УДК 004.93 **DOI** 10.15622/ia.20.5.5

А.А. Двойникова, М.В. Маркитантов, Е.В. Рюмина, Д.А. Рюмин, А.А. Карпов

АНАЛИТИЧЕСКИЙ ОБЗОР АУДИОВИЗУАЛЬНЫХ СИСТЕМ ДЛЯ ОПРЕДЕЛЕНИЯ СРЕДСТВ ИНДИВИДУАЛЬНОЙ ЗАЩИТЫ НА ЛИЦЕ ЧЕЛОВЕКА

Двойникова А.А., Маркитантов М.В., Рюмина Е.В., Рюмин Д.А., Карпов А.А. **Аналитический обзор аудиовизуальных систем для определения средств индивидуальной защиты на лице человека.**

Аннотация. Начиная с 2019 года все страны мира столкнулись со стремительным распространением пандемии, вызванной коронавирусной инфекцией COVID-19, борьба с которой продолжается мировым сообществом и по настоящее время. Несмотря на очевидную эффективность средств индивидуальной защиты органов дыхания от заражения коронавирусной инфекцией, многие люди пренебрегают использованием защитных масок для лица в общественных местах. Поэтому для контроля и своевременного выявления нарушителей общественных правил здравоохранения необходимо применять современные информационные технологии, которые будут детектировать защитные маски на лицах людей по видео- и аудиоинформации. В статье приведен аналитический обзор существующих и разрабатываемых интеллектуальных информационных технологий бимодального анализа голосовых и лицевых характеристик человека в маске. Существует много исследований на тему обнаружения масок по видеоизображениям, также в открытом доступе можно найти значительное количество корпусов, содержащих изображения лиц как без масок, так и в масках, полученных различными способами. Исследований и разработок, направленных на детектирование средств индивидуальной защиты органов дыхания по акустическим характеристикам речи человека пока достаточно мало, так как это направление начало развиваться только в период пандемии, вызванной коронавирусной инфекцией COVID-19. Существующие системы позволяют предотвратить распространение коронавирусной инфекции с помощью распознавания наличия/отсутствия масок на лице, также данные системы помогают в дистанционном диагностировании COVID-19 с помощью обнаружения первых симптомов вирусной инфекции по акустическим характеристикам. Однако, на сегодняшний день существует ряд нерешенных проблем в области автоматического диагностирования симптомов COVID-19 и наличия/отсутствия масок на лицах людей. В первую очередь это низкая точность обнаружения масок и коронавирусной инфекции, что не позволяет осуществлять автоматическую диагностику без присутствия экспертов (медицинского персонала). Многие системы не способны работать в режиме реального времени, из-за чего невозможно производить контроль и мониторинг ношения защитных масок в общественных местах. Также большинство существующих систем невозможно встроить в смартфон, чтобы пользователи могли в любом месте произвести диагностирование наличия коронавирусной инфекции. Еще одной основной проблемой является сбор данных пациентов, зараженных COVID-19, так как многие люди не согласны распространять конфиденциальную информацию.

Ключевые слова: определение защитных масок, голосовые характеристики дикторов, COVID-19, средства индивидуальной защиты, обнаружение кашля, лицевые характеристики.

1. Введение. С конца 2019 года все страны столкнулись со стремительным распространением пандемии, вызванной коронавирусной инфекцией COVID-19, борьба с которой продолжается мировым сообществом и по настоящее время. Аномально высокий уровень заражения COVID-19 связан с возможностью его передачи не только воздушно-капельным путем (при кашле, чихании и дыхании инфицированным человеком), но и контактным путем через дверные ручки, поручни в транспорте и другие загрязненные предметы и поверхности. Именно поэтому многие ведущие научные группы и мировые промышленные корпорации из различных научных областей, таких как медицина, биология, информатика и другие, сосредоточились на решении глобальной проблемы уменьшения числа зараженных людей по всему миру [1] и активно ведут исследования и разработки интеллектуальных технологий, которые позволят создать эффективные решения по предотвращению распространения коронавирусной инфекции COVID-19.

Во многих исследованиях, например, [2–4], описывается эффективность использования средств индивидуальной защиты (СИЗ) органов дыхания для предотвращения респираторных заболеваний. Ношение медицинских масок в общественных местах необходимо для обеих категорий людей: для инфицированных, чтобы уменьшить выделения возбудителя инфекции из респираторного тракта и предотвратить дальнейшее распространение болезни, а также для здоровых людей, чтобы избежать попадания вредоносных вирусов и бактерий в организм и минимизировать возможность заражения такими заболеваниями, как туберкулез [5], грипп, острое респираторное заболевание (ОРЗ) или острая респираторная вирусная инфекция (ОРВИ) [6]. Кроме этого, к респираторным заболеваниям относится и коронавирусная инфекция COVID-19 [7], которая легко передается от человека к человеку в воздушной среде, и поэтому для предотвращения пандемии [8] и снижения смертности [9] необходимо обеспечить соблюдение масочного режима в общественных местах. Так, в отечественной работе [10], которая посвящена исследованию по выявлению корреляции между числом заболеваний COVID-19 и плотностью населения, устанавливается линейная зависимость между количеством зараженных людей и численностью определенного субъекта Российской Федерации. Главным способом предотвращения распространения пандемии в странах с высокой плотностью населения является коллективное ношение средств индивидуальной защиты и регулярная их смена [11]. В качестве средств защиты органов дыхания используются маски различных видов: медицинские, тканевые, респираторы и другие. В зависимости

от вида маски уровень их эффективности различается, так, например, группа австралийских ученых в своей работе [12] доказывает, что вероятность попадания различных инфекций в организм человека ниже при ношении специализированных медицинских масок и респираторов относительно тканевых вариаций защитных масок. В другом проведенном исследовании [9] авторы утверждают, что даже малоэффективные маски (уровень задержания вирусов которых не более 20%) могут быть полезны как для профилактики заболеваний, связанных в той или иной степени с COVID-19, у здоровых людей, так и для предотвращения бессимптомной передачи данной коронавирусной инфекции.

Несмотря на эффективность средств индивидуальной защиты органов дыхания от заражения вирусными инфекциями, многие люди пренебрегают использованием защитных масок в общественных местах. На эту тему в 2020 году в Индии проводилось исследование [13], в котором опрашивали людей, пребывавших на домашнем карантине в связи с обнаружением у них респираторного заболевания. Результаты опроса показали, что треть участников (34,1%) не готовы носить маски в общественных местах даже при наличии строгих законов об использовании СИЗ. В свою очередь в работе [14] был произведен анализ и предоставлена статистика ношения масок среди пешеходов на разных улицах городов Ирана, методом наблюдения было установлено, что 45,6% из 10440 людей не используют защитные маски на улицах, а в вечернее время число таких людей увеличивается до 56,1%. В июне 2020 года в штате Висконсин (США) проводилось статистическое наблюдение за покупателями в продуктовых магазинах [15], результаты которого показали, что только 41,5% из 5517 людей правильно носили маски на лицах, остальные 58,5% пренебрегали ношением масок в общественном месте. Во Франции [16] проводили опрос 1012 респондентов, направленный на извлечение мнений людей об обязательных мерах в общественных местах, в том числе ношении масок, 58% опрошенных выступили против наличия в их стране ограничительных мер. Группа исследователей из Института этнологии и антропологии Российской академии наук [17] в первые недели режима самоизоляции провела опрос среди 239 россиян, направленный на получение мнений людей о том, какое процентное соотношение людей носят маски в общественных местах, и результаты опроса показали, что всего 6,3% респондентов считают, что каждый человек надевает защитную маску, 26,8% думают, что защитную маску носит каждый второй, однако большинство респондентов 61,1% говорят, что маски используют только 1-2 человека из толпы, а мнение о том, что маски не носятся

вовсе, разделили между собой 5,9% опрошенных. Во время соблюдения карантинных требований населением для снижения распространения COVID-19 были неизбежны контакты между большим количеством людей, и риск их заражения был достаточно высок, поэтому наличие на лице средств индивидуальной защиты жизненно необходимо. Однако, как показывает статистика [18], только 54,68% водителей такси в Эфиопии использовали маски во время работы. Противоположную статистику можно заметить в медицинских учреждениях, где использование защитных масок является обязательным правилом нахождения на территории помещения. Так, ученые из Малайзии [19] оценили степень распространенности медицинских масок на лицах посетителей районной специализированной больницы во время пандемической вспышки COVID-19. Результаты показали, что среди 1625 человек, включенных в окончательный анализ, почти 97% людей были зарегистрированы в каких-либо масках для лица, из которых 72% носили именно медицинскую маску.

Использование защитных масок помогает уменьшить распространение коронавирусной инфекции COVID-19, однако большинство людей, находясь в общественных местах, пренебрегают средствами индивидуальной защиты. Поэтому для контроля и своевременного выявления нарушителей общественных правил здравоохранения необходимо разрабатывать и применять новые информационные технологии, которые будут детектировать маски на лице человека по видео- и аудиоинформации.

В последние годы в связи с распространением эпидемии, вызванной коронавирусной инфекцией COVID-19, ученые-информатики начали активно заниматься разработкой автоматических систем по детектированию масок на лице человека. Такие системы помогают обеспечить контроль соблюдения людьми законов об обязательном ношении средств индивидуальной защиты органов дыхания в общественных местах. Определить наличие/отсутствие маски на лице человека можно посредством:

1. Видеоинформации (например, видеонаблюдение в общественных местах, камеры смартфонов).

2. Акустической информации (например, телефонные разговоры).

Целью статьи является проведение сравнительного анализа подходов к автоматическому распознаванию наличия/отсутствия масок на лице человека по видео- и аудиоинформации, подходов к распознаванию респираторных заболеваний, в том числе COVID-19 по речи и звукам человека, а также корпусов изображений/видео и аудиоинфор-

мации для решения задач автоматического определения средств индивидуальной защиты. Предложенный сравнительный анализ помогает выделить достоинства и недостатки рассмотренных систем, выявить решенные и нерешенные проблемы в области мониторинга ношения масок в общественных местах, а также выдвинуть универсальные требования к разрабатываемым автоматическим системам определения наличия/отсутствия масок на лице человека.

2. Системы анализа видеоинформации. Одним из самых эффективных способов построения автоматических систем распознавания наличия/отсутствия масок на лицах людей является анализ видеоинформации. Преимуществом таких систем является возможность дистанционного контроля соблюдения масочного режима людей в общественных местах. С помощью видеоинформации можно анализировать ношение масок как конкретных людей (например, сотрудников офиса, клиентов в банке), так и множества людей в целом (например, для сбора статистики ношения масок в общественных местах). Для построения систем детектирования масок на лице человека необходимы корпусы с изображениями людей как в масках, так и без них.

2.1. Корпусы изображений/видео для распознавания наличия/отсутствия масок на лице человека. В последнее время в связи со стремлением многих ученых помочь снизить скорость распространения инфекции COVID-19 было создано несколько корпусов, содержащих в себе изображения людей в масках, для построения систем детектирования наличия/отсутствия СИЗ на лицах людей в общественных местах. Однако еще до возникновения и глобального распространения пандемии коронавирусной инфекции COVID-19 группа китайских ученых [20] собрала и аннотировала корпус MAFA (MAsked FAces), который включает изображения лиц с окклюзиями различными предметами, минимальный размер изображений составляет 80 пикселей. Окклюзия лица – перекрытие лица различного характера и степени сложности. Каждое изображение лица аннотировано по трем элементам: тип маски (простая, сложная, перекрытие лица различной областью тела человека, гибридная – комбинация различных типов маски) и степень окклюзий (низкая, средняя, высокая), а также ориентация лица в кадре (слева, слева фронтально, фронтально, справа фронтально, справа).

Medical Masks Dataset (MMD) включает в себя изображения людей в общественных местах. На каждом снимке изображено несколько людей с масками на лице. Face Mask Dataset (FMD) содержит в себе изображения как одного человека, так и нескольких. Разметка изображений производится на 3 класса: «лицо в маске», «лицо без маски»,

«лицо с неправильно надетой маской». Авторы работы [21] собрали корпус изображений людей из регионов Китая, России и Италии под названием MOXA3k. Разметка производилась на 2 класса: «лицо в маске» и «лицо без маски».

Корпус Medical Mask detection содержит изображения лиц людей разной национальности, возрастов и регионов. Изображения имеют разметку 3 классов: «лицо в маске», «лицо без маски» и «лицо с неправильно надетой маской», помимо этого в корпусе также присутствует разметка по 20 объектам, например, шлем, хиджаб, противогаз и другие. Другой корпус – Face mask classification (FMC) включает изображения лиц людей, находящихся на зашумленном фоне. Половина изображений всего объема FMC – это «лица в маске», а другая половина – «лица без маски». Авторы работы [22] собрали корпус Face Mask Detection Video Dataset изображений/видео, содержащих лица людей в масках и без них.

В период пандемии сбор данных, содержащих изображения лиц людей в масках, является затруднительным процессом, так как большинство стран ввели ограничительные меры на свободное передвижение людей, поэтому многие люди соблюдали режим строгой самоизоляции. По этой причине некоторые исследователи прибегали к получению данных изображений лиц в масках синтетическим путем с помощью компьютерного наложения изображения маски на лица людей. Так, авторы работы [23] для построения системы детектирования защитных масок разработали 3 корпуса, которые объединили в один под названием Masked Face Recognition Dataset (RMFD). Он включает в себя корпусы: Masked Face Detection Dataset (MFDD), содержащий в себе 24771 взятых из Интернета изображений лиц в маске, Real-world Masked Face Recognition Dataset (RMFRD) – содержащий 5000 фотографий 525 лиц в масках и 90000 изображений тех же субъектов без маски и Simulated Masked Face Recognition Dataset (SMFRD) – содержащий 500 тыс. изображений лиц, на которые наложили смоделированные медицинские маски.

Simulated Masked Face Dataset (SMFD) содержит в себе фотографии лиц без масок, на которые авторы корпуса наложили изображение обычной медицинской маски. Face Mask ~12K Images Dataset содержит в себе изображения, обрезанные в области лица людей на различных фонах. Корпус Face Mask detection представляет собой изображения лиц в масках и без них, причем маска может быть как реальной, так и наложенной синтетическим путем. Авторы работы [24] использовали корпус CASIA-Webface [25], который содержит фото-

графии лиц знаменитостей, взятых с сайта IMDb[1], для симулирования различных окклюзий на лице. Различные предметы, такие как 6 видов очков и 3 вида маски, в том числе медицинская, с 6 различными текстурами, и их комбинации накладываются на лица людей с учетом ключевых точек изображений лица. Авторы работы [26] создали синтетический корпус путем наложения изображения медицинской маски на лица, используя корпус Labeled Faces in the Wild (LFW) [27], который содержит в себе изображения лиц известных личностей.

На рисунке 1 показаны примеры изображений лиц в реальных и синтетических масках, а также без них из описанных выше корпусов.

Рис. 1. Примеры изображений лиц из различных корпусов

Подробное описание всех упомянутых выше корпусов, а также ссылки для скачивания представлены в таблице 1, из которой видно, что существует достаточно большое количество корпусов изображений лиц как в масках, так и без них. Однако в общей сложности изображений лиц людей, у которых надета маска, значительно меньше, чем лиц без масок, это связано с тем, что сбор данных является сложным процессом из-за введенных во многих странах ограничительных мер. Поэтому некоторые ученые используют синтетический способ получения необходимых данных, а именно компьютерное наложение изображений масок на лица людей. Стоит отметить, что такие данные не всегда выглядят естественно, так как в некоторых случаях маски не перекрывают необходимые области лица, либо, наоборот, перекрывают значительно большую часть лица.

2.2. Подходы к автоматическому распознаванию маски на лице человека. За последние годы было проведено множество исследований, направленных на обнаружение защитных масок на лице человека с использованием видеоинформации. Авторы корпуса MAFA [20] предложили метод LLE-CNN, который состоит из 3 этапов: сначала извлекаются координаты точек лицевых ориентиров, затем применяется локально-линейное встраивание, потом выполняется идентификация потенциальных областей лица и уточнение их позиций путем совместного выполнения задач классификации и регрессии в рамках сверточной нейронной сети (англ. Convolutional Neural Network, CNN).

[1] https://www.imdb.com/

В работе [28] предлагается новый подход к детектированию лиц AOFD (Adversarial Occlusion-aware Face Detection), который позволяет одновременно обнаруживать лицо с большим процентом окклюзии и сегментировать его закрытые области. В работе [29] используется более новый корпус Medical Face Mask Detection Dataset (MFMD)[2], который включает в себя 2 других: MMD и FMD для обнаружения лиц в медицинских масках.

Таблица 1. Корпусы изображений людей в различных защитных масках

Название корпуса [ссылка на источник]	**Объем корпуса (изображений)**	**Количество экземпляров в каждом классе**	**Способ наложения масок**
MAFA[3] [20]	~30 тыс.	~ 35 тыс. лиц в масках	Реальный
MMD[4]	682	3000 лиц в масках	
FMD[5]	853	3232 лиц в масках, 717 лиц без масок, 123 с неправильно надетой маской	
MOXA3k[6] [21]	3000	3015 лиц в масках, 9161 лиц без масок	
Medical Mask detection[7]	6000	6769 лиц в масках, 2086 лиц без масок, 227 с неправильно надетой маской	
FMC[8]	438	219 лиц в масках, 219 лиц без масок	
Face Mask Detection Video Dataset[9] [22]	4357	8306 лиц в масках, ~13 тыс. лиц без масок	
SMFD[10]	1570	785 лиц в масках, 785 лиц без масок	Синтетический
CASIA-Webface[11] [24]	~804 тыс.	Нет данных	
LFW[12] [13][26]	26 тыс.	13 тыс. лиц в масках 13 тыс. лиц без масок	
RMFD[14] [23]	~505 тыс.	1025 лиц в масках, 95 тыс. лиц без масок	Реальный и синтетический
Face Mask ~12K Images Dataset[15]	~12 тыс.	5883 лиц в масках, 5909 лиц без масок	Реальный и синтетический
Face Mask detection[16]	4000	1915 лиц в масках, 1918 лиц без масок	

[2] https://www.kaggle.com/mloey1/medical-face-mask-detection-dataset
[3] https://github.com/ThanhNguyenFG/Masked-Face-Detection
[4] https://www.kaggle.com/vtech6/medical-masks-dataset
[5] https://www.kaggle.com/andrewmvd/face-mask-detection
[6] https://shitty-bots-inc.github.io/MOXA/index.html
[7] https://www.kaggle.com/humansintheloop/medical-mask-detection
[8] https://www.kaggle.com/dhruvmak/face-mask-detection
[9] https://data.mendeley.com/datasets/v3kry8gb59/1
[10] https://github.com/prajnasb/observations
[11] https://github.com/Baojin-Huang/Webface-OCC
[12] http://vis-www.cs.umass.edu/lfw/
[13] https://www.kaggle.com/muhammeddalkran/lfw-simulated-masked-face-dataset
[14] https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset
[15] https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset
[16] https://www.kaggle.com/aneerbanchakraborty/face-mask-detection-data

Для эффективной работы систем детектирования различных защитных масок на лицах людей необходим большой объем обучающих данных. На сегодняшний день существуют несколько корпусов с изображениями людей в масках. Многие исследователи, работающие в области искусственного интеллекта, для обнаружения СИЗ на лицах используют несколько корпусов для обучения. В статье [26] предлагается применять гибридную модель, использующую глубокое обучение – нейронную сеть ResNet-50 и классическое обучение – метод опорных векторов (англ. Support Vector Machine, SVM), для обнаружения СИЗ на лицах людей. Для обучения, проверки и тестирования такой системы авторы объединяют 2 корпуса: RMFD [23] и SMFD. И дополнительно для тестирования системы авторы используют корпус LFW с изображениями лиц знаменитей с синтетически наложенными масками. Недостатком представленных систем является то, что авторы работ используют корпусы, содержащие лица в медицинских масках, и остается непонятным, какой результат покажут системы, если лицо будет закрыто другим предметом, например смартфоном.

Авторами данной статьи было проведено исследование по обнаружению СИЗ на лицах людей с апробацией системы при окклюзии на лице иным предметом, отличным от защитной маски [30]. Исследование проводилось на трех корпусах: Medical Mask detection была выбрана для обучения и проверки, а RMFRD и MAFA (после дополнительной разметки) – для тестирования. В работе был предложен гибридный метод, в основе которого лежит объедение двух наборов признаков: глубокие признаки, извлеченные с помощью нейронной сети ResNet-50 [31], и характеристики распределения интенсивности пикселей на изображениях. В качестве классификатора выступает простая полносвязная нейросеть (англ. Fully Connected Neural Network, FCNN). Схема предложенного метода представлена на рисунке 2. Для увеличения точности работы метода используется аугментация данных, а именно аффинные преобразования и расширение обучающих данных посредством Интернета. Вероятность ложного срабатывания предложенного метода при окклюзии на лице другим предметом на корпусе MAFA составляет 45,48%. Эксперименты показали, что проблема ложного обнаружения защитной маски при окклюзии на лице другим предметом остается, и для ее решения необходимо использовать сложные экземпляры на этапе обучения, как класс «лицо без маски».

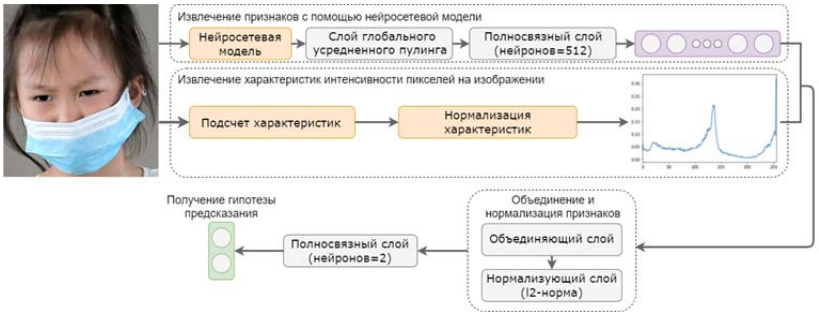

Рис. 2. Функциональная схема метода распознавания наличия/отсутствия защитной маски на лице человека [30]

Системы детектирования защитных масок на лице человека необходимы для того, чтобы в режиме реального времени осуществлять контроль над соблюдением рекомендаций от Всемирной организации здравоохранения (ВОЗ), касаемых ношения средств индивидуальной защиты органов дыхания в общественных местах. Проблемы построения таких систем заключаются в том, что они должны быть небольшого размера из-за ограниченной памяти периферийных устройств, иметь высокую скорость обработки данных и при этом показывать высокую точность обнаружения масок на лицах людей. Для решения данных проблем авторы работы [21] предложили использовать модель обнаружения объектов YOLO v3 Tiny [32] и легкую нейронную сеть MobileNet v2 [33]. С применением предложенных моделей авторам удалось получить максимальную скорость обработки видеоизображений – 138 кадров в секунду. В статье [34] для обнаружения масок в реальном времени предлагается подход под названием SSDMNV2, который включает в себя модель обнаружения объектов Single Shot Multibox Detector (SSD) [35] с базовой архитектурой ResNet-10 [36] и нейронную сеть с архитектурой MobileNet v2 [33] в качестве классификатора. С помощью подхода SSDMNV2 удалось достичь скорости обработки – 15,71 кадров в секунду. Авторы [37] для построения системы обнаружения СИЗ на лице человека использовали 4 различных корпуса: Face Mask ~12K Images Dataset, FMC и 2 корпуса для обучения и тестирования (OpenMV Dataset), собранных авторами статьи [37], содержащих изображения с камеры OpenMV Cam H7, объемами 1979 и 594, соответственно. Для увеличения объема обучающих данных к изображениям применялось масштабирование с различной интерполяцией. В качестве классификатора было предложено использовать CNN, с помощью которой достигается скорость обработки данных – 30 кадров в секунду.

Краткая характеристика рассмотренных работ для задачи распознавания наличия/отсутствия маски на лице отображена в таблице 2. В ней описываются используемые корпусы и методы, а также представлены показатели эффективности предложенных методов.

Таблица 2. Сравнительная характеристика работ по распознаванию наличия/отсутствия масок на лицах с использованием видеоинформации

Работа	**Корпусы**	**Методы**	**Показатель**	**Результат**
Ge S. et al. [20]	MAFA	LLE-CNN	Average precision	76,14%
Chen Y. et al. [28]	MAFA	AOFD		76,14%
Loey M. et al. [29]	MMD, FMD, MAFA	YOLO v2, ResNet-50		81,00%
Roy B. et al. [21]	MOXA3k	YOLO v3 Tiny, MobileNet v2		56,27%
Nagrath P. et al. [34]	MMD, RMFD	SSDMNV2		94,00%
Loey M. et al. [26]	MMD, FMD, MAFA	ResNet-50, SVM	Accuracy	99,64% (RMFD) 99,49% (SMFD) 100% (LFW)
Mohan P. et al. [37]	Face Mask ~12K Images Dataset, FMC, OpenMV Dataset	CNN		99,83%
Ryumina E. et al. [30]	Medical Mask detection, RMFRD, MAFA	Гибридный метод извлечения признаков, FCNN	UAR	98,12% (RMFRD) 97,68% (MAFA)

Обзор существующих современных исследований показал, что в большинстве случаев используются нейросетевые либо гибридные подходы, с помощью которых можно достичь высоких результатов обнаружения защитных масок у человека. Направление детектирования СИЗ на лицах в области машинного обучения стало востребованным относительно недавно, поэтому на сегодняшний день недостаточно изучены современные подходы и их комбинации, которые показывают высокую эффективность на смежных задачах. Также стоит отметить тот факт, что сравнение эффективности методов не всегда возможно, так как при обучении систем обнаружения масок используются различные корпусы, а также их комбинации. Несмотря на полученные высокие результаты работы систем, предложенных различными исследователями, работающими в области детектирования СИЗ на лицах

людей, эффективность систем при тестировании на лицах с высокой степенью окклюзии различными предметами остается неизвестной.

Несмотря на большое количество корпусов с изображениями лиц для детектирования защитных масок, затруднительно обучить надежную систему определения СИЗ у человека, так как в общей сложности изображений с реальными, а не синтетическими масками на лицах, в несколько раз меньше, чем лиц без масок. Генерация синтетических масок является эффективным способом в случае, когда сбор реальных данных является затруднительным процессом, однако наложение синтетических масок на изображения лиц зачастую выглядит неестественно и самое главное не всегда правильно перекрывает необходимые части лица.

3. Системы анализа аудиоинформации. Акустические характеристики речевых высказываний могут отражать полезную информацию, например о физиологическом [38] и психологическом [39] здоровье человека, о проявлениях различных эмоциональных экспрессий [40]. Для предотвращения борьбы с пандемией COVID-19 также можно использовать анализ акустической информации для решения различных задач [41], например, распознавания респираторных заболеваний, в том числе COVID-19, распознавания наличия/отсутствия маски и другие. Распознавание наличия/отсутствия маски на лице человека по акустическим данным является более сложной задачей, но не менее актуальной на сегодняшний день по сравнению с детектированием СИЗ по видеоданным. Кроме того, выявление различных респираторных заболеваний, кашля и коронавирусной инфекции по голосовым характеристикам человека позволяют значительно снизить скорость распространения COVID-19. Для построения эффективных систем анализа аудиоинформации в различных задачах необходимы наборы обучающих данных.

3.1. Корпусы аудиоданных для задач распознавания заболеваний и наличия/отсутствия масок. Респираторные заболевания передаются от человека к человеку воздушно-капельным путем. Одним из основных признаков таких заболеваний является кашель. С помощью анализа аудиоинформации речевых высказываний можно детектировать у человека наличие/отсутствие кашля. Такие системы могут применяться в области здравоохранения, помогая врачам дистанционно наблюдать за состоянием пациента, а также для незамедлительного оказания медицинской помощи при проявлении первых признаков заболеваний. Для разработки таких автоматических систем нужны новые корпусы, позволяющие решать задачу детектирования кашля и респираторных заболеваний. Так, авторы работы [42] создали 2 речевых

корпуса для обнаружения кашля. Первый корпус – Synthetic Cough Sounds содержит полусинтетические данные. Авторы записывали речь с кашлем и без, а также различные фоновые шумы (звуки ветра, дыхания, одышки, различных помещений, офисов, ресторанов, вокзала, многолюдной улицы), которые впоследствии накладывались на записанные речевые высказывания. Помимо кашля корпус содержит также звуки прочистки горла, сопения, чихания, отрыжки, дыхания, отдышки, смеха, храпа, глотания и речи. Данный корпус содержит 26 высказываний с длительностью от 15 до 155 секунд. Количество информантов, их возраст и соотношение полов не уточняются. Информант – человек, служащий источником информации для различных исследовательских целей. Второй корпус – Real Cough Sounds содержит записи пациентов Королевской клиники Эдинбурга (Великобритания), которые болели бронхитом, астмой или хронической обструктивной болезнью легких во время записи корпуса. Корпус записывался в 3 сценариях: запись речи с низким уровнем шума; с фоновым шумом; с шумом, издаваемым информантом во время выполнения различных заданий, например, печатание на клавиатуре, открытие/закрытие окон и так далее. Возраст информантов варьируется от 45 до 72 лет, среди них 3 мужчины и 10 женщин. Всего корпус содержит 78 высказываний.

Отсутствие проявления кашля не может гарантировать здоровое физиологическое состояние человека. При наличии простудных заболеваний у человека может и не проявляться явных респираторных признаков, таких как кашель и чихание, однако акустические характеристики речи человека могут содержать в себе информацию о нездоровом состоянии органов дыхания. Так, в 2017 году проходили международные соревнования по компьютерной паралингвистике INTERSPEECH Computational Paralinguistics Challenge 2017 (ComParE 2017) [43], на которых участникам предлагалось определить наличие/отсутствие простуды по акустическим характеристикам речи. Для этого организаторы предоставили корпус аудиоданных инфекции верхних дыхательных путей (англ. Upper Respiratory Tract Infection Corpus, URTIC), который был разработан Вуппертальским университетом (Германия). Данный корпус содержит высказывания 382 мужчин и 248 женщин. Возраст информантов варьируется от 16 до 84 лет (средний возраст – 29,5 лет). Запись проходила в тихих комнатах. Участники выполняли различные задания: рассказывали истории (последние выходные, отпуск), описывали картинки, произносили различные команды, проговаривали числа, читали короткие рассказы. Сессии длились от 15 минут до 2-х часов, при этом количество заданий варьировалось для каждого информанта. Всего корпус содержит 28652

высказывания с длительностью от 3 до 10 секунд. В обучающую и валидационную выборки включались записи только 210 информантов, из которых 37 были простужены.

В период пандемии анализ акустической информации для обнаружения коронавирусной инфекции является особенно востребованной задачей. В начале 2020 года индийские ученые [44] в целях обнаружения COVID-19 по кашлю, дыханию и голосу разработали новый речевой корпус Coswara, который был собран с помощью платформы краудсорсинга с использованием веб-приложения. С помощью данного приложения информанты добровольно записывали речь, кашель и дыхание на свои электронные устройства, а также проходили анкетирование, в котором указывали параметры записанной речи (тип аудио: кашель, речь, дыхание; отсутствие или присутствие шумов на заднем плане), а также наличие/отсутствие коронавируса у информанта. Данный корпус[17] содержит высказывания 1123 мужчин и 363 женщин, возрастом от 7 до 80 лет.

Ученые из Кембриджского университета (Великобритания) [45] также разработали подобный речевой корпус с использованием приложения для мобильных устройств, записывая информантов с помощью их собственных смартфонов и планшетов. Информанты также предоставляли результаты тестирования на COVID-19, основную демографическую и медицинскую информацию. Кембриджский корпус Cambridge COVID-19 Sound содержит около 10000 аудиозаписей 6613 информантов (4525 мужчин и 2069 женщин). У 406 из них были документально подтвержденные положительные тесты на COVID-19. Кроме того, 691 информант записывались более одного раза. Возрастной диапазон информантов варьировался от 16 до 80 лет. В начале 2021 года количество аудиозаписей и информантов возросло[18]. В записи данного корпуса участвовали жители из Греции, Великобритании, Италии, Германии, Испании, Ирана, США, Бангладеша, Индии и Франции.

Корпус Sonde Health COVID-19 2020 (SHC) [46] состоит из записей речи, собранных из бесплатного приложения Sonde Health[19] на смартфоне. Запись данных происходила в неклинических условиях, например дома, в автомобилях, на рабочих и тому подобных местах. Каждый участник записывал по 3 речевых высказывания, заранее подготовленных организаторами, а также заполнял анкету, в которой ука-

[17] https://iiscleap.github.io/coswara-blog/coswara/2020/11/23/visualize_coswara_data_metadata.html
[18] https://www.covid-19-sounds.org/en/blog/voice_covid_icassp.html
[19] https://www.sondehealth.com/sondeone-page

зывались пол, возраст, субъективная оценка качества записей, результат теста на COVID-19.

Для предотвращения передачи инфекции между людьми воздушно-капельным путем необходимо использовать средства индивидуальной защиты, например, медицинские маски, как здоровым людям, так и инфицированным. Для задачи обнаружения наличия/отсутствия маски на лице информантов по голосовым характеристикам существуют несколько речевых корпусов. Одним из них является речевой корпус Хельсинкского университета (Финляндия) (англ. Speech corpus of University of Helsinki) [47]. Он был разработан для распознавания информантов в масках с целью сокращения преступлений. Корпус содержит речевые высказывания 4 мужчин и 4 женщин, возрастом от 21 до 28 лет. Все они являются носителями финского языка и студентами университета. Корпус записывался в звукоизолированной комнате площадью около 5 квадратных метров. Данные записывались одновременно с 3 микрофонов. Информанты в 2 сеансах зачитывали ряд предложений, а также имитировали спонтанную речь, описывая картинки из комиксов. Каждым информантом было записано 60 аудиофайлов, длительность которых составляет от 60 до 90 секунд.

В рамках международной конференции INTERSPEECH 2020 на соревновании ComParE 2020 [48] был впервые представлен Аугсбургский речевой корпус (англ. Mask Augsburg Speech Corpus, MASC). Он состоит из речевых высказываний людей, находящихся в медицинской одноразовой маске и без нее. Корпус содержит аудиозаписи носителей немецкого языка (16 женщин и 16 мужчин, возрастом от 20 до 41 года, средний возраст 25,6). Записи были сделаны в звукоизоляционной аудиостудии с использованием конденсаторного микрофона AKG C4500 BC. Исходные аудиозаписи (48 кГц, 24 бит на отсчет) в дальнейшем преобразовывались в аудиофайлы с частотой дискретизации 16 кГц, 16 бит на отсчет. Во время записи участники выполняли разные речевые задания в маске и без: они отвечали на вопросы, читали текст с различными медицинскими терминами, а также описывали словами рисунки со спортивными мероприятиями, семьями, детьми и едой. Аудиофайлы были разделены на фрагменты продолжительностью в 1 секунду без наложения.

Сбор естественной речи является затруднительным процессом, поэтому некоторые ученые прибегают к сбору синтезированной речи. Речевой корпус Mask Sorbonne Speech Corpus (MSSC) [49] был собран одной из команд участников соревнований ComParE 2020 для увеличения размера обучающего корпуса MASC. Для имитации человеческого голоса авторы использовали акустическую систему Bose Sound-

Link micro. Процесс записи корпуса проходил в два этапа. Сначала с помощью динамика проигрывались 1000 высказываний от 30 информантов (15 мужчин и 15 женщин), которые были выбраны из речевого корпуса German Distant Speech Data Corpus [50], затем на колонку надевалась медицинская маска и снова проигрывались эти же высказывания. Таким образом авторы создали синтетический корпус, который содержит высказывания людей в масках и без них. Процесс записи происходил в безэховой камере.

В таблице 3 представлены основные параметры рассмотренных речевых корпусов: количество информантов, аудиозаписей, длительность, и количество экземпляров в каждом классе.

Таблица 3. Речевые корпусы для детектирования кашля, наличия/отсутствия маски и COVID-19 (н/д – не доступно)

Название	**Кол-во информ.**	**Длительность, часов**	**Количество экземпляров в каждом классе**
Детектирование кашля и простуды			
Synthetic Cough Sounds [42]	н/д	~3	н/д
Real Cough Sounds [42]	13	~26	н/д
Upper Respiratory Tract Infection Corpus, URTIC [43]	630	~45	1987 (речь простуженных людей), 17070 (речь здоровых людей), + 9551 (скрытая выборка)
Детектирование COVID-19 по речи и звукам			
Coswara[20] [44]	1486	н/д	1498 (речь здоровых людей), 157 (речь людей с COVID-19), 117 (речь людей с респ. инфекцией)
Cambridge COVID-19 Sound [45]	32 тыс.	н/д	н/д
Sonde Health COVID-19 2020 (SHC) [46]	66	0,33	66 (речь инфицированных людей), 132 (речь здоровых людей)
Детектирование наличия/отсутствия маски на лице информанта			
Speech corpus of University of Helsinki [47]	8	~12	96 (речь в шлеме), 96 (речь в резиновой маске), 96 (речь в медицинской маске), 96 (речь в шарфе), 96 (речь без маски)
Mask Augsburg Speech Corpus, MASC [48]	32	~13	~18 тыс. (речь с маской) ~19 982 (речь без маски)
Mask Sorbonne Speech Corpus, MSSC [49]	30	н/д	1000 (речь без маски) 1000 (речь с маской)

[20] https://github.com/iiscleap/Coswara-Data

Наличие корпусов, содержащих акустические характеристики речевых высказываний людей, говорит о том, что распознавание кашля, наличия/отсутствия маски на лице человека, а также обнаружение коронавирусной инфекции COVID-19 являются актуальными задачами в области анализа речи и аудиосигналов, особенно в условиях пандемии.

3.2. Подходы к автоматическому распознаванию респираторных заболеваний. Еще задолго до наступления пандемии COVID-19 ученые начали проводить исследования по выявлению различных респираторных заболеваний по голосовым характеристикам человека. Так, в работе [51] представлена система, построенная с применением скрытых марковских моделей (англ. Hidden Markov model, HMM), обнаружения кашля у амбулаторных больных, которые носили на груди звукозаписывающее устройство и микрофон. Но ношение специального периферийного устройства для записи кашля для больных пациентов становится затруднительно, поэтому предложено детектировать наличие кашля из аудиозаписей со смартфона. Авторы статьи [42] предложили использовать локальные моментные характеристики, полученные из мел-частотных кепстральных коэффициентов (англ. Mel-Frequency Cepstral Coefficients, MFCCs), и классификатор на основе метода k-ближайших соседей (kNN) для обнаружения кашля по записям со смартфона. Зачастую запись на смартфоны происходит в достаточно зашумленных условиях, поэтому обнаружение кашля по аудиозаписям становится затруднительным. В работе [52] также использовался kNN для обнаружения кашля в зашумленных звуковых сигналах. Для проверки эффективности предложенного метода авторы собрали корпус аудиозаписей, содержащих кашель, и наложили на них различные шумовые звуки.

Победители соревнований INTERSPEECH ComParE-2017 в своей статье [53] достигли наилучших результатов, используя комбинацию нескольких классификаторов: глубокие нейронные сети (англ. Deep Neural Networks, DNN) для извлечения акустических признаков и SVM в качестве классификатора. Системы, направленные на распознавание заболеваний дыхательных путей по аудиосигналам человека, находят активное применение в телемедицине. Такая система, работающая на смартфоне или на любом другом портативном устройстве, может помочь врачам осуществлять дистанционный мониторинг пациентов, находящихся как на территории медицинских учреждений, так и в домашних условиях.

3.3. Подходы к автоматическому распознаванию COVID-19 по речи и звукам человека. В период пандемии анализ акустической

информации для обнаружения коронавирусной инфекции является особенно востребованной задачей. Поэтому в 2021 году в рамках международной конференции INTERSPEECH 2021 было анонсировано несколько соревнований по определению COVID-19 по кашлю, дыханию и речи. На ежегодном соревновании по компьютерной паралингвистике ComParE 2021 [54] организаторы предложили исследователям две задачи: обнаружение коронавируса по кашлю и по речи людей (инфицированных COVID-19 и заболевших иными респираторными инфекциями). Участникам предоставлялись около 2000 записей речевого корпуса Cambridge COVID-19 Sound [45]. Базовый подход, с помощью которого удалось достичь точности по показателю UAR=73,9% и UAR=72,1% для задач распознавания COVID-19 по кашлю и по речи, соответственно, представлен в статье [54]. В данных соревнованиях приняли участие несколько команд. В работе [55] в качестве признаков используется необработанный звук в сочетании с различными вариациями спектрограмм, в качестве классификаторов используется ансамбль CNN моделей с различной архитектурой. Авторы работы [56] применяют фонетический анализ и модель Гауссовой смеси (англ. Gaussian Mixture Model, GMM) для распознавания COVID-19 по речи. В работе [57] проводятся исследования с тем же корпусом, но уже не в рамках соревнований. Авторы работы отбирают данные с учетом сбалансированности классов и выделяют свои обучающую, валидационную и тестовую выборки. В качестве признаков авторы используют спектрограммы, прошедшие через 128 мел-фильтров, такое представление спектрограммы известно как мел-спектрограмма.

Также в рамках INTERSPEECH-2021 проводится соревнование DiCOVA [58], включающее в себя несколько задач: детектирование COVID-19 по кашлю и по дыханию. Для секции обнаружения COVID-19 по кашлю предоставлен корпус Coswara [44]. Авторы соревнований предложили использовать MFCCs аудиопризнаки совместно с 3 различными классификаторами на основе: логистической регрессии (англ. Logistic Regression, LR), многослойного перцептрона (англ. Multilayered Perceptron, MP) и случайного леса (англ. Random Forest, RF), с помощью которых достигается базовый результат по показателю площади под кривой (англ. Area Under Curve, AUC) в 69,71%.

Краткая характеристика рассмотренных работ для задачи детектирования наличия/отсутствия кашля и простуды и COVID-19 представлена в таблице 4. В ней описываются используемые признаки и классификаторы для построения систем, а также приведены показатели и их числовые значения.

Таблица 4. Сравнительная характеристика работ по анализу аудиоинформации в задачах распознавания респираторных заболеваний

Работа	Аугментация и признаки	Машинный классификатор	Количественный показатель и его значение
Детектирование кашля и респираторных заболеваний			
Matos S. et al. [51]	н/д	HMM	Accuracy = 82,00%
Monge-Álvarez J. et al. [42]	Локальные моментные характеристики	kNN	Accuracy = 95,28%
Monge-Álvarez J. et al. [52]	MFCCs	kNN	Accuracy = 85,71%
Gosztolya G. [53]	Признаки из DNN	SVM	UAR=64,00%
Детектирование COVID-19 по речи и звуку			
Schuller B. et al. [54]	openSMILE, BoAW, auDeep, DeepSpectrum	SVM	UAR= 73,90% (кашель), UAR= 72,10% (речь)
Schuller B. et al. [55]	Необработанный звук с различными спектрограммами	CNN	UAR = 76,10% (кашель), UAR = 73,10% (речь)
Klumpp P. et al. [56]	Спектрограммы	GMM, SVM, LR	UAR = 64,2%
Xia T. et al. [57]	Мел-спектрограммы	CNN	ROC-AUC = 74,00%
Muguli A. et al. [58]	MFCCs	LR, RF, MP	AUC = 69,71%

3.4. Подходы к автоматическому распознаванию наличия/отсутствия маски на лице информанта. Помимо детектирования респираторных заболеваний, также немаловажно контролировать наличие/отсутствие защитной маски на лице инфицированного человека. Анализ видеоинформации не всегда может показывать необходимую эффективность в детектировании маски на лице человека, а в некоторых случаях может быть вообще невозможен, например, при плохом освещении в помещении, удаленности камеры от человека, либо в клиентских сервисах, где камера рассчитана на одного человека, а общение происходит с несколькими, также в государственных учреждениях, где видеосъемка запрещена. Поэтому в случаях отсутствия возможности получить видеоинформацию о наличии/отсутствии маски на лице можно анализировать акустические характеристики голоса человека, которые содержат информацию о наличии/отсутствии защитной маски. Было произведено множество исследований, анализирующих влияние защитной маски на качество и разборчивость речи. В 2008 г. была опубликована статья [59] об анализе изменений акустических характеристик речи медицинского персонала при использовании защитной маски. Спектральные характеристики аудиозаписей показывают различия речевого сигнала при использовании маски и без

нее, тем не менее, наличие маски на лице врачей почти никак не отражается на восприятии и понимании его речи другим персоналом при отсутствии проблем со слухом. С наступлением пандемии COVID-19 начали активно проводиться исследования по разборчивости речи человека в маске. Авторы работы [60] выбрали 156 фраз на английском языке с низкой предсказуемостью из корпуса Speech Perception in Noise (SPIN) [61], которые произносили 2 информанта по несколько раз в масках и без маски, и с различным стилем произношения речи: четким, небрежным и позитивно эмоциональным. Затем эти аудиозаписи давали прослушать 63 экспертам-аудиторам, которые должны были идентифицировать слова в речи информанта, далее эти текстовые данные записывались в один файл[21]. Исследования показывают, что говорящие адаптируют свою речь в зависимости от наличия маски на лице. Разборчивость четко произносимой речи даже лучше, когда человек находится в маске, нежели без нее, при эмоциональной речи получились противоположные результаты, с наличием маски у информантов речь становится хуже воспринимаемой, а при небрежной речи, как при наличии маски, так и без нее, значение разборчивости слов и звуков речи не меняется. В работе [62] представлены исследования влияния различных типов масок на разборчивость речи: одноразовых медицинских и многоразовых тканевых масок, а также респираторов. Для экспериментов использовался имитатор человеческой головы и рта, на который надевали различные маски и записывали синтезированную речь как в масках, так и без них. Результаты анализа показали, что все виды масок ухудшают разборчивость речи, более затруднительно речь воспринимается, когда информант носит двухсторонние тканевые маски. Аналогичные результаты показало исследование, описанное в работе [63], при использовании таких типов «масок» как мотоциклетного шлема, резиновой маски, шарфа и одноразовой медицинской маски.

Организаторы соревнований ComParE 2020 [48] для задачи распознавания маски на лице информанта предоставили базовые признаки речевых высказываний, такие как openSMILE ComParE [64], BoAW [65], auDeep [66], DeepSpectrum [67]. В данных международных соревнованиях приняли участие несколько команд. Большинство команд в качестве признаков использовали спектрограммы. Один из предложенных методов [68] для решения данной задачи заключался в использовании нескольких SVM, которые обучались на MFCCs, векторах Фишера (англ. Fisher Vectors, FV) и всех базовых признаках. Затем на полученных вероятностях обучался еще один SVM классификатор. В

[21] https://data.mendeley.com/datasets/74p6w8xx5r/1

работах [69, 49] помимо акустических характеристик речи анализировали также фонетические характеристики. В работе [69] использовали нейронные сети в качестве классификатора, такие как CNN и рекуррентные нейронные сети (англ. Recurrent Neural Network, RNN). Также был представлен ряд исследований [70-72], в которых использовались различные методы аугментации данных. Так, в [71] был рассмотрен нейросетевой подход на основе генеративно-состязательных нейронных сетях (англ. Generative Adversarial Networks, GAN) [73], который не смог превзойти по точности простые методы аугментации (SpecAugment и Mixup). В статье [72] в качестве классификаторов использовались предварительно обученные нейронные сети (англ. Pretrained Audio Neural Networks, PANNs). Победителем оказалась команда Аризонского университета [74]. Американский коллектив в качестве признаков использовал изображения спектрограмм линейного масштаба, а в качестве классификаторов были предложены нейросетевые подходы с использованием предварительно обученных моделей, таких как VGG [75], ResNet [31], DenseNet [76] и AlexNet [77]. Авторы провели анализ специфичных для классов областей изображений спектрограмм с помощью метода создания тепловых карт (англ. Class Activation Mapping, CAM). Основываясь на этом анализе, они предположили, что спектрограммы линейного масштаба работали лучше, потому что они более растянуты в диапазоне частот 3-5 кГц, которые имеют решающее значение для данной задачи классификации, в то время как на спектрограмме в мел-частотном масштабе данный диапазон составляет около 15%.

Авторы настоящей статьи совместно с нидерландскими и немецкими коллегами также принимали участие в данном соревновании [78]. Использование метода кросс-валидации и тщательный выбор предварительно обученных нейросетевых моделей позволили получить самые современные ансамбли DNN, точность распознавания и обобщающая способность которых значительно увеличились в сравнении с базовым существующим подходом. Мы использовали 10 кросс-валидационных выборок, на которых обучалась предобученная модель ResNet18v2 при различных алгоритмах оптимизации: 10 из них с методом адаптивной оценки моментов (англ. Adaptive Moment Estimation, ADAM) и 10 – со стохастическим градиентным спуском (англ. Stochastic gradient descent, SGD). Общая схема данного метода представлена на рисунке 3. В результате мы обучили 20 нейронных сетей ResNet18v2. После этого в каждой кросс-валидационной выборке данных мы взвешивали прогнозы двух сетей (ResNet18v2 с ADAM и ResNet18v2 с SGD). Затем рассчитывалось среднее предсказание меж-

ду полученными значениями на каждой кросс-валидационной выборке. Все нейронные сети минимизировали целевую функцию бинарной кросс-энтропии.

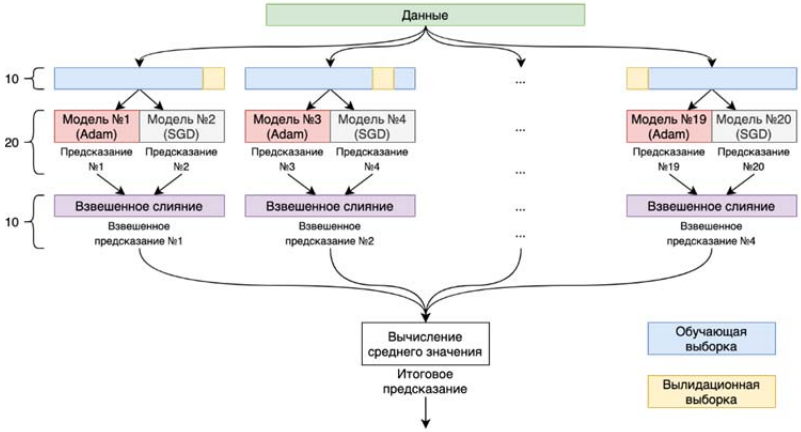

Рис. 3. Схема метода для обнаружения медицинской маски по голосовым характеристикам информанта [78]

Краткая характеристика рассмотренных работ для задачи детектирования наличия/отсутствия маски представлена в таблице 5.

Направление детектирования защитных масок на лицах по акустическим характеристикам речевых высказываний человека является достаточно новой задачей, оно стало востребованным только в период пандемии коронавирусной инфекции COVID-19. Со временем актуальность данного направления только возрастает, так как эффективно работающая система обнаружения наличия/отсутствия инфекции у человека по его кашлю, речи, а также детектирования защитных масок с помощью акустической информации позволит обеспечить контроль состояния человека дистанционно, а также предотвратить процесс распространения вирусной инфекции за счет мониторинга наличия/отсутствия СИЗ на лицах людей. К тому же, данные системы легко можно использовать на устройствах, которыми люди пользуются ежедневно, это поможет осуществлять беспрерывное наблюдение за человеком, а также не будет составлять дискомфорт пользователю при ношении данного устройства.

Таблица 5. Сравнительная характеристика работ по анализу аудиоинформации в задаче детектирования маски

Работа	**Аугментация и признаки**	**Машинный классификатор**	**Количественный показатель и его значение**
Illium et al. [70]	Аугментация (скорость, шум, громкость, смещение, SpecAugment). Спектрограммы	CNN	UAR = 71,50%
Базовое решение ComParE 2020 [48]	openSMILE ComParE, BoAW, auDeep, DeepSpectrum	SVM	UAR = 71,80%
Ristea et al. [71]	Аугментация с помощью GAN. Спектрограммы	ResNet + SVM	UAR = 74,60%
Yang et al. [68]	MFCCs, FV, openSMILE ComParE, BoAW, auDeep, DeepSpectrum	SVM	UAR = 75,10%
Klumpp et al. [69]	Спектрограммы	CNN, RNN	UAR = 75,40%
Markitantov M. et al. [78]	Спектрограммы	ResNet с разными оптимизаторами	UAR = 75,90%
Koike et al. [72]	SpecAugmen, Mixup.Спектрограммы	PANNs	UAR = 77,50%
Montacié et al. [49]	Признаки из ResNet50 и низкоуровневых дескрипторы (openSMILE)	kNN, SVM	UAR = 77,70%
Szep et al. [74]	Линейные спектрограммы	ResNet, DenseNet	UAR = 80,10%

Помимо детектирования респираторных заболеваний и наличия/отсутствия маски у информанта, аудиоинформация помогает определить у человека его возраст, наличие головной боли, отсутствие аппетита и другое. Все эти метаданные являются вспомогательными признаками коронавируса, например люди пожилого возраста более подвержены заражению, поэтому с помощью таких данных можно осуществлять начальный скрининг заражения вирусной инфекцией. Подробный аналитический обзор по методам, используемым в данных задачах, представлен в статье [79].

4. Аудиовизуальные корпусы. На сегодняшний день известны только 2 многомодальных корпуса для задачи распознавания наличия/отсутствия маски на лице человека. Корпус Audio-Visual Face Cover Corpus (AVFCC) [80] содержит аудиовизуальные записи информан-

тов с различными «масками», такими как мотоциклетный шлем, балаклава, обычная медицинская маска. Одновременно записывались две непрерывные видеозаписи в формате HD (1280×720). Звук данного корпуса был записан на портативный 4-х канальный рекордер Edirol R-4 Pro. Чтобы облегчить синхронизацию двух видеозаписей, в начале каждой записи производились хлопки. Данный корпус состоит из 6120 высказываний, которые были записаны 10 носителями английского языка. Корпус содержит аудиозаписи 5 мужчин и 5 женщин с возрастным диапазоном от 21 до 36 лет (средний возраст – 26,5 лет).

Корпус аудиовизуальных русскоязычных данных людей в защитных масках (англ. Biometric Russian Audio-Visual Extended MASKS corpus, BRAVE-MASKS) [81], созданный авторами настоящей статьи, содержит разноракурсные изображения лиц людей в различных вариациях защитных масок, а также аудиозаписи слитной русской речи людей в масках. Корпус предназначен для анализа лицевых и голосовых характеристик информантов в маске и без, для обучения и тестирования системы автоматического распознавания наличия/отсутствия маски, распознавания типа маски, а также бимодальной верификации и идентификации информантов. Запись корпуса проводилась посредством двух смартфонов Apple iPhone XS Max и планшета Apple iPad Pro в офисных условиях на разнородном фоне. Все видеофайлы были обработаны экспертом и разбиты на группы в зависимости от используемой информантом маски. Записанный корпус имеет следующие параметры: 30 информантов (15 женщин и 15 мужчин) разных возрастов (от 19 до 86 лет, средний возраст – 40 лет). Общее количество вариаций защитных масок и респираторов составляет 33 штуки (одноразовые медицинские маски разных цветов, многоразовые тканевые маски различных цветов с рисунками и без, медицинские и строительные респираторы и другие средства защиты слизистых поверхностей, в которых лицо остается частично видимым). Каждому информанту предоставлялось 5 разных защитных масок. Каждый информант проговаривал 83 высказывания на русском языке в течение 6 сессий, в общей сложности BRAVE-MASKS содержит 498 высказываний. Суммарная длительность всех речевых высказываний с 3-х устройств записи для каждого информанта составляет приблизительно 130 минут. Затем, используя эти же маски, записывались вращения головой с 7 различных точек в комнате (с расстояния 1, 2 и 3 метра под разными углами). Оптическое разрешение видеоданных – 4K 3840x2160 пикселей, частота кадров – 30 (для планшета) и 60 (для смартфонов) кадров в секунду, цветность – 24 бита на пиксель. Аудиