DOI: 10.55648/1998-6920-2024-18-2-43-56 УДК 004.056.53

Подход к оценке защищенности речевой акустической информации с применением нейронных сетей

Н. А. Волков, А. В. Иванов

Самарский государственный технический университет (СамГТУ)

Аннотация: В работе рассматривается методика оценки защищенности речевой акустической информации при подготовке помещений для проведения закрытых переговоров. Авторами предложена структурная схема этапов создания интеллектуальной системы, в которой с учетом недостатков существующих подходов используются методы распознавания, основанные на сверточных нейронных сетях. Описывается процесс формирования обучающего набора данных в формате аудиозаписей с наложенными шумами с различными отношениями сигнал/шум. Рассматриваются возможности аудиоредактора Adobe Audition и библиотек Python для формирования наборов данных. Предлагается классифицировать спектрограммы либо мел-частотные кепстральные коэффициенты аудиозаписей с помощью нейронной сети по процентам разборчивости речи с целью автоматизации процесса оценки защищенности речевой акустической информации. Для достижения требуемого результата планируется обучить нейронную сеть на различных данных, провести сравнительный анализ с существующим подходом, оценить производительность системы и провести валидацию результатов. Предложенный подход и его практическое применение позволят значительно повысить качество и расширить условия применения оценки защищенности речевой акустической информации.

Ключевые слова: глубокие нейронные сети, сверточные нейронные сети, отношение сигнал/шум, зашумленность аудиозаписи, распознавание речи, спектрограммы, мел-частотные кепстральные коэффициенты, оценка защищенности речевой акустической информации.

Для цитирования: Волков Н. А., Иванов А. В. Подход к оценке защищенности речевой акустической информации с применением нейронных сетей // Вестник СибГУТИ. 2024. Т. 18, № 2. С. 43–56. https://doi.org/10.55648/1998-6920-2024-18-2-43-56.

Статья поступила в редакцию 13.10.2023;
переработанный вариант – 20.12.2023;
принята к публикации 22.12.2023.

1. Введение

Одной из ключевых задач в области обеспечения информационной безопасности является защита акустической речевой информации при проведении переговоров. Закрытая информация может быть подслушана при помощи технических средств акустической и виброакустической разведки [1].

Для оценки уровня защищенности от утечки информации по техническим каналам в помещении, которое предназначено для проведения закрытых переговоров, чаще всего используется общепринятый инструментально-расчетный подход, предложенный А. А. Хоревым,

В. К. Желязняком и Ю. К. Макаровым [2] на основании результатов экспериментальных исследований, проведенных Н. Б. Покровским [3]. Важно отметить, что формантный метод Н. Б. Покровского был предложен для оценки качества телефонных линий связи, поэтому условия проведения экспериментов существенно отличаются от условий оценки защищенности речевой информации. В связи с этим данный подход обладает рядом недостатков при его использовании для оценки защищенности помещений от утечки по техническим каналам акустической речевой информации [4]. Перечислим основные из них:

1) в качестве принимающей стороны не рассматривался злоумышленник, обладающий возможностью записи сигнала, многократного прослушивания, обработки сигнала (шумоочистка);

2) экспериментальные исследования (артикуляционные испытания, на основании которых были выведены все зависимости, используемые в методике [3]) проводились с использованием таблиц (слоговых, словесных и т.д.), где элементы были максимально некоррелированы между собой. Однако в реальных ситуациях мы имеем дело со связными, осмысленными текстами, что позволяет злоумышленнику понять те слова, которые не были им распознаны по общему содержанию переговоров;

3) не рассматривались условия с высоким уровнем маскирующих шумов, что оказывает влияние на зависимости восприятия речевых сигналов;

4) не учитываются индивидуальные особенности речи говорящих, всё основывается на усредненном спектре речи;

5) все зависимости амплитудного состава речи Н. Б. Покровским были получены при условии, что источник звука находился в 8 см от микрофона, в задачах же защиты информации все измерения и запись сигнала производятся на расстояниях более 1 м. Из этого следует, что использовать интегральные уровни по методу Н. Б. Покровского в наших задачах нельзя. Кроме того, спектры сигналов, измеренные на расстояниях 8 см и 1 м от микрофона, могут отличаться;

6) не рассматривался эффект форсирования речи (повышение уровня речи, вызванное усиленным напряжением голосовых связок) [5];

7) маскировка речи учитывается только с помощью помех на основе белого шума, исключая возможность оценки защищенности от помех типа «речевой хор» (такая помеха формируется путем смешения фрагментов речи нескольких дикторов) [6];

8) не учитывается влияние конфигурации помещения на оценку разборчивости речи (форма помещения, размер помещения, реверберация, поглощение звука) [7, 8].

Указанные недостатки подчеркивают необходимость корректировки существующих подходов к оценке защищенности речевой акустической информации. Корректировки должны позволить учесть реальные условия защиты информации, связность и осмысленность речи дикторов, разнообразие помех и индивидуальные особенности дикторов.

Реализация скорректированного подхода становится возможной при переходе от формантного метода к оценке защищенности речевой акустической информации с помощью интеллектуальной системы на основе нейронной сети. В случае реализации подобного подхода станет возможным учесть ряд существующих недостатков.

2. Нейросетевой подход к оценке защищенности речевой акустической информации

При оценке защищенности помещения от утечки информации по техническим каналам оператором проводится инструментально-расчетная оценка разборчивости речи в контрольных точках помещения. При проведении оценки используется комплект контрольно-измерительной аппаратуры, состоящий из генератора тестовых сигналов (в качестве которого чаще всего используется генератор белого шума), усилителя мощности, акустического излучателя, измерительного датчика (микрофон и акселерометр) и измерительного устройства

– шумомера со встроенными фильтрами (октавными). Также в контрольной точке устанавливают средство защиты информации для создания маскирующих помех. Состав измерительной установки приведен на рис. 1.

Рис. 1. Состав измерительной установки при оценке защищенности акустической речевой информации в помещении

Оценка защищенности производится в следующей последовательности:

1) измеряется уровень тестового сигнала, излучаемого акустической системой;

2) измеряется уровень фоновых шумов или шумов от средств защиты информации (СЗИ) в контрольной точке;

3) измеряется уровень смеси сигнал+шум в контрольной точке;

4) рассчитывается значение словесной разборчивости;

5) делается вывод о соответствии полученного значения норме.

