Научная статья
УДК 004.773.5:[004.032.26+534.83+534.442]
https://doi.org/10.24143/2072-9502-2025-1-93-102
EDN UGLCHQ

Тестирование эффективности гибридной методики шумоподавления в речевом сигнале для системы видеоконференций

Сергей Валерьевич Белов✉, Сергей Сергеевич Катунин, Иван Юрьевич Кучин, Михаил Федорович Руденко

Астраханский государственный технический университет,
Астрахань, Россия, ssbelov@yandex.ru✉

Аннотация. Рассмотрен вопрос качества аудиосигнала во время проведения видеоконференций. Описано влияние шумов на качество и разборчивость речевого сигнала. Проведен анализ процесса шумоподавления в аудиосигнале в реальном времени. Выделены основные подходы к решению задачи шумоподавления в реальном времени, а именно подход с распознаванием и устранением шумов и подход с распознаванием голоса и устранением звуков, отличающихся от речевого сигнала. Разработан алгоритм интеллектуального шумоподавления с применением методов искусственного интеллекта, отличающийся наличием механизмов регулировки «интенсивности» воздействия шумоподавления на речевой сигнал, а также регулировки «чувствительности» к шуму, позволяющий влиять на эффективность работы алгоритма для специфичных условий эксплуатации без необходимости переобучения модели нейросети алгоритма. Разработано программное обеспечение, реализующее гибридную методику шумоподавления, в виде модуля, который был внедрен в ранее разработанный программный комплекс для организации и проведения видеоконференций. Выделено три типовых сценария, для каждого из которых отобраны эталонный (незашумленный) речевой сигнал, а также набор шумов для этих сценариев для проведения двух видов испытаний с целью тестирования эффективности модуля шумоподавления в различных условиях его применения. Проведены испытания с тестом эффективности шумоподавления сегментов шума во время речевых пауз с применением методов среднеквадратичного значения громкости сигнала RMS и визуального анализа спектрограммы с подсчетом процента количества сэмплов шумов в выходном сигнале. Проведены испытания с тестом эффективности шумоподавления зашумленного речевого сигнала с использованием объективных методов оценки качества речевого сигнала ViSQOL и NISQA. Результаты испытаний представлены в виде таблиц.

Ключевые слова: видеоконференция, речевой сигнал, шумоподавление, качество сигнала, спектр, фильтрующий коэффициент, полоса частот, шкала Барка, рекуррентные нейронные сети

Для цитирования: *Белов С. В., Катунин С. С., Кучин И. Ю., Руденко М. Ф.* Тестирование эффективности гибридной методики шумоподавления в речевом сигнале для системы видеоконференций // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2025. № 1. С. 93–102. https://doi.org/10.24143/2072-9502-2025-1-93-102. EDN UGLCHQ.

Original article

Testing a hybrid noise reduction technique effectiveness in a speech signal for a video conferencing system

Sergey V. Belov✉, Sergey S. Katunin, Ivan Yu. Kuchin, Mikhail F. Rudenko

Astrakhan State Technical University,
Astrakhan, Russia, ssbelov@yandex.ru✉

Abstract. The issue of audio signal quality during video conferences is considered. The effect of noise on the quality and intelligibility of the speech signal is described. The analysis of the noise reduction process in the audio signal in real time is carried out. The main approaches to solving the problem of noise reduction in real time are highlighted, namely the approach with recognition and elimination of noise and the approach with voice recognition and elimina-

tion of sounds different from the speech signal. An algorithm for intelligent noise reduction using artificial intelligence methods has been developed, characterized by the presence of mechanisms for adjusting the "intensity" of the noise reduction effect on the speech signal, as well as adjusting the "sensitivity" to noise, allowing to influence the efficiency of the algorithm for specific operating conditions without the need to retrain the neural network model of the algorithm. Software has been developed that implements a hybrid noise reduction technique in the form of a module that was implemented in a previously developed software package for organizing and conducting video conferences. Three typical scenarios have been identified, for each of which a reference (noiseless) speech signal has been selected, as well as a set of noises for these scenarios for conducting two types of tests in order to test the effectiveness of the noise reduction module in various conditions of its application. Experiments were carried out with a test of the effectiveness of noise reduction of noise segments during speech pauses using the methods of RMS signal volume and visual spectrogram analysis with calculation of the percentage of the number of samples of rhythms in the output signal. Experiments were carried out with a test of the effectiveness of noise reduction of a noisy speech signal using objective methods for evaluating the quality of the speech signal ViSQOL and NISQA. The test results are presented in the form of tables.

Keywords: video conferencing, speech signal, noise reduction, signal quality, spectrum, filtering coefficient, frequency band, Bark scale, recurrent neural networks

For citation: Belov S. V., Katunin S. S., Kuchin I. Yu., Rudenko M. F. Testing a hybrid noise reduction technique effectiveness in a speech signal for a video conferencing system. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics*. 2025;1:93-102. (In Russ.). https://doi.org/10.24143/2072-9502-2025-1-93-102. EDN UGLCHQ.

Введение

В эпоху цифровизации, а также в связи с недавней коронавирусной пандемией и последовавшей за ней самоизоляцией видеоконференцсвязь стала востребованным решением на современном рынке технологий [1–3]. По данным исследования аналитической компании *Fortune business insights*, объем рынка видеоконференцсвязи к 2032 г. увеличится вдвое и составит 60 млрд долл. Также одним из драйверов роста эксперты называют использование технологий искусственного интеллекта (ИИ) [4]. Однако во время проведения видеоконференций нередко возникают проблемы с качеством аудиосигнала, исходящего от участников. Речевой сигнал – основной способ передачи информации при естественном общении между людьми. Все звуки, не являющиеся человеческой речью, могут быть отнесены к категории «шум». Качество воспроизведения важно для любого слушателя: наличие в звуке шумов, которые не имеют отношения к содержанию (помехи, стуки, шипение, электрические щелчки и трески, гул, и т. д.), мешают восприятию звуковой информации. Сильные шумы и искажения способны не только ухудшить качество воспринимаемых сигналов, но и привести к снижению разборчивости речи [5, 6].

Поскольку процесс шумоподавления в системах видеоконференций рассматривается в условиях реального времени, то в процессе анализа и обработки аудиосигнала существуют задержки в несколько десятков миллисекунд и отсутствует возможность полноценного анализа спектра входящего аудиосигнала. Для решения этих проблем предлагается использовать методы искусственного интеллекта.

Однако по-прежнему наблюдается потребность в создании новых и в усовершенствовании существующих методов шумоподавления ввиду их эвристического характера. Например, с 2018 г. по настоящее время ежегодно в рамках международной программы ICASSP DNS CHALLENGE проводятся различные мероприятия для развития инноваций в сфере шумоподавления с использованием методов ИИ [10]. В определенных случаях при снижении зашумленности речевого сигнала существующие методы могут даже ухудшать его разборчивость. Также для некоторых методов характерны различные артефакты или новые искажения, осложняющие восприятие информации [6, 11].

Шумоподавление в речевом сигнале в реальном времени

Классические методы решения данной проблемы показывают низкую эффективность в условиях их применения в режиме реального времени, поскольку процесс шумоподавления осложняется необходимостью обработки сигнала с минимальной задержкой без возможности полноценного анализа спектра входящего аудиосигнала. В режиме реального времени допустимы задержки только в несколько десятков миллисекунд, за которые необходимо проводить анализ и обработку аудиосигнала.

Чаще всего полоса шумовых частот переменна и динамична, т. е. помимо статического фонового шума во входящий аудиосигнал могут попадать и различные шумы и иные нежелательные звуки, которые не были предусмотрены цифровым фильтром.

Для решения данной проблемы существуют алгоритмы шумоподавления, основанные на использовании методов машинного и глубокого обучения, способные анализировать спектр аудиосигнала за минимальный промежуток времени, например 10–20 мс [7–9].

На практике при решении задачи шумоподавления в реальном времени часто применяются эвристические методы ввиду сложности решаемой задачи.

Можно выделить следующие концептуально разные подходы:

– алгоритм анализирует спектр входящего аудиосигнала и при обнаружении шума исключает из выходного аудиосигнала найденный шум;

– алгоритм анализирует спектр входящего аудиосигнала и при обнаружении голоса (речевой сигнал) исключает все остальные звуки, не совпадающие с этим голосом.

Оба подхода имеют свои преимущества и недостатки. Например, подход с обнаружением голоса менее универсален, поскольку потребует дополнительного (и, возможно, длительного) обучения новому образцу голоса говорящего, но может давать более качественный результат, поскольку алгоритм будет устранять шумы, которые им изначально не были предусмотрены, при условии корректного определения голоса говорящего.

Первый подход в формальном виде представляет собой решение задачи бинарной классификации. В таком случае спектр сигнала анализируется и классифицируется с определенной вероятностью принадлежности к классу (уверенности) «голоса» либо «шума». В основе классификации лежит правило, согласно которому и происходит разделение заданного множества на две группы. Данная задача часто решается статистически с применением методов машинного обучения. Это вид обучения с учителем, когда классы предопределены, а данные помечаются заранее.

Таким образом, выделение признаков и формирование правила классификации происходит с помощью обучения на двух наборах данных посредством применения бинарной кросс-энтропии.

Второй же подход практически не встречается на практике как по причине своей низкой универсальности, так и по причине более сложной технической реализации. Этот подход подразумевает решение задачи одноклассной классификации, т. е. так называемой задачи выявления аномалий. Данная классификация во многих моментах схожа с кластеризацией, однако существуют гибридные подходы с применением классических статистических техник, таких как метод опорных векторов, байесовские сети или изолирующие леса, совместно с методами глубокого обучения с применением нейросетей [12].

Возвращаясь к первому подходу, следует отметить, что задача бинарной классификации часто решается с помощью алгоритмов, основанных на использовании методов машинного и глубокого обучения, таких как рекуррентные нейронные сети (РНС), которые были специально разработаны для решения задачи прогнозирования временных рядов [9]. Такие алгоритмы характеризуются способностью быстро адаптироваться к динамичному спектру шума, однако качество таких алгоритмов зависит от качества обучения модели нейронной сети, лежащей в основе алгоритма [7, 8].

Одной из самых используемых на практике РНС с применением архитектуры GRU является библиотека RNNoise [13, 14], которая является программной реализацией интеллектуального подхода к решению задачи шумоподавления.

Одним из выходов нейросети является множество фильтрующих коэффициентов для разных полос частот согласно шкале Барка. Другой выход нейросети является значением вероятности голосовой активности в сигнале, которое не используется для шумоподавления, но может быть применимо в других сценариях. Однако при всех своих достоинствах библиотека RNNoise имеет и ряд недостатков. Качество шумоподавления сильно зависит от качества и объема обучающей выборки, соответственно, может возникнуть ситуация, когда фильтр не устранит этот определенный шум из аудиосигнала по той причине, что обучающий набор данных не включает в себя этот звук шума. Также возможна и обратная ситуация, когда фильтр обрабатывает исходный звук слишком «агрессивно», негативно влияя на разборчивость и качество речевого сигнала.

Таким образом, становится актуальной задача модификации RNNoise, при достижении которой изменения позволят регулировать «интенсивность» шумоподавления с целью снижения воздействия алгоритма шумоподавления на выходной сигнал.

Также существует потребность в механизме, позволяющем усилить чувствительность алгоритма к шуму в реальном времени и, соответственно, увеличить эффект шумоподавления без необходимости переобучения модели на другом наборе данных.

Формирование понятия качества речевого сигнала и процесс его оценки

Под качеством речевого сигнала понимается процент соответствия параметров и характеристик исходного звука речи итоговому записанному речевому сигналу, а также процент зашумленности и количество посторонних звуков (не относящихся к исходному речевому сигналу) в записанном речевом сигнале [6, 11].

Также существенным фактором влияния на качество передаваемого речевого сигнала являются технические характеристики микрофона, а именно тип микрофона, диапазон воспринимаемых частот (Гц), чувствительность, характеристика направленности, уровень собственных шумов микрофона.

Следует отметить и окружающую обстановку (окружающие звуки), в условиях которой передается речевой сигнал, например, если звук передается в условиях жилого помещения, это могут быть фоновая музыка или фоновая человеческая речь, шумы, вызываемые передвижениями внутри домов, шум технического оборудования, наружные звуки от транспорта и пр. [5].

Существует множество различных методов оценки качества речевого сигнала: MOS (метод экспертных оценок) [15, 16], SNR (отношение сигнал/шум) [17], PESQ (объективный метод определения качества речи) [18, 19].

Субъективные методы оценки качества основываются на статистической обработке результатов работы большого числа слушателей-экспертов. Эти оценки существенно зависят от возраста и пола диктора, скорости произнесения фраз и других обстоятельств. Тесты при получении субъективных оценок проводят с имитацией реальных условий, например посторонний шум, фоновая речь других людей и т. п. Количественные результаты этих тестов отображают усредненное качество, уровень усилий слушателя, разборчивость, естественность звучания.

Наиболее широко используемая методика субъективной оценки качества описана в Рекомендации МСЭ Р. 800 и известна как методика MOS [16]. В соответствии с ней качество речи, получаемое при прохождении речевого сигнала от говорящего через систему связи к слушателю, оценивается как арифметическое среднее от всех оценок, выставляемых экспертами после прослушивания тестируемого канала передачи.

Задача любого объективного метода оценки качества речевого сигнала состоит в том, чтобы достичь высокой степени корреляции с субъективно-статистическими испытаниями, которые до сих пор остаются наиболее точной оценкой качества речи.

Большинство методов основано на сравнении оригинального и обработанного сигналов с помощью некоторой психоакустической модели. Оценивается степень заметности искажений в обработанном сигнале для человека. Под психоакустической моделью подразумевается модель, которая преобразует звуковой сигнал в его внутреннее представление с точки зрения слухового аппарата человека, которое и сравнивается с внутренним представлением исходного сигнала.

Наиболее распространенным объективным методом является оценка PESQ, определенная в рекомендации МСЭ-Т Р. 862 [18]. Она представляет собой объективную методику определения качества речевой связи в телефонных системах, которая прогнозирует результаты субъективной оценки качества этого вида связи слушателями-экспертами.

Для определения качества передачи речи в PESQ предусмотрено сравнение входного, или эталонного, сигнала с его искаженной версией на выходе системы связи. Результатом сравнения входного и выходного сигналов является оценка качества связи, которая аналогична усредненной субъективной оценке MOS.

Одной из открытых некоммерческих альтернатив метода PESQ является метод ViSQOL (Virtual Speech Quality Objective Listener), разработанный компанией Google. ViSQOL – это объективный показатель воспринимаемого качества звука. Он использует спектрально-временную меру сходства между эталонным и тестовым речевым сигналом для получения оценки MOS-LQO. Баллы MOS-LQO варьируют от 1 (худший результат) до 5 (лучший результат). Однако авторы метода отмечают, что ViSQOL разработан как метод для оценки ухудшения качества кодеков и VoIP-сети с помощью субъективного теста, аналогичного ITU-T P.800, и для других вариантов использования, таких как оценка качества речи после шумоподавления, регрессионное тестирование при предварительной обработке и генеративные модели, основанные на глубоком обучении, ViSQOL может показывать себя эффективно с некоторыми из них и неудачно с другими [20].

Также современным подходом является разработка методов объективной оценки качества речевого сигнала с применением методов ИИ, например метод NISQA – модель глубокого обучения, разработанная для прогнозирования качества речи. Отличительной особенностью NISQA является то, что она не требует наличия эталонного сигнала для предсказания оценки MOS. Помимо общего качества речи, NISQA также предоставляет прогнозы по параметрам качества: шум, окраска, прерывистость и громкость, – чтобы лучше понять причину ухудшения общей оценки качества MOS [21].

Гибридная методика шумоподавления

Гибридная методика представляет собой последовательность действий для шумоподавления в речевом сигнале (рис.) [22]:

1. Представление спектра входящего аудиосигнала (20 мс) в виде амплитудно-частотной характеристики (АЧХ) (960 элементов) алгоритма быстрого преобразования Фурье (БПФ, или FFT).

2. Выделение признаков звука в виде спектральной мощности полос частот согласно шкале Барка.

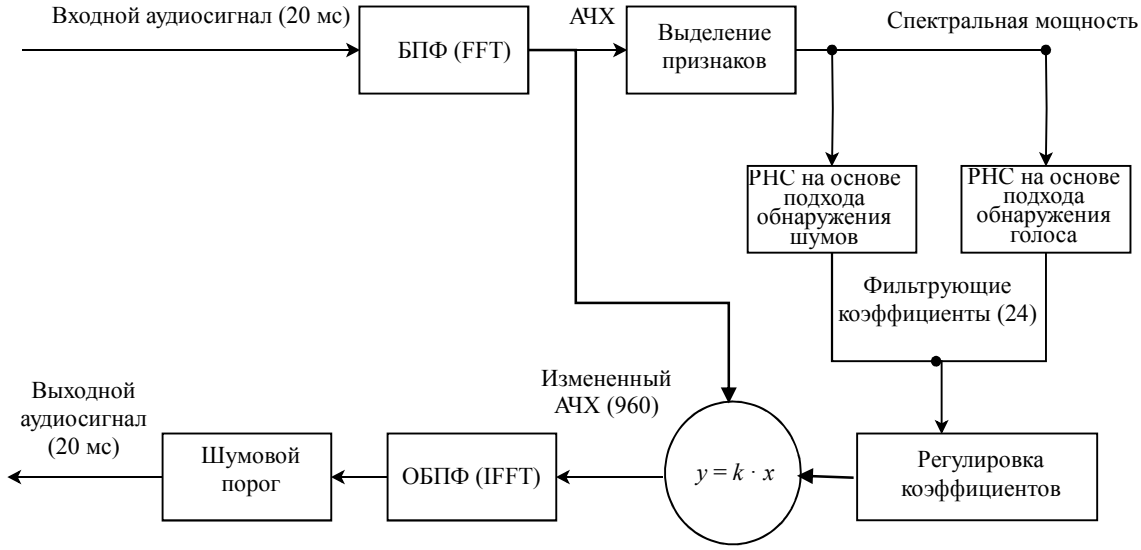
3. Применение РНС, основанной на подходе с обнаружением шумов либо на подходе с обнаружением голоса, и вычисление последовательности (24 элемента) фильтрующих коэффициентов со значениями, лежащими в диапазоне [0, 1].

4. Корректировка фильтрующих коэффициентов в соответствии с параметрами «интенсивности» и «чувствительности» алгоритма шумоподавления.

5. Изменение формы (АЧХ) аудиосигнала за счет использования фильтрующих коэффициентов.

6. Получение выходного аудиосигнала с помощью алгоритма обратного быстрого преобразования Фурье (ОБПФ, или IFFT).

7. Отсеивание шумов, громкость которых ниже громкости речевого сигнала, на основе метода шумового порога.



Концептуальное представление гибридной методики шумоподавления

Conceptual representation of a hybrid noise reduction technique

В методике предусмотрены механизмы, позволяющие влиять на эффективность работы алгоритма для специфичных условий эксплуатации без необходимости переобучения модели нейросети алгоритма путем регулировки фильтрующих коэффициентов: «интенсивности» воздействия шумоподавления на речевой сигнал, а также «чувствительности» к шуму.

Параметр интенсивности должен позволить пользователю снизить эффект шумоподавления в соответствии с установленным значением k_I, а параметр чувствительности – установить пороговое значение вероятности принадлежности текущего звука к классу «голос», т. е. полностью отсекать все звуки в тех полосах частот, значение коэффициентов которых меньше пороговой вероятности k_S.

Выходные фильтрующие коэффициенты РНС с предыдущего шага являются степенью уверенности нейросети в принадлежности к тому или иному классу («голос» или «шум»). Таким образом, при высоком проценте распознавания голоса данный параметр позволяет заглушать те звуки шумов, которые отсутствовали в обучающей выборке или которые недостаточно точно определяются РНС (например, если начения коэффициентов для таких звуков равно 0,5, то РНС в равной степени уверена в том, что текущий обрабатываемый звук является либо голосом, либо шумом):

$$g_M(b) = \begin{cases} g(b) + (1 - g(b)) \cdot (1 - k_I), & k_S \le g(b) \le 1 \\ 0, & 0 \le g(b) < k_S \end{cases},$$

где b – номер полосы частот от 1 до 24 в соответствии со шкалой Барка; $g_M(b)$ – модифицированный коэффициент фильтра на полосе частот b; $g(b)$ – оригинальный коэффициент фильтра на полосе частот b; k_I – степень интенсивности воздействия шумоподавления на речевой сигнал; k_S – параметр чувствительности алгоритма к шуму.

Тестирование эффективности гибридной методики шумоподавления

После анализа предметной области появляется возможность выделить три типовых сценария использования системы видеоконференций, в которых происходит передача речевых аудиопотоков:

– телемедицина. Для этого сценария характерен шум от медицинского оборудования (компрессор, насос, сплит-система, аппарат ИВЛ, кислородный концентратор и т. п.);

– дистанционное образование. Виды шумов: шум от компьютера/ноутбука, шум периферии (клавиатура, мышь), фоновая речь;

– общий случай, для которого характерен фоновый статический шум.

Для каждого из сценариев были отобраны эталонный (незашумленный) речевой сигнал и набор шумов для этих сценариев, после чего были проведены испытания с целью тестирования эффективности модуля шумоподавления в различных условиях его применения.

Были проведены два вида испытаний:

– тест эффективности шумоподавления сегментов шума во время речевых пауз;

– тест эффективности шумоподавления зашумленного речевого сигнала.

Шумоподавление сегментов шума во время речевых пауз, очевидно, является более простой задачей, чем шумоподавление зашумленного речевого сигнала, т. е. сигнала, в котором одновременно присутствуют речь и шум.

Для первого вида испытаний выходной сигнал в идеальном случае должен стать тишиной. Соответственно, измерение эффективности для первого вида испытаний проводится как среднеквадратичное значение громкости сигнала RMS (чем ближе оно к нулю в линейном виде либо к –100 Дб по шкале децибелов, тем эффективнее работает алгоритм шумоподавления).

Оценка производилась также методом визуального анализа спектрограммы (эффективнее тот алгоритм, у которого выходной сигнал на спектрограмме имеет меньшее количество звуков) и подсчетом процента количества сэмплов шумов в выходном сигнале по формуле

$$A_P = \frac{1}{N}\sum_{i=0}^{N} a_i \text{ при } |a_i| > 0,$$

где A_p – процент (доля) ненулевых значений амплитуд в сигнале; N – количество сэмплов в фрагменте звуковой дорожки; a_i – ненулевая амплитуда сигнала.

В сравнении применялась стандартная универсальная модель (матрица весов нейросети) RNNoise версии 0.1.1, при этом сравнивались классический алгоритм RNNoise и модифицированный алгоритм, где были введены степень «интенсивности» и степень «чувствительности» к шуму.

Результаты испытаний для тестирования эффективности шумоподавления сегментов шума во время речевых пауз представлены в табл. 1.

Таблица 1

Table 1

Результаты тестирования эффективности шумоподавления сегментов шума

Test results of testing the effectiveness of noise reduction of noise segments

Вид шума	Алгоритм	RMS, дБ	% шума
Клавиатура и мышь	Отсутствует	–44,9837	99,75
	Стандартный	–62,2252	12,98
	$k_S = 0{,}7$	–69,5511	0,62
Статический шум	Отсутствует	–60,5006	99,45
	Стандартный	–80,1755	8,27
	$k_S = 0{,}7$	–86,558	0,23
Стуки и удары	Отсутствует	–16,1699	99,86
	Стандартный	–30,9318	69,27
	$k_S = 0{,}7$	–38,1688	9,61
Аппарат ИВЛ (насос, компрессор)	Отсутствует	–25,5314	99,95
	Стандартный	–46,189	90,65
	$k_S = 0{,}7$	–56,6124	0,43

Исходя из данных, полученных при испытаниях, можно сделать вывод о том, что степень «чувствительности» к шуму позволила снизить как уровень громкости шумов, так и их количество в сигнале. В табл. 2 приведено сравнение результатов испытаний для различных сценариев.

Таблица 2

Table 2

Сравнительная таблица результатов тестирования шумоподавления сегментов шума

Comparative table of test results for noise reduction testing of noise segments

Вид шумов	RMS, дБ			% шума		
	Стандарт	k_S = 0,7	Разница, %	Стандарт, %	k_S = 0,7, %	Разница, %
Клавиатура и мышь	–62,2252	–69,5511	11,7 7	12,98	0,62	12,36
Статический шум	–80,1755	–86,558	7,96	8,27	0,43	7,84
Стуки и удары	–30,9318	–38,1688	23,39	69,27	9,61	59,66
Аппарат ИВЛ (насос, компрессор)	–46,189	–56,6124	22,56	90,65	0,43	90,22
Среднее значение, %	16,42			42,52		

Для оценки эффективности шумоподавления в зашумленном речевом сигнале были применены два метода: ViSQOL и NISQA, позволяющие объективно оценить качество сигнала по той же метрике MOS (субъективная оценка экспертами). С использованием данных можно получить воспроизводимые результаты, а также снизить роль человеческого фактора при испытаниях.

С помощью программной утилиты Audacity были получены тестовые образцы как результат объединения (микширования) эталонного речевого сигнала с различными шумами, соответствующие проводимому эксперименту.

Таким образом, были получены аудиодорожки той же длины, что и аудиодорожка с эталонным речевым сигналом.

Результаты испытаний для тестирования эффективности шумоподавления в зашумленном речевом сигнале представлены в табл. 3.

Таблица 3

Table 3

Результаты тестирования эффективности шумоподавления зашумленной речи

Test results for testing the effectiveness of noise reduction of noisy speech

Вид шума	Алгоритм	ViSQOL	NISQA (MOS)	NISQA (Noise)
Клавиатура и мышь	Отсутствует	2,29276	2,9913785	2,95772
	Стандартный	3,23077	3,2656085	3,8569636
	k_S = 0,35	3,277	3,7113245	4,0137954
Статический шум	Отсутствует	2,44946	3,7336133	3,0041766
	Стандартный	4,04534	4,0719028	4,281976
	k_S = 0,35	4,05035	4,2777042	4,2982836
Стуки и удары	Отсутствует	4,14048	3,7915282	4,0385704
	Стандартный	4,30956	4,0489206	4,364556
	k_S = 0,35	4,32137	3,671685	4,180684
Аппарат ИВЛ (насос, компрессор)	Отсутствует	2,71487	2,7581673	2,073441
	Стандартный	3,59706	2,8426402	3,443428
	k_S = 0,35	3,64842	3,170225	3,8501952

Анализируя полученные результаты экспериментов, можно прийти к выводу, что с использованием степени «чувствительности» к шуму на значении 0,35 удалось добиться улучшения качества речи на ~5–10 % относительно стандартного метода RNNoise.

Заключение

Разработана гибридная методика шумоподавления в реальном времени, которая отличается сочетанием классического метода шумового порога и двух алгоритмов шумоподавления, основанных на рекуррентной нейронной сети. Разработана программная реализация модуля шумоподавления для системы видеоконференций согласно предложенной методике. Проведено тестирование эффективности шумоподавления.

Анализ собранной статистики показал, что в среднем по выбранным наборам шумов для типовых сценариев при шумоподавлении сегментов шума уровень громкости шумов был снижен на ~16 %, а количество шумов в сигнале на ~42 %. При этом было достигнуто улучшение качества речи (согласно методам ViSQOL и NISQA) при шумоподавлении зашумленной речи в среднем на 5–10 %. Успешная апробация разработанной методики позволила сделать вывод об обоснованности выдвинутых положений и достоверности результатов работы.

Дальнейшее развитие исследования может быть связано с реализацией в методике подхода с обнаружением голоса, а именно подхода с одноклассной классификацией.

Список источников

1. Рудых Л. Г. Дистанционное обучение в вузе: проблемы и перспективы // Молодеж. вестн. ИрГТУ. 2020. Т. 10. № 2. С. 158–162.

2. Демина Н. В., Сабанова Л. В., Сабанова В. А. Видеоконференции и дистанционное обучение как основные виды телемедицинских услуг // Науч.-метод. электрон. журн. «Концепт». 2019. № V2. С. 28–33.

3. Савельев А. И. Архитектуры, алгоритмы и программные средства обработки потоков многомодальных данных в пиринговых веб-приложениях видеоконференцсвязи: автореф. дис. … канд. техн. наук. СПб., 2016. 17 с.

4. Video Conferencing Market Size, Share, Growth & Trends // Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/industry-reports/video-conferencing-market-100293 (дата обращения: 20.05.2024).

5. Бысько М. В. Шумология // ЭНЖ «Медиамузыка». 2014. № 3. С. 6.

6. Топников А. И. Оценка разборчивости и обработка речевых сигналов в задаче шумоподавления: автореф. дис. … канд. техн. наук. Владимир, 2012. 16 с.

7. Valin J.-M. A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement. URL: https://jmvalin.ca/papers/rnnoise_mmsp2018.pdf (дата обращения: 23.10.2022).

8. Yong Xu, Jun Du, Li-Rong Dai, Chin-Hui Lee. A Regression Approach to Speech Enhancement Based on Deep Neural Networks // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015. V. 23. Iss. 1. P. 7–19. URL: https://ieeexplore.ieee.org/document/6932438 (дата обращения: 18.11.2022).

9. Дубенко Ю. В., Дышкант Е. Е. Нейросетевой алгоритм выбора методов для прогнозирования временных рядов // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2019. № 1. С. 51–60.

10. Dubey H., Aazami A., Gopal V., Naderi B., Braun S., Cutler R., Ju A., Zohourian M., Tang M., Gamper H., Golestaneh M., Aichner R. ICASSP 2023 deep noise suppression challenge // arXiv.org e-Print archive. 9 May 2023. URL: https://arxiv.org/pdf/2303.11510 (дата обращения: 15.02.2024).

11. Афанасьев А. А. Модели и методы анализа и обработки речевого сигнала в системах связи: автореф. дис. … д-ра техн. наук. Орел, 2018. 16 с.

12. Raghavendra Chalapathy, Aditya Krishna Menon, Sanjay Chawla. Anomaly detection using one-class neural networks // arXiv.org e-Print archive. January 14, 2019. URL: https://arxiv.org/pdf/1802.06360 (дата обращения: 14.04.2024).

13. Канев А. И., Назаров М. М., Терентьев В. О., Усков Д. Ю. Исследование различных архитектур нейронных сетей для удаления шумов из аудио и видео // Искусственный интеллект в автоматизированных системах управления и обработки данных: сб. ст. Всерос. науч. конф. (Москва, 27–28 апреля 2022 г.). М.: МГТУ им. Н. Э. Баумана, 2022. Т. 2. С. 298–304.

14. Сайгин А. А., Плотникова Н. П. Обзор нейронных сетей для решения задачи разделения источников звука // Тенденции развития науки и образования. 2022. № 86-1. С. 89–93.

15. Полторак В. П., Моргаль О. М., Заика Ю. А. Оценка качества передачи речи в IP-телефонии // Молодой ученый. 2014. № 4. С. 121–123.

16. Methods for subjective determination of transmission quality // Recommendation ITU-T. 1996. P. 800 URL: https://www.itu.int/rec/T-REC-P.800-199608-I (дата обращения: 23.10.2022).

17. Топников А. И., Нестеров М. С., Новоселов С. А., Приоров А. Л. Неэталонная оценка разборчивости зашумленных речевых сигналов // Цифровая обработка сигналов. 2015. № 1. С. 39–44.

18. Perceptual evaluation of speech quality (PESQ) // Recommendation ITU-T P.862. 2001. URL: https://www.itu.int/rec/T-REC-P.862-200102-I (дата обращения: 23.10.2022).

19. Берко Г. А., Галич С. А., Пасюк А. О., Семенов Е. С. Применение алгоритма PESQ для оценки качества передачи речи по IP-сетям // Огарёв-online. 2015. № 11 (52). С. 3.

20. Chinen M., Lim F. S. C., Skoglund J., Gureev N., O'Gorman F., Hines A. ViSQOL v3: An Open Source Production Ready Objective Speech and Audio Metric // arXiv.org e-Print archive. 20 April 2020. URL: https://arxiv.org/pdf/2004.09584 (дата обращения: 19.04.2024).

21. Mittag G., Möller S. Non-intrusive Speech Quality Assessment for Super-wideband Speech Communication Networks // ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing

(ICASSP) (12–17 May 2019). URL: https://ieeexplore.ieee.org/document/8683770 (дата обращения: 25.04.2024).

22. Белов С. В., Катунин С. С. Гибридная методика шумоподавления в речевом сигнале для системы видеоконференций // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2023. № 1. С. 36–42.

References

1. Rudyh L. G. Distancionnoe obuchenie v vuze: problemy i perspektivy [Distance learning at the university: problems and prospects]. *Molodezhnyj vestnik IrGTU*, 2020, vol. 10, no. 2, pp. 158-162.

2. Demina N. V., Sabanova L. V., Sabanova V. A. Videokonferencii i distancionnoe obuchenie kak osnovnye vidy telemedicinskih uslug [Videoconferencing and distance learning as the main types of telemedicine services]. *Nauchno-metodicheskij elektronnyj zhurnal «Koncept»*, 2019, no. V2, pp. 28-33.

3. Savel'ev A. I. *Arhitektury, algoritmy i programmnye sredstva obrabotki potokov mnogomodal'nyh dannyh v piringovyh veb-prilozheniyah videokonferencsvyazi. Avtoreferat dissertacii ... kand. tekhn. nauk* [Architectures, algorithms, and software tools for processing multimodal data streams in peer-to-peer video conferencing web applications. Abstract of the dissertation ... Candidate of Technical Sciences]. Saint Petersburg, 2016. 17 p.

4. *Video Conferencing Market Size, Share, Growth & Trends. Fortune Business Insights*. Available at: https://www.fortunebusinessinsights.com/industry-reports/video conferencing-market-100293 (accessed: 20.05.2024).

5. Bys'ko M. V. *Shumologiya* [Noisology]. ENZH «Mediamuzyka», 2014, no. 3, p. 6.

6. Topnikov A. I. *Ocenka razborchivosti i obrabotka rechevyh signalov v zadache shumopodavleniya. Avtoreferat dissertacii ... kand. tekhn. nauk* [Evaluation of intelligibility and processing of speech signals in the noise reduction task. Abstract of the dissertation ... Candidate of Technical Sciences]. Vladimir, 2012. 16 p.

7. Valin J.-M. *A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement*. Available at: https://jmvalin.ca/papers/rnnoise_mmsp2018.pdf (accessed: 23.10.2022).

8. Yong Xu, Jun Du, Li-Rong Dai, Chin-Hui Lee. A Regression Approach to Speech Enhancement Based on Deep Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, vol. 23, iss. 1, pp. 7-19. Available at: https://ieeexplore.ieee.org/document/6932438 (accessed: 18.11.2022).

9. Dubenko Yu. V., Dyshkant E. E. Nejrosetevoj algoritm vybora metodov dlya prognozirovaniya vremennyh ryadov [Neural network algorithm for selecting methods for time series forecasting]. *Vestnik Astrahanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika*, 2019, no. 1, pp. 51-60.

10. Dubey H., Aazami A., Gopal V., Naderi B., Braun S., Cutler R., Ju A., Zohourian M., Tang M., Gamper H., Golestaneh M., Aichner R. *ICASSP 2023 deep noise suppression challenge. arXiv.org e-Print archive*. 9 May 2023. Available at: https://arxiv.org/pdf/2303.11510 (accessed: 15.02.2024).

11. Afanas'ev A. A. *Modeli i metody analiza i obrabotki rechevogo signala v sistemah svyazi. Avtoreferat dissertacii ... d-ra tekhn. nauk* [Models and methods of speech signal analysis and processing in communication systems. Abstract of the dissertation ... Doctor of Technical Sciences]. Orel, 2018. 16 p.

12. Raghavendra Chalapathy, Aditya Krishna Menon, Sanjay Chawla. Anomaly detection using one-class neural networks. *arXiv.org e-Print archive*. January 14, 2019. Available at: https://arxiv.org/pdf/1802.06360 (accessed: 14.04.2024).

13. Kanev A. I., Nazarov M. M., Terent'ev V. O., Uskov D. Yu. Issledovanie razlichnyh arhitektur nejronnyh setej dlya udaleniya shumov iz audio i video [Investigation of various neural network architectures for removing noise from audio and video]. *Iskusstvennyj intellekt v avtomatizirovannyh sistemah upravleniya i obrabotki dannyh: sbornik statej Vserossijskoj nauchnoj konferencii (Moskva, 27–28 aprelya 2022 g.)*. Moscow, MGTU im. N. E. Baumana, 2022. Vol. 2. Pp. 298-304.

14. Sajgin A. A., Plotnikova N. P. Obzor nejronnyh setej dlya resheniya zadachi razdeleniya istochnikov zvuka [An overview of neural networks for solving the problem of sound source separation]. *Tendencii razvitiya nauki i obrazovaniya*, 2022, no. 86-1, pp. 89-93.

15. Poltorak V. P., Morgal' O. M., Zaika Yu. A. Ocenka kachestva peredachi rechi v IP-telefonii [Evaluation of the quality of speech transmission in IP telephony]. *Molodoj uchenyj*, 2014, no. 4, pp. 121-123.

16. Methods for subjective determination of transmission quality. *Recommendation ITU-T*, 1996, p. 800 Available at: https://www.itu.int/rec/T-REC-P.800-199608-I (accessed: 23.10.2022).

17. Topnikov A. I., Nesterov M. S., Novoselov S. A., Priorov A. L. Neetalonnaya ocenka razborchivosti zashumlennyh rechevyh signalov [Non-electronic assessment of intelligibility of noisy speech signals]. *Cifrovaya obrabotka signalov*, 2015, no. 1, pp. 39-44.

18. Perceptual evaluation of speech quality (PESQ). *Recommendation ITU-T P.862*. 2001. Available at: https://www.itu.int/rec/T-REC-P.862-200102-I (accessed: 23.10.2022).

19. Berko G. A., Galich S. A., Pasyuk A. O., Semenov E. S. Primenenie algoritma PESQ dlya ocenki kachestva peredachi rechi po IP-setyam [Application of PESQ algorithm to evaluate the quality of speech transmission over IP networks]. *Ogaryov-online*, 2015, no. 11 (52), p. 3.

20. Chinen M., Lim F. S. C., Skoglund J., Gureev N., O'Gorman F., Hines A. ViSQOL v3: An Open Source Production Ready Objective Speech and Audio Metric. *arXiv.org e-Print archive*. 20 April 2020. Available at: https://arxiv.org/pdf/2004.09584 (accessed: 19.04.2024).

21. Mittag G., Möller S. Non-intrusive Speech Quality Assessment for Super-wideband Speech Communication Networks. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (12-17 May 2019)*. Available at: https://ieeexplore.ieee.org/document/8683770 (accessed: 25.04.2024).

22. Belov S. V., Katunin S. S. Gibridnaya metodika shu-

mopodavleniya v rechevom signale dlya sistemy videokonferencij [Hybrid noise reduction technique in a speech signal for a video conferencing system]. *Vestnik Astrahanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika i informatika*, 2023, no. 1, pp. 36-42.

Статья поступила в редакцию 15.10.2024; одобрена после рецензирования 23.12.2024; принята к публикации 20.01.2025
The article was submitted 15.10.2024; approved after reviewing 23.12.2024; accepted for publication 20.01.2025

Информация об авторах / Information about the authors

Сергей Валерьевич Белов – кандидат технических наук, доцент; директор института информационных технологий и коммуникаций; Астраханский государственный технический университет; ssbelov@yandex.ru

Sergey V. Belov – Candidate of Technical Sciences, Assistant Professor; Director of Institute of Information Technologies and Communications; Astrakhan State Technical University; ssbelov@yandex.ru

Сергей Сергеевич Катунин – ассистент кафедры автоматизированных систем обработки информации и управления; Астраханский государственный технический университет; sulmpx60@yandex.ru

Sergey S. Katunin – Lecturer of the Department of Automated Control and Data Processing Systems; Astrakhan State Technical University; sulmpx60@yandex.ru

Иван Юрьевич Кучин – кандидат технических наук; доцент кафедры информационной безопасности; Астраханский государственный технический университет; iri@astu.org

Ivan Yu. Kuchin – Candidate of Technical Sciences; Assistant Professor Department of Information Security; Astrakhan State Technical University; iri@astu.org

Михаил Федорович Руденко – доктор технических наук, профессор; профессор кафедры безопасности жизнедеятельности и инженерной экологии; Астраханский государственный технический университет; mf.rudenko@mail.ru

Mikhail F. Rudenko – Doctor of Technical Sciences, Professor; Professor of the Department of Life Safety and Environmental Engineering; Astrakhan State Technical University; mf.rudenko@mail.ru