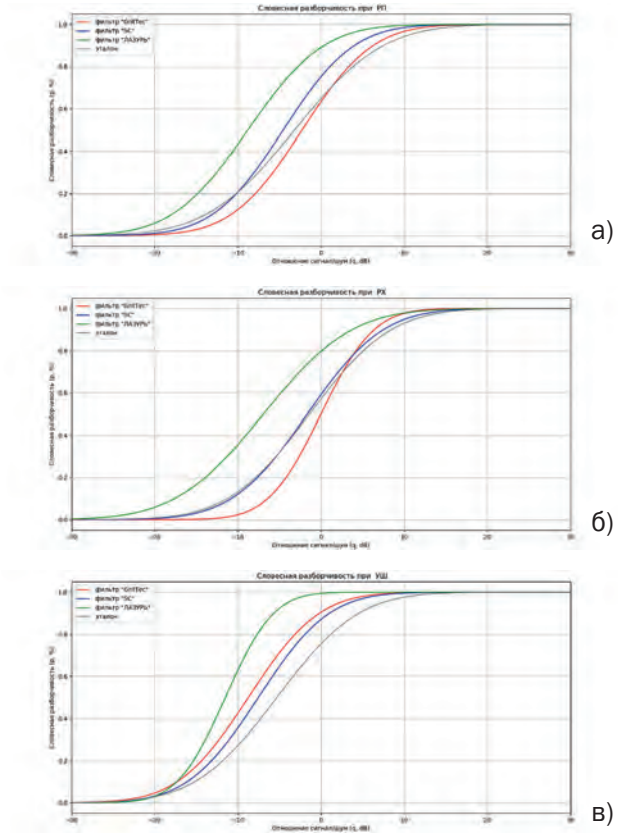

Рис. 5. Графики зависимости словесной разборчивости речи от SNR для исходной записи («эталон») и записей очищенными различными средствами шумоочистки (для средних значений) с помехами: шум со спектром, близким к нормированному спектру речи (а), «речевой хор» (б) и «уличный шум» (в) для средств шумопонижения: «GritTec's Noise Cancellation», «Sound Cleaner» и «Лазурь»

Исходя из зависимостей, представленных на рисунке 5, можно сделать вывод, что для аудиозаписей, полученных в условиях шума со спектром, близким

к нормированному спектру речи, наиболее эффективным является средство шумоочистки «Лазурь», обеспечивающее увеличение словесной разборчивости речи в области низких отношений сигнал/шум (от -16 до -6 дБ) на 22–36%. Несколько меньшую эффективность демонстрирует средство шумоочистки «Sound Cleaner». Применение средства шумоочистки «GritTec's Noise Cancellation» приводит к снижению словесной разборчивости речи.

Применение для аудиозаписей, полученных в условиях помехи типа «речевой хор», средства шумоочистки «Лазурь» обеспечивает увеличение словесной разборчивости речи в области низких отношений сигнал/шум (от -14 до -6 дБ) до 25–40%. Средство шумоочистки «Sound Cleaner» понижает словесную разборчивость при отношении сигнал/шум ниже -4 дБ, и незначительно повышает разборчивость при отношении сигнал/шум выше -4 дБ. Применение средства шумоочистки «GritTec's Noise Cancellation» приводит к снижению словесной разборчивости речи.

При этом наилучшие результаты шумоочистки аудиозаписей, полученных в условиях «уличного шума», демонстрирует средство шумоочистки «Лазурь», обеспечивающее увеличение словесной разборчивости речи в области низких отношений сигнал/шум (от -16 до -6 дБ) на 8–28%.

Проведенный эксперимент демонстрирует, что работа некоторых алгоритмов шумоочистки с сигналами может приводить к некоторому снижению разборчивости речи. Кроме того, по результатам анализа зависимостей словесной разборчивости речи от отношения «сигнал/шум», представленных на рис. 6, становится очевидным, что помеха типа «речевой хор» демонстрирует большую устойчивость по отношению к применяемым механизмам шумопонижения, нежели шум со спектром, близким к нормированной речи, и «уличный шум».

Стоит отметить, что на рисунке 5 представлены аппроксимированные функции зависимости словесной разборчивости речи от SNR. Представленные на указанном рисунке значения словесной разборчивости ввиду наличия аппроксимации несколько отличаются от фактически полученных в ходе эксперимента значений.

Таблица 1

Оценка словесной разборчивости речи и эффективности средств шумоочистки для шума со спектром, близким к нормированному спектру речи

№ п/п	Отношение сигнал/шум, дБ	Разборчивость без шумоочистки, %	Эффективность шумоочистки ΔW, %		
			«GritTec»	«Sound Cleaner»	«Лазурь»
1	- 20	0	0	0	0
2	- 18	0	0	1	2
3	- 16	2	-2	-1	5

Таблица 2

Оценка словесной разборчивости речи и эффективности средств шумоочистки для помехи типа «речевой хор»

№ п/п	Отношение сигнал/шум, дБ	Разборчивость без шумоочистки, %	Эффективность шумоочистки ΔW, %		
			«GritTec»	«Sound Cleaner»	«Лазурь»
1	- 20	0	0	0	0
2	- 18	0	0	0	0
3	- 16	0	0	0	6

Таблица 3

Оценка словесной разборчивости речи и эффективности средств шумоочистки для «уличного шума»

№ п/п	Отношение сигнал/шум, дБ	Разборчивость без шумоочистки, %	Эффективность шумоочистки ΔW, %		
			«GritTec»	«Sound Cleaner»	«Лазурь»
1	- 20	0	3	0	0
2	- 18	1	5	0	0
3	- 16	3	8	7	7

Обратим внимание на фактически полученные в ходе эксперимента значения словесной разборчивости РС без применения и с применением средств шумоочистки к аудиозаписям с отношением сигнал/шум менее -14 дБ (табл. 1–3).

Согласно результатам, представленным в табл. 1–3, применение различных методов шумоочистки к аудиозаписям с изначально малыми значениями отношения сигнал/шум не приводит к существенному увеличению словесной разборчивости. Приращение разборчивости на 3–5% крайне незначительно. Данные значения сопоставимы с ошибкой аппроксимации и попадают в доверительный интервал (0,95%) при построении зависимости словесной разборчивости от *SNR*.

Выводы

Необходимо отметить, что результаты работы различных методов шумопонижения и реконструкции РС, приведенные в проанализированных в ходе текущего исследования источниках, получены путем математического моделирования и вычислительных экспериментов. По результатам физического эксперимента методы шумопонижения и реконструкции РС по очевидным причинам продемонстрировали бы гораздо менее впечатляющие результаты.

Более того, стоит отметить, что важным аспектом корректной оценки полученных при оценке методов шумопонижения результатов является неоднозначность измерений среднеквадратической мощности (*RMS*) сформированных акустических сигналов, что не позволяет с должной точностью определить значения *SNR* смеси сигнала с шумом.

Для корректного расчета значений *SNR*, применительно к потоку слитной речи, необходимо вводить соответствующие поправки или лимитировать длительность пауз тестовых сигналов согласно характеристикам и параметрам слитной речи.

Ввиду сказанного очевидной становится практическая невозможность достаточно точного определения степени улучшения восприятия прошедшей процедуры шумопонижения фонограммы по приращению значений *SNR*.

Улучшение качества фонограммы, по отношению к которой применялись алгоритмы шумопонижения, можно оценить только субъективно (на слух), однако результаты оценки улучшения разборчивости в большинстве указанных исследований не приведены.

Более того, в подавляющем большинстве исследований отсутствуют сведения о диапазоне частот, в котором проводился анализ результатов работы методов шумопонижения.

Очевидно также, что прямое сравнение всех указанных методов по сведениям, приведенным в источниках, малоинформативно. Каждый из описанных методов оставляет в обработанных фонограммах свои артефакты, которые влияют на разборчивость речи ввиду субъективности восприятия, слушающего по-разному. Поэтому оценка эффективности того или иного алгоритма через объективный показатель *SNR* не дает полного представления о степени улучшения восприятия РС слушающим.

Кроме того, в рассмотренных работах не указано, каким образом проводилась оценка *SNR* для сигналов, прошедших процедуру шумопонижения.

Стоит также отметить, что нижняя граница *SNR*, значимых для задач оценки защищенности акустической речевой информации, находится в диапазоне около -25...-20 дБ.

Согласно работам Ю. С. Быкова[27] нижняя граница разборчивости РС в каналах связи для коррелированных тестов (команд) достигается при *SNR* -14 дБ. Судя по результатам, приведенным в проанализированных источниках, а также результатам экспериментальных исследований, методы шумопонижения и реконструкции РС также не могут преодолеть указанное Ю. С. Быковым пороговое значение, и при значении SNR исходной фонограммы менее -14 дБ не являются эффективными. Аналогичные выводы о невозможности применения механизмов шумопонижения к записям с низким *SNR* сделаны исследователями в [15].

В настоящее время алгоритмы быстрого вейвлет-преобразования демонстрируют большую эффективность нежели алгоритмы шумопонижения на основе быстрого преобразования Фурье или стационарного вейвлет-преобразования. При этом очевидно, что все указанные методы наилучшим образом работают с фонограммами, в которых наблюдается квазистационарный шум. При зашумлении фонограмм, например, речеподобной помехой указанные методы будут малоэффективны.

Методы фильтрации, основанные на оценке спектральных характеристик помех, могут быть эффективны только при попытке устранения стационарной помехи, спектральный состав которой заранее известен.

Методы коррекции и сглаживания спектра согласно результатам исследований демонстрируют весьма скромную эффективность. При этом существенным недостатком указанных методов шумопонижения является следующее: все алгоритмы коррекции спектра основаны на предположении, что уровень энергии спектра шума распределен по всему диапазону частот. Т.е. при применении для защиты информации речеподобной помехи указанные методы могут оказаться неэффективными.

27 Быков Ю. С. Теория разборчивости и повышения эффективности радиотелефонной связи / Ю.С. Быков. – М.: Госэнергоиздат, 1959. – 352 с.

Наиболее перспективными наравне с вейвлет-алгоритмами являются методы улучшения качества РС, использующие характеристики источника речи, в частности выходной сигнал ФЛП. Даже при достаточно низких *SNR* (порядка -10 дБ) данные методы демонстрируют относительно высокую эффективность.

При помощи методов реконструкции РС по трекам первых гармоник (следов фонообъектов) можно добиться значительного улучшения разборчивости РС (практически 100%) даже при очень маленьких значениях *SNR*. Однако в случае отсутствия треков первых гармоник на сонограмме сигнала указанный метод практически не может быть реализован.

Очевидно, что у всех рассмотренных методов шумопонижения и реконструкции РС есть частные, специфичные для каждого метода границы его применимости. Но у них также есть общее ограничение: все рассмотренные методы будут показывать весьма малую эффективность в случае применения для зашумления фонограммы корректно сформированной речеподобной помехи.

Постоянно меняющиеся частоты основного тона голосов дикторов, создающих помеху, делают практически невозможной задачу спектрального вычитания или фильтрации. Методы вейвлет-анализа и линейного предсказания в случае речеподобной помехи не способны разделить коэффициенты, соответствующие речеподобной помехе и РС. В случае применения методов реконструкции РС на сонограмме будет очень трудно выделить треки первых гармоник диктора на фоне треков гармоник голосов, формирующих речеподобную помеху.

Недостатки работы указанных алгоритмов с речеподобной помехой были продемонстрированы авторами в ходе эксперимента. Показано, что для двух программных сред «Sound Cleaner», «GritTec's Noise Cancellation» попытки применения алгоритмов шумопонижения записей с помехой типа «речевой хор» приводят к снижению словесной разборчивости речи.

Более того, в ходе практического эксперимента получены результаты, показывающие, что улучшение комфортности восприятия РС при применении процедур шумопонижения и реконструкции РС носит ограниченный характер. При хорошем качестве исходного РС после процедур шумопонижения из-за сопутствующих им искажений разборчивость речи может снижаться.

Следует также обратить внимание на спектральную структуру сигналов, прошедших процедуры шумопонижения: в них зачастую преобладают (наилучшим образом восстанавливаются) гармонические составляющие РС, т.е. вокализованные фонемы (аллофоны). Данное явление согласуется с тем фактом, что 60-70% разборчивости речи в потоке слитной речи дают именно вокализованные аллофоны.

Ввиду изложенного результаты проведенной работы по оценке возможностей методов шумопонижения и реконструкции РС, зашумляемых помехой типа «речевой хор», могут стать основой дальнейших исследований и разработок в области защиты речевой информации.

Литература

1. *Thimmaraja Y., Nagaraja B., Jayanna H., A spatial procedure to spectral subtraction for speech enhancement, Multimedia Tools and Applications volume 81, pages 23633–23647, 2022, https://doi.org/10.1007/s11042-022-12152-3.*
2. *Y. Yang, P. Liu, H. Zhou and Y. Tian, A Speech Enhancement Algorithm combining Spectral Subtraction and Wavelet Transform, 2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 2021, pp. 268–273, doi: 10.1109/AUTEEE52864.2021.9668622.*
3. *G. Ioannides, V. Rallis, Real-Time Speech Enhancement Using Spectral Subtraction with Minimum Statistics and Spectral Floor, 2023, https://doi.org/10.48550/arXiv.2302.10313.*
4. *A. Li, C. Zheng, R. Peng, and X. Li, On the importance of power compression and phase estimation in monaural speech dereverberation, JASA Express Lett., vol. 1, no. 1, pp. 014802, 2021.*
5. *T. Peer and T. Gerkmann, Phase-aware deep speech enhancement: It's all about the frame length, JASA Express Lett., vol. 2, no. 10, pp. 104802, 2022.*
6. *Бабурин А. В., Глущенко Л. А., Корзун А. М., Вейвлет-технологии для шумоочистки речевых сигналов в оптико-электронных каналах передачи информации, Информация и безопасность, 2023, Т. 26, вып. 1, с. 45–52, DOI 10.36622/VSTU.2023.26.1.006.*
7. *P. Kuwalek, W. Jesko, Speech Enhancement Based on Enhanced Empirical Wavelet Transform and Teager Energy Operator, Electronics 2023, 12(14), 3167; https://doi.org/10.3390/electronics12143167.*
8. *Лепендин А. А., Ильяшенко И. Д., Насретдинов Р. С., Применение обучаемого дискретного вейвлет-преобразования с адаптивными порогами для шумоочистки речевых сигналов, Высокопроизводительные вычислительные системы и технологии, том 4 (1), 2020, с. 51–57.*
9. *M. Talbi and M. S. Bouhlel, A New Speech Enhancement Technique Based on Stationary Bionic Wavelet Transform and MMSE Estimate of Spectral Amplitude Hindawi, Security and Communication Networks, vol. 2021, Article ID 9968275, 11 pages, 2021, https://doi.org/10.1155/2021/9968275.*
10. *X. Feng, N. Li, Z. He, Y. Zhang and W. Zhang, DNN-Based Linear Prediction Residual Enhancement for Speech Dereverberation, 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Tokyo, Japan, 2021, pp. 541–545.*

11. *Yang Liu, Na Tang, Xiaoli Chu, Yang Yang, Jun Wang, LPCSE: Neural Speech Enhancement through Linear Predictive Coding, Audio and Speech Processing, 2022, https://doi.org/10.48550/arXiv.2206.06908.*
12. *С. М. Горошко, С. Н. Петров. Метод шумоочистки речевых сигналов на основе мел-частотных кепстральных коэффициентов с использованием фильтрации Калмана / С. М. Горошко, С. Н. Петров // Известия Гомельского государственного университета имени Ф. Скорины. – 2019. – № 6 (117). – С. 103–107.*
13. *K. Tan, Z.-Q. Wang, and D. Wang, Neural spectrospatial filtering, IEEE/ACM Trans. Audio. Speech, Lang. Process., vol. 30, pp. 605–621, 2022.*
14. *Дворянкин С. В., Дворянкин Н. С., Устинов Р. А. Речеподобная помеха, стойкая к шумоочистке, как результат скремблирования защищаемой речи. Вопросы кибербезопасности, 2022, № 5(51). DOI: 10.21681/2311-3456-2022-5-14-27.*
15. *Иванов А. В., Волков Н. А. Применение методов шумоочистки для обработки речевой акустической информации, Сборник избранных статей научной сессии ТУСУР, номер 1–3, 2021, с. 34–37.*