На рис. 2 представлена предлагаемая интеллектуальная система оценки защищенности речевой акустической информации, в составе которой: источник речевого тестового сигнала, усилитель мощности, акустический излучатель, измерительный датчик (микрофон или акселерометр) и модуль оценки защищенности речевой акустической информации на основе нейронной сети. Также в контрольной точке установлено средство защиты информации.

Рис. 2. Схема предлагаемой интеллектуальной системы оценки защищенности речевой акустической информации

Алгоритм работы оператора с предлагаемой интеллектуальной системой оценки защищенности речевой акустической информации почти не отличается от общепринятой инструментально-расчетной оценки разборчивости речи в контрольных точках помещения – трудоемкость оператора не увеличивается, но вместо генератора тестовых сигналов мы используем источник речевого тестового сигнала; роль измерительного устройства теперь выполняет модуль оценки защищенности речевой акустической информации на основе нейронной сети.

Опишем более подробно этапы разработки интеллектуальной системы оценки защищенности речевой акустической информации на основе нейронной сети. Данные этапы представлены на рис. 3.

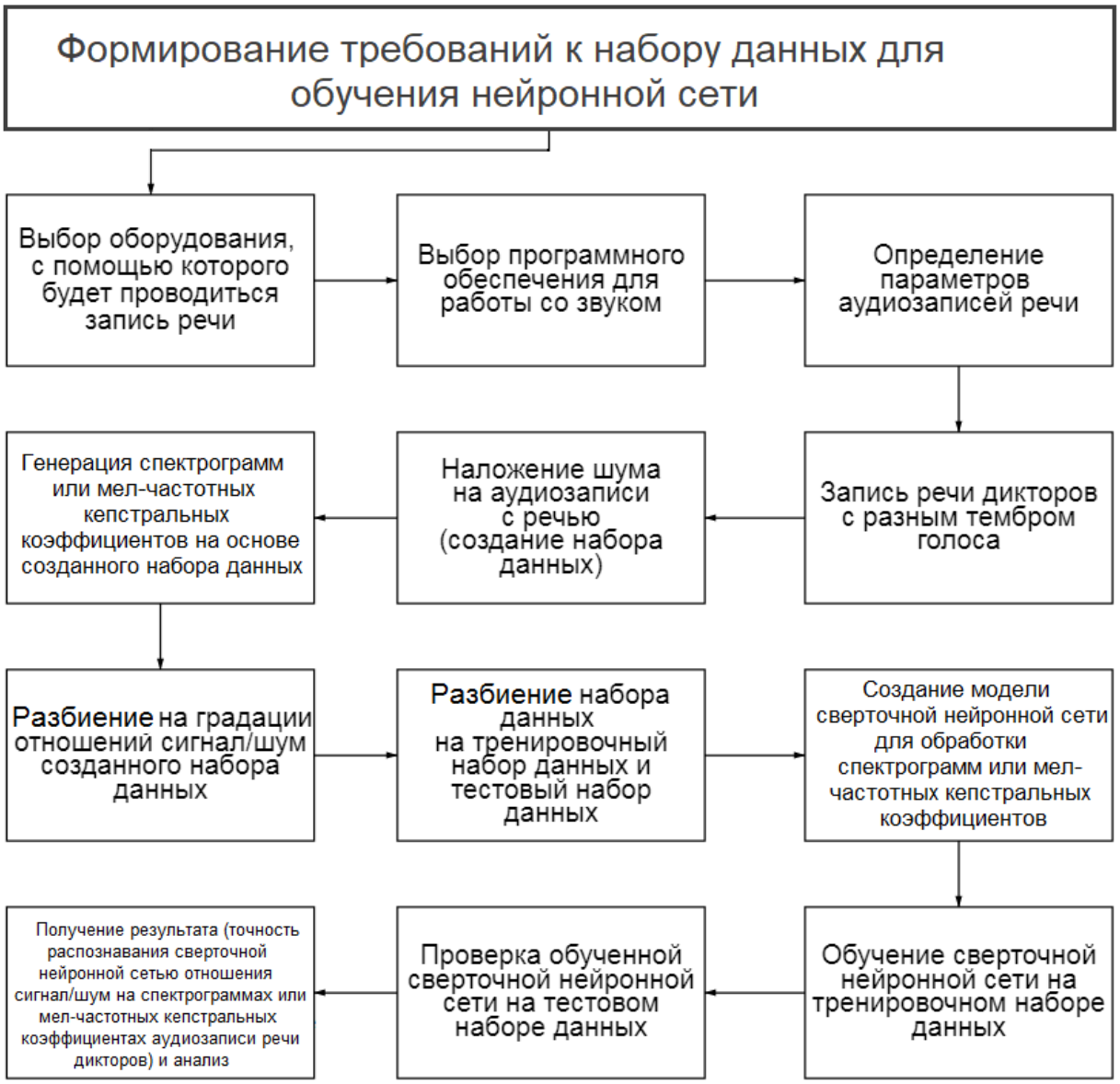

Рис. 3. Структурная схема этапов создания интеллектуальной системы оценки защищенности речевой акустической информации на основе нейронной сети

Рассмотрим более подробно приведенные этапы.

Для начала необходимо сформулировать требования к набору данных для обучения нейронной сети, а именно:

– какой вид данных будет использоваться для обучения нейронной сети – аудиозаписи или изображения;

– какое общее количество образцов будет использоваться в наборе данных;

– по каким классификациям необходимо разделить набор данных, а также необходимо определить количество таких классификаций;

– сколько образцов будет содержаться в тренировочном и тестовом наборах данных.

Для проведения записи голосов дикторов необходимо использовать программное обеспечение, которое имеет функционал обработки звука. Adobe Audition является широко признанным аудиоредактором, который пользуется популярностью среди исследователей в области защиты речевой информации [9, 10, 11]. Это программное обеспечение поддерживает все современные аудиоформаты и предлагает разнообразные функции. Встроенный спектральный анализатор позволяет анализировать спектр звука, а графический эквалайзер с тридцатью полосами предоставляет гибкую возможность регулировки звука. Программа позволяет корректировать фазу, тон и скорость воспроизведения, а также может генериро-

вать различные типы шумов (белый, розовый, коричневый). Дополнительно Adobe Audition обладает функциями исправления и удаления дефектов, таких как щелчки и шумоподавление, а также предоставляет множество других возможностей [12].

Для записи речи без искажения частотных характеристик и контроля уровня сигналов необходимо использовать измерительные микрофоны и шумомер с функцией записи сигнала либо с модулями аналого-цифрового преобразования для акустики (например, LTR 24-2).

Отметим также, что необходимо определить, для какого количества дикторов и с каким тембром голоса нужно провести запись речи (мужского низкого голоса, мужского высокого голоса, женского низкого голоса, женского высокого голоса).

Прежде чем проводить запись речи дикторов, необходимо определить параметры получаемых аудиодорожек, а именно:

– Частота дискретизации аудиозаписи. Обычно используется частота дискретизации, равная 44.1 кГц или 48 кГц.

– Длительность аудиозаписи. Определяется экспериментальным путем, начиная с длительности в 10 секунд, оценивая влияния длительности на результат обучения нейронной сети. Кроме того, необходимо определить, будет ли аудиозапись речи диктора иметь паузы или необходимо отредактировать аудиозапись и вырезать все паузы.

– Формат, в котором необходимо сохранять файлы. Предлагается использовать формат wav, так как он обеспечивает сохранение звука без потерь [19].

После проведения записи голосов дикторов мы получим несколько аудиозаписей – набор данных чистой речи.

Набор данных для обучения должен представлять из себя аудиозаписи с наложенными шумами различных спектров (для первоначального эксперимента предлагается взять белый и розовый шумы), имитирующими случай с оценкой защищенности речевой информации при использовании СЗИ. Для генерации шумов и наложения на первоначальном этапе предлагается воспользоваться программным обеспечением Adobe Audition. Но в дальнейшем для набора достаточного количества данных для обучения, поскольку процесс наложения шумов на аудиозаписи речи дикторов является довольно трудоемким, необходимо разработать программное обеспечение, которое при задании отношения сигнал/шум и выборе вида шума будет автоматически его накладывать на аудиозаписи речи дикторов. Предлагается использовать язык программирования Python для разработки такого программного обеспечения. Для обработки аудиофайлов в разрабатываемом программном обеспечении необходимо использовать соответствующие библиотеки, применяемые в данной области. Одной из важных библиотек для обработки звука является PyAudio [13]. Основная функциональность PyAudio заключается в записи аудиопотока в объекты типа “bytes”. В дальнейшем эти данные могут быть сохранены в формате wav с использованием библиотек, таких как SciPy или Wave.

Также существует библиотека Librosa [14], которая предназначена для сбора и воспроизведения аудио. Librosa является модулем на языке Python, который специализируется на анализе звуковых сигналов, обработке звука и голоса. Она предоставляет функции для воспроизведения аудиофайлов непосредственно в среде разработки Python, а также возможность визуализации аудиодорожки. В рамках данной библиотеки также имеются функции для построения спектрограммы – графического представления спектральных характеристик звукового сигнала.

В работе [15] рассматривается библиотека pyAudioAnalysis, которая предоставляет широкий спектр функциональных возможностей анализа звука с помощью простого в использовании комплексного программного дизайна. Данная библиотека может быть использована для извлечения аудиофайлов, обучения и применения аудиоклассификаторов, сегментации аудиопотока с использованием контролируемых или неконтролируемых методологий и визуализации аудиодорожки.

В исследовании [16] авторами представляется модуль PYO, предназначенный для цифровой обработки звука. Эта библиотека позволяет быстро интегрировать аудиопроцессы в

другие задачи программирования, такие как математические вычисления, сетевые коммуникации или программирование графических интерфейсов.

Следующим этапом разработки интеллектуальной системы является генерация спектрограмм или мел-частотных кепстральных коэффициентов на основе зашумленных аудиозаписей речи диктора. В настоящее время существуют различные подходы к анализу речевых сигналов как во временной, так и в частотной областях. В частности, широко применяются методы, основанные на создании спектрограмм и вычислении мел-частотных кепстральных коэффициентов. Спектрограмма представляет собой графическое отображение изменения спектральной плотности мощности сигнала в зависимости от времени. Для создания спектрограммы используется оконное преобразование Фурье, которое позволяет разбить сигнал на интервалы с определенными диапазонами частот и времени, характеризуемыми амплитудой сигнала на каждом интервале. Визуализация амплитуды на спектрограмме обычно осуществляется путем использования яркости или цвета (чаще всего более яркие области соответствуют большей амплитуде). Все последовательности разбиваются на одинаковые по частотному диапазону сегменты, а затем эти сегменты разделяются на временные отрезки, поскольку длительность сигналов может быть различной. Для обработки полученных фрагментов данных применяется быстрое преобразование Фурье [17]. На рис. 4 показана спектрограмма аудиозаписи речи диктора длиной 10 секунд без наложения шума.

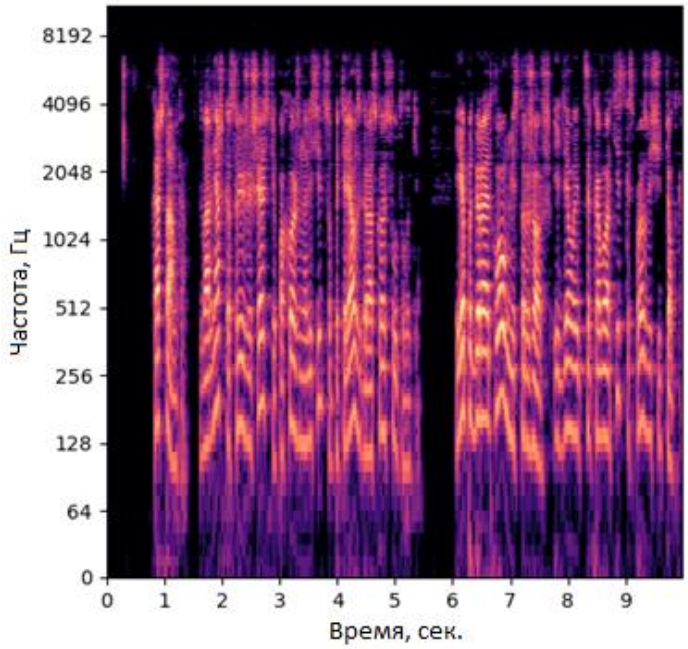

Рис. 4. Спектрограмма аудиозаписи речи диктора длиной 10 секунд без наложения шума

Мел-частотные кепстральные коэффициенты (MFCC) являются набором относительно небольших значений, которые описывают частотные характеристики сигнала. Этот метод основывается на равномерном распределении частотных полос по логарифмической шкале, соответствующей способности человека воспринимать звуки [18]. В результате преобразования получается величина, называемая мелом. Для вычисления MFCC используется оконное преобразование Фурье, а при обработке записи голоса, представленной одним словом, преобразование может производиться над всем образцом сигнала. Полученные спектральные значения проходят через треугольные оконные фильтры, после чего к полученным значениям в каждом окне применяется дискретное косинусное преобразование. Амплитуды результирующего спектра являются итоговыми мел-частотными кепстральными коэффициентами [19]. На рис. 5 показан пример мел-частотных кепстральных коэффициентов, полученных из аудиозаписи длиной 10 секунд без наложения шума.

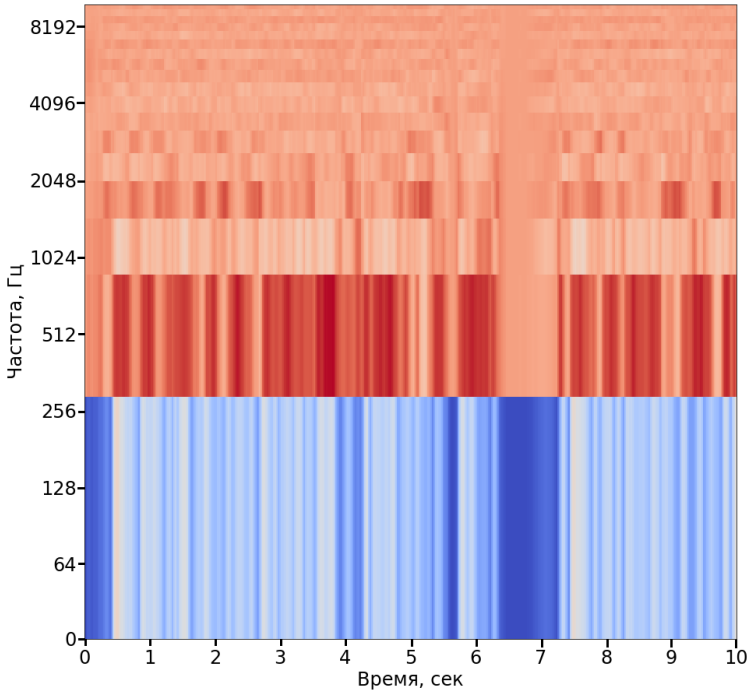

Рис. 5. Мел-частотные кепстральные коэффициенты аудиозаписи речи диктора длиной 10 секунд без наложения шума

Для того чтобы правильно сгенерировать спектрограмму или мел-частотные кепстральные коэффициенты, необходимо определить следующие параметры для зашумленной аудиозаписи речи:

– частота дискретизации;

– вид окна Фурье – обычно используются окна Хэмминга, Ханна или прямоугольные окна для создания спектрограммы или мел-частотных кепстральных коэффициентов;

– разрешение изображения – необходимо выбрать размер выходной спектрограммы или мел-частотных кепстральных коэффициентов в пикселях, учитывая баланс между размером данных и детализацией;

– количество признаков, которые необходимо выделить из аудиозаписи (при генерации мел-частотных кепстральных коэффициентов);

– временной отрезок, который необходимо обработать, – можно выбрать полную длительность аудиозаписи или определенный временной отрезок для обработки;

– отображаемые параметры по осям графика;

– вид представления изображения (линейный, логарифмический);

– возможно, для расширения набора данных провести «отзеркаливание» по горизонтали спектрограмм или мел-частотных кепстральных коэффициентов.

В языке программирования Python доступны библиотеки NumPy и Matplotlib, позволяющие работать с вышеперечисленными параметрами. NumPy представляет собой библиотеку, которая добавляет поддержку многомерных массивов и матриц, а также обширную коллекцию высокоуровневых (и очень быстрых) математических функций для работы с этими массивами [20]. Matplotlib, в свою очередь, является библиотекой, используемой для создания графиков и упрощения процесса их построения. Она часто используется совместно с библиотекой NumPy [21].

Для реализации инструмента определения соответствия значения разборчивости речи доле примеси шумового сигнала в речевом сигнале (а это основная задача в оценке защищенности) предлагается применить нейронную сеть, которая работает с изображениями.

В современном мире всё больше исследователей в области обработки звука используют машинное обучение в своих научных целях [22, 23, 24]. Особенно широкое распространение в данной области получил такой класс машинного обучения, как глубокое обучение. Всё ча-

ще исследователи используют архитектуры нейронных сетей, такие как генеративные состязательные сети (GAN), трансформеры и сверточные нейронные сети (CNN) [24, 25, 26]. Генеративные состязательные сети широко применяются в задачах, где необходимо сгенерировать новые данные, например, изображения, звуки, тексты. Они широко используются в области искусственного интеллекта для создания высококачественных и реалистичных данных [27]. В нашей задаче мы уже имеем данные и генерировать новые нет необходимости. Такой вид нейронных сетей, как трансформеры, используется для обработки последовательных данных – естественный язык, машинный перевод, задачи обработки текста [28]. В последние годы были разработаны модификации трансформеров, такие как Vision Transformer (ViT), которые успешно применяются в задачах обработки изображений [29]. Однако предобученный трансформер имеет фиксированное разрешение входного изображения, что может стать ограничением для ряда сценариев. Трансформеры изначально разрабатывались для работы с последовательными данными и не всегда эффективны при обработке пространственных зависимостей в изображениях. Это может привести к трудностям в анализе локальных структур и объектов на изображениях. Также стоит отметить, что при обработке больших изображений трансформерам может потребоваться разделение изображения на более мелкие фрагменты, что увеличивает вычислительную сложность и требует дополнительной обработки [30, 31]. Согласно последним исследованиям высокую эффективность в различных задачах распознавания имеют сверточные нейронные сети [32, 33]. Сверточные нейронные сети показали свою эффективность при использовании в задачах компьютерного зрения или распознавания изображений. Они эффективно улавливают локальные пространственные зависимости в данных, что является ключевым требованием в задачах компьютерного зрения. Сверточные нейронные сети, особенно при использовании предварительного обучения на больших наборах данных, могут быть вычислительно более эффективными и требовать меньше ресурсов – это важно при работе с большими объемами данных [26, 34, 35]. Кроме того, такие нейронные сети могут использоваться для распознавания речи, если в наборе данных использовать изображения спектрограмм или мел-частотных кепстральных коэффициентов, сгенерированных на основе зашумленных аудиозаписей. Сверточная нейронная сеть состоит из разных видов слоев: сверточные слои, субдискретизирующие слои (подвыборка) и выходной слой [36]. Пример архитектуры сверточной нейронной сети представлен на рис. 6, где обрабатывается спектрограмма аудиозаписи речи диктора с наложением белого шума с результатом распознавания нейронной сетью на выходе.

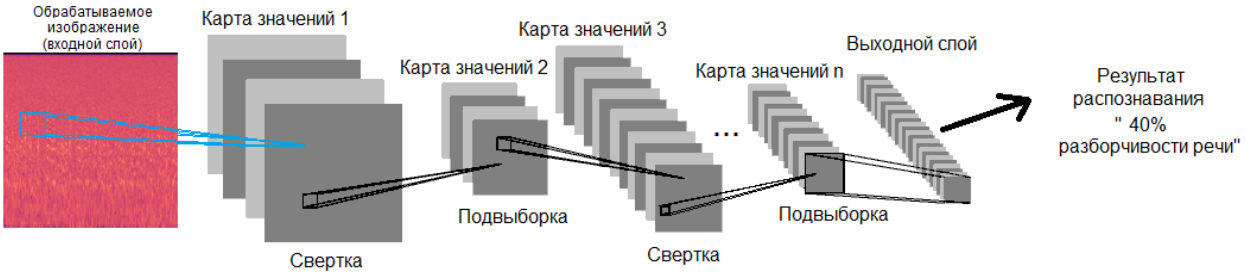

Рис. 6. Архитектура сверточной нейронной сети

Отметим, что при создании модели сверточной нейронной сети необходимо выделить её параметры и параметры обучения, а именно:

– количество слоев сверточной нейронной сети, включая сверточные слои, слои объединения и полносвязные слои;

– сколько необходимо полных прохождений обучения на наборе данных (эпох), которые позволят модели достичь сходимости. Можно использовать методы ранней остановки для предотвращения переобучения;

– вид, в котором будет представлена информация об успешном обучении, – какие метрики и графики будут использоваться для оценки успешности обучения модели, например, точность, потери, кривые обучения и валидации;

– вид, в котором подается на вход модели сверточной нейронной сети на распознавание тестовая информация – по одному изображению или пакетами.

Также необходимо разделить набор данных, представленный в виде спектрограмм или мел-частотных кепстральных коэффициентов, на тренировочный и тестовый. Такое разделение позволяет тестировать работу нейронной сети на ранее не использовавшихся данных и, соответственно, оценивать её точность распознавания. Чтобы не охватывать весь диапазон от 100 процентов разборчивости речи (аудиозапись речи без наложения шума) до 0 процентов (в аудиозаписи речи невозможно распознать ни одного слова), для начальной проверки работоспособности нейронной сети можно выделить несколько градаций процента разборчивости речи: например, от 20 до 30 процентов (каждой градации будет соответствовать 1 процент разборчивости речи). В соответствии с подходами для проведения артикуляционных испытаний [37] необходимо дать на прослушивание аудиторам зашумленную аудиозапись речи диктора – для нашего испытания ограничимся двумя мужчинами и двумя женщинами. Прослушивая зашумленную аудиозапись, аудитор оценивает, сколько слов он смог распознать. Зная заведомо количество слов в прослушиваемой аудиозаписи речи и сравнивая его с результатом распознавания аудитором количества слов, можно рассчитать процент разборчивости речи. Если взять четыре результата распознавания слов аудиторами, то получим среднее значение процента разборчивости речи. Полагаясь на это значение, необходимо провести градацию зашумленных аудиозаписей в наборе данных. В таком случае результатом работы предложенной реализации нейронной сети будет процент разборчивости речи, определенный в указанном диапазоне. Количество заданных градаций будет соответствовать итоговому количеству классов распознавания нейронной сети.

В дальнейшем для проверки предложенного подхода предлагается сравнить результат инструментально-расчетного подхода оценки разборчивости речи, разработанного А. А. Хоревым, В. К. Желязняком и Ю. К.Макаровым, с результатами работы интеллектуальной системы оценки защищенности речевой акустической информации на основе нейронной сети.

3. Заключение

В результате работы предложен подход к оценке защищенности речевой акустической информации с применением нейронных сетей. Для дальнейшей реализации данного подхода было сделано следующее:

1) предложена структурная схема этапов создания интеллектуальной системы оценки защищенности речевой акустической информации на основе нейронной сети;

2) сформулированы рекомендации по формированию наборов данных для обучения сверточной нейронной сети;

3) произведен выбор оборудования для записи речи диктора, программного обеспечения для работы с аудиофайлами и среды разработки программного обеспечения для реализации интеллектуальной системы оценки защищенности речевой акустической информации на основе нейронной сети;

4) предложены варианты представления данных для обучения сверточной нейронной сети.

Однако для получения оптимальных результатов необходимо обеспечить обучение и тестирование интеллектуальной системы на разнообразных данных (произвольные спектры помех и речи), чтобы она могла обрабатывать различные условия и шумовые сценарии. Также необходим сравнительный анализ результатов, полученных с помощью разработанной системы, и результатов, полученных с использованием традиционного инструментально-расчетного подхода. Результаты данного исследования могут быть использованы для совершенствования процесса оценки защищенности речевой акустической информации.

Литература

1. *Сагдеев К. М., Петренко В. И.* Методика оценки технической защищенности речевой информации в выделенных помещениях // Известия ЮФУ. Технические науки. 2012. № 12 (137). С. 121–129.
2. *Железняк В. К., Макаров Ю. К., Хорев А. А.* Некоторые методические подходы к оценке эффективности защиты речевой информации // Специальная техника. 2000. № 4. С. 39–45.
3. *Покровский Н. Б.* Расчет и измерение разборчивости речи. М.: Связьиздат, 1962. 392 с.
4. *Трушин В. А., Рева И. Л., Иванов А. В.* Экспериментальная оценка разборчивости речи в задачах защиты информации на основе модифицированных артикуляционных измерений // Материалы X-й Междунар. конф. «Актуальные проблемы электронного приборостроения», НГТУ, Новосибирск, 2010. Т. 3. С. 133–136.
5. *Иванов А. В., Рева И. Л., Трушин В. А., Тудэвдагва У.* Корректировка методики оценки защищенности речевой информации от утечки по техническим каналам в условиях форсирования речи // Научный вестник Новосибирского государственного технического университета. 2014. № 2 (55). С. 183–189.
6. *Макаров Ю. К., Хорев А. А.* К оценке эффективности защиты акустической (речевой) информации // Специальная техника. 2000. № 5. С. 46–56.
7. *Иванов А. В., Салимов Ш. Р.* О возможности применения технологий распознавания речи в задачах оценки защищенности акустической информации от утечки по техническим каналам // Динамика систем, механизмов и машин. 2020. Т. 8, № 2. С. 109–114.
8. *Жабыко Е. И., Рублевская Н. И.* Акустическое проектирование залов многоцелевого назначения: учебное пособие. Владивосток: Издательство ДВГТУ, 2008. 89 с.
9. *Хорев А. А., Поров И. С.* Методика вероятностной оценки разборчивости // Защита информации. Инсайд. 2020. № 2 (92). С. 44–52.
10. *Трушин В. А., Заводовская А. И., Овешников И. А., Топорищев Э. В.* Исследование воздействия речеподобной помехи на психоэмоциональное состояние человека // Динамика систем, механизмов и машин. 2020. Т. 8, № 2. С. 138–144.
11. *Иванов А. В., Рева И. Л., Шемшетдинова Э. Э.* Исследование влияния различий в спектрах речи на результат оценки разборчивости // Динамика систем, механизмов и машин. 2017. Т. 5, № 4. С. 65–70.
12. Adobe Audition. Профессиональная студия звукозаписи [Электронный ресурс]. URL: `https://www.adobe.com/ru/products/audition.html` (дата обращения: 10.09.2023).
13. PyAudio 0.2.13 – Python Package Index [Электронный ресурс]. URL: `https://pypi.org/project/PyAudio/` (дата обращения: 10.09.2023).
14. Librosa 0.10 – Librosa - audio and music processing in Python [Электронный ресурс]. URL: `https://librosa.org/doc/latest/index.html` (дата обращения: 10.09.2023).
15. *Giannakopoulos T.* PyAudioAnalysis: An open-source python library for audio signal analysis // PLoS ONE. 2015. № 10 (12). P. 1–17.
16. *Bélanger O.* Pyo, the python DSP toolbox // Proc. ACM Multimedia Conference, 2016. P. 1214–1217.
17. *Умняшкин С. В.* Основы теории цифровой обработки сигналов: учебное пособие. М.: Техносфера, 2016. 528 с.
18. *Tyagi V., Wellekens C.* On desensitizing the Mel-cepstrum to spurious spectral components for robust speech recognition // Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. P. 1–21.
19. *Герасимов С. М., Жаринов О. О.* Исследование методов анализа речевых сигналов // Сборник докладов 73-й Международной студенческой научной конференции ГУАП, 2020. Ч. 3. С. 36–41.

20. NumPy documentation [Электронный ресурс]. URL: https://numpy.org/doc/stable/ (дата обращения: 10.09.2023).
21. Matplotlib 3.7.2 documentation [Электронный ресурс]. URL: https://matplotlib.org/stable/tutorials/index.html (дата обращения: 10.09.2023).
22. *Li J., Deng L., Haeb-Umbach R., Gong Y.* Robust automatic speech recognition. A bridge to practical applications, 2016. 286 p.
23. *Fang Z., Yin B., Du Z. et al.* Fast environmental sound classification based on resource adaptive convolutional neural network // Scientific Reports. 2022. № 12. P. 1–18.
24. *Madhu A., Kumaraswamy S.* EnvGAN: a GAN-based augmentation to improve environmental sound classification // Artificial Intelligence Review. 2022. № 55. P. 6301–6320.
25. *Kim B., Kim J., Ye J. C.* Task-Agnostic Vision Transformer for Distributed Learning of Image Processing // Transactions on Image Processing. 2023. № 32. P. 203–218.
26. *Ullah R., Asif M., Shah W. A., Anjam F., Ullah I., Khurshaid T., Wuttisittikulkij L., Shah S., Ali S. M., Alibakhshikenari M.* Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer // Sensors. 2023. № 13. P. 1–20.
27. *Porkodi S. P., Sarada V., Maik V. et al.* Generic image application using GANs (Generative Adversarial Networks): A Review // Evolving Systems. 2023. № 14. P. 903–917.
28. *Song Q., Sun B., Li S.* Multimodal Sparse Transformer Network for Audio-Visual Speech Recognition // IEEE Transactions on Neural Networks and Learning Systems. 2023. V. 34, № 12. P. 10028–10038.
29. Vision Transformer: What It Is & How It Works (2023 Guide) [Электронный ресурс]. URL: https://www.v7labs.com/blog/vision-transformer-guide/ (дата обращения: 13.12.2023).
30. *Tay Y., Dehghani M., Gupta J., Bahri D., Aribandi V., Qin Z., Metzler D.* Are Pre-trained Convolutions Better than Pre-trained Transformers? 2022. arXiv: 2105.03322 [cs.CL].
31. *Sanford C., Hsu D., Telgarsky M.* Representational Strengths and Limitations of Transformers. 2023. arXiv:2306.02896 [cs.LG].
32. *Abdel-Hamid O., Mohamed A.-R., Jiang H., Penn G.* Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition // Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2012. P. 4277–4280.
33. *Сикорский О. С.* Обзор свёрточных нейронных сетей для задачи классификации изображений // Новые информационные технологии в автоматизированных системах. 2017. № 20. С. 1–8.
34. *Ciretan D. C., Giusti A., Gambardella L. M., Schmidhuber J.* Deep neural networks segment neuronal membranes in electron microscopy images // Proc. NIPS. 2012. P. 1–9.
35. *Ciretan D. C., Meier U., Gambardella L. M., Schmidhuber J.* Deep, Big, Simple Neural Nets for Handwritten Digit Recognition // MIT Press. 2010. V. 22, № 12. P. 3207–3220.
36. Что такое свёрточная нейронная сеть – Хабр [Электронный ресурс]. URL: https://habr.com/ru/articles/309508/ (дата обращения: 10.09.2023).
37. *Трушин В. А.* Информационно-измерительная модель формантного метода определения разборчивости речи // Труды Научно-исследовательского института радио. 2017. № 4. С. 2–9.

Волков Никита Андреевич
аспирант кафедры «Электронные системы и информационная безопасность», Самарский государственный технический университет (СамГТУ, 443011, Самара, ул. Молодогвардейская, д. 244), e-mail: volkovnikandr@gmail.com, ORCID ID: 0000-0002-0644-6332.

Иванов Андрей Валерьевич
к.т.н., доцент кафедры «Электронные системы и информационная безопасность», Самарский государственный технический университет (СамГТУ, 443011, Самара, ул. Молодогвардейская, д. 244), тел. +7 846 337-31-96, e-mail: andrej.ivanov@corp.nstu.ru, ORCID ID: 0000-0003-2002-8572.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.
Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

An Approach to Assessing the Security of Speech Acoustic Information Using Neural Networks

Nikita A. Volkov, Andrey V. Ivanov

Samara State Technical University (SamSTU)

Abstract: The paper is devoted to the consideration of the methodology for assessing the security of speech acoustic information in the preparation of premises for private negotiations. Taking into account the disadvantages of existing approaches it is proposed to apply recognition methods based on convolutional neural networks. The paper proposes a block diagram of the stages for creating an intelligent system. The process of creating a training dataset in audio recording format with superimposed noises with different signal-to-noise ratios is described. The possibilities of the Adobe Audition audio editor and Python libraries for generating datasets are considered. It is proposed to classify spectrograms or mel-frequency cepstral coefficients of audio recordings using a neural network by the percentage of speech intelligibility in order to automate the process of assessing the security of speech acoustic information. To achieve the desired result, it is planned to train a neural network on various data, conduct a comparative analysis with the existing approach, evaluate the performance of the system and validate the results. The proposed approach and its practical application will significantly improve the quality and expand the conditions for the application of the security assessment of speech acoustic information.

Keywords: deep neural networks, convolutional neural networks, signal-to-noise ratio, audio recording noise, speech recognition, spectrograms, mel-frequency cepstral coefficients, assessment of the security of speech acoustic information.

For citation: Volkov N. A., Ivanov A. V. An approach to assessing the security of speech acoustic information using neural net-works (in Russian). *Vestnik SibGUTI*, 2024, vol. 18, no. 2, pp. 43-56. https://doi.org/10.55648/1998-6920-2024-18-2-43-56.

The article was submitted: 13.10.2023;
revised version: 20.12.2023;
accepted for publication 22.12.2023.

References

1. Sagdeev K. M., Petrenko V. I. Metodika otsenki tekhnicheskoi zashchishchennosti rechevoi informatsii v vydelennykh pomeshcheniyakh [Methodology for assessing the technical security of speech information in dedicated rooms]. *Izvestiya YuFU. Tekhnicheskie nauki*, 2012, no. 12 (137), pp. 121-129.
2. Zheleznyak V. K., Makarov Yu.K., Khorev A. A. Nekotorye metodicheskie podkhody k otsenke effektivnosti zashchity rechevoi informatsii [Some methodological approaches to evaluating the effectiveness of speech information protection]. *Spetsial'naya tekhnika*, 2000, no. 4, pp. 39-45.
3. Pokrovskii N. B. *Raschet i izmerenie razborchivosti rechi* [Calculation and measurement of speech intelligibility]. Moscow, Svyaz'izdat, 1962. 392 p.
4. Trushin V. A., Reva I. L., Ivanov A. V. Eksperimental'naya otsenka razborchivosti rechi v zadachakh zashchity informatsii na osnove modifitsirovannykh artikulyatsionnykh izmerenii [Experimental assessment of speech intelligibility in information security tasks based on modified articulation measurements]. *Aktual'nye problemy elektronnogo priborostroeniya: materialy X Mezhdunar. konf. (APEP 2010)*, Novosibirsk: NGTU, 2010, vol. 3, pp. 133–136.
5. Ivanov A. V., Reva I. L., Trushin V. A., Tudevdagva U. Korrektirovka metodiki otsenki zashchishchennosti rechevoi informatsii ot utechki po tekhnicheskim kanalam v usloviyakh forsirovaniya rechi [Correction of the methodology for assessing the security of speech information from leakage through technical channels in conditions of speech forcing]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta*, 2014, no. 2(55), pp. 183-189.
6. Makarov Yu. K., Khorev A. A. K otsenke effektivnosti zashchity akusticheskoi (rechevoi) informatsii [To assess the effectiveness of the protection of acoustic (speech) information]. *Spetsial'naya tekhnika*, 2000, no. 5, pp. 46–56.
7. Ivanov A. V., Salimov Sh. R. O vozmozhnosti primeneniya tekhnologii raspoznavaniya rechi v zadachakh otsenki zashchishchennosti akusticheskoi informatsii ot utechki po tekhnicheskim kanalam [On the possibility of using speech recognition technology in the tasks of assessing the security of acoustic information from leakage through technical channels]. *Dinamika sistem, mekhanizmov i mashin*, 2020, vol. 8, no 2, pp. 109–114.
8. Zhabyko E. I., Rublevskaya N. I. *Akusticheskoe proektirovanie zalov mnogotselevogo naznacheniya* [Acoustic design of multi-purpose halls]. Vladivostok, Izdatel'stvo DVGTU, 2008, 89 p.
9. Khorev A. A., Porev I. S. Metodika veroyatnostnoi otsenki razborchivosti [The method of probabilistic assessment of intelligibility]. *Zashchita informatsii. Insaid*, 2020, no. 2(92), pp. 44-52.
10. Trushin V. A., Zavodovskaya A. I., Oveshnikov I. A., Toporishchev E. V. Issledovanie vozdeistviya rechepodobnoi pomekhi na psikhoemotsional'noe sostoyanie cheloveka [Investigation of the impact of speech-like interference on the psychoemotional state of a person]. *Dinamika sistem, mekhanizmov i mashin*, 2020, vol. 8, no. 2, pp. 138-144.
11. *Ivanov A. V., Reva I. L., Shemshetdinova E. E.* Issledovanie vliyaniya razlichii v spektrakh rechi na rezul'tat otsenki razborchivosti [Investigation of the effect of differences in speech spectra on the result of the intelligibility assessment]. *Dinamika sistem, mekhanizmov i mashin*, 2017, vol. 5, no. 4, pp. 65–70.
12. *Adobe Audition. Professional'naya studiya zvukozapisi* [Adobe Audition. Professional recording studio], available at: `https://www.adobe.com/ru/products/audition.html` (accessed: 10.09.2023).
13. *PyAudio 0.2.13 – Python Package Index*, available at: `https://pypi.org/project/PyAudio/` (accessed: 10.09.2023).
14. *Librosa 0.10 – Librosa – audio and music processing in Python*, available at: `https://librosa.org/doc/latest/index.html` (accessed: 10.09.2023).
15. Giannakopoulos T. PyAudioAnalysis: An open-source python library for audio signal analysis. *PLoS ONE*, 2015, no. 10(12), pp. 1-17.
16. Bélanger O. Pyo, the python DSP toolbox. *MM 2016 - Proceedings of the 2016 ACM Multimedia Conference*, 2016, pp. 1214-1217.
17. Umnyashkin S. V. *Osnovy teorii tsifrovoi obrabotki signalov* [Fundamentals of the theory of digital signal processing]. Moscow, Tekhnosfera, 2016. 528 p.
18. Tyagi V., Wellekens C. On desensitizing the Mel-cepstrum to spurious spectral components for robust speech recognition. *Proceedings (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 1-21.
19. Gerasimov S. M., Zharinov O. O. Issledovanie metodov analiza rechevykh signalov [Research of methods of speech signal analysis]. *Sbornik dokladov sem'desyat tret'ei mezhdunarodnoi studencheskoi nauchnoi konferentsii GUAP*, 2020, vol. 3, pp. 36-41.

20. *NumPy documentation*, available at: https://numpy.org/doc/stable/ (accessed: 10.09.2023).
21. *Matplotlib 3.7.2 documentation*, available at: https://matplotlib.org/stable/tutorials/index.html (accessed: 10.09.2023).
22. Li J., Deng L., Haeb-Umbach R., Gong Y. *Robust automatic speech recognition. A bridge to practical applications*, 2016, 286 p.
23. Fang Z., Yin B., Du Z. et al. Fast environmental sound classification based on resource adaptive convolutional neural network. *Scientific Reports*, 2022, no. 12, pp. 1-18.
24. Madhu A., Kumaraswamy S. EnvGAN: a GAN-based augmentation to improve environmental sound classification. *Artificial Intelligence Review*, 2022, no. 55, pp. 6301-6320.
25. Kim B., Kim J., Ye J. C. Task-Agnostic Vision Transformer for Distributed Learning of Image Processing. *Transactions on Image Processing*, 2023, no. 32, pp. 203-218.
26. Ullah R., Asif M., Shah W. A., Anjam F., Ullah I., Khurshaid T., Wuttisittikulkij L., Shah S., Ali S. M., Alibakhshikenari M. Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. *Sensors 23*, 2023, no. 13, pp. 1-20.
27. Porkodi S. P., Sarada V., Maik V. et al. Generic image application using GANs (Generative Adversarial Networks): A Review. *Evolving Systems 14*, 2023, pp. 903-917.
28. Song Q., Sun B., Li S. Multimodal Sparse Transformer Network for Audio-Visual Speech Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, vol. 34, no. 12, pp. 10028-10038.
29. *Vision Transformer: What It Is & How It Works (2023 Guide)*, available at: https://www.v7labs.com/blog/vision-transformer-guide/ (accessed: 13.12.2023).
30. Tay Y., Dehghani M., Gupta J., Bahri D., Aribandi V., Qin Z., Metzler D. Are Pre-trained Convolutions Better than Pre-trained Transformers? 2022, *arXiv: 2105.03322 [cs.CL]*.
31. Sanford C., Hsu D., Telgarsky M. Representational Strengths and Limitations of Transformers. 2023, *arXiv:2306.02896 [cs.LG]*.
32. Abdel-Hamid O., Mohamed A.-R., Jiang H., Penn G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2012, pp. 4277-4280.
33. Sikorskii O. S. Obzor svertochnykh neironnykh setei dlya zadachi klassifikatsii izobra-zhenii [Overview of convolutional neural networks for image classification problem], *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2017, no. 20, pp. 1-8.
34. Ciretan D. C., Giusti A., Gambardella L. M., Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. *Proc. NIPS*, 2012, pp. 1-9.
35. Ciretan D. C., Meier U., Gambardella L. M., Schmidhuber J. Deep, Big, Simple Neural Nets for Handwritten Digit Recognition, *MIT Press*, 2010, vol. 22, no. 12, pp. 3207-3220.
36. Chto takoe svertochnaya neironnaya set' – Khabr [What is a convolutional neural network – Habr], available at: https://habr.com/ru/articles/309508/ (accessed: 10.09.2023).
37. Trushin V. A. Informatsionno-izmeritel'naya model' formantnogo metoda opredeleniya razborchivosti rechi [Information and measurement model of the formant method for determining speech intelligibility]. *Trudy Nauchno-issledovatel'skogo instituta radio*, 2017, no. 4, pp. 2-9.

Nikita A. Volkov

Postgraduate student of the Department of Electronic Systems and Information Security, Samara State Technical University (SamSTU, Russia, 443011, Samara, Molodogvordeyskaya St. 244), phone: +7 846 337-31-96, e-mail: volkovnikandr@gmail.com, ORCID ID: 0000-0002-0644-6332.

Andrey V. Ivanov

Cand. of Sci. (Engineering), Associate Professor of the Department of Electronic Systems and Information Security, Samara State Technical University (SamSTU, Russia, 443011, Samara, Molodogvordeyskaya St. 244), phone: +7 846 337-31-96, e-mail: andrej.ivanov@corp.nstu.ru, ORCID ID: 0000-0003-2002-8572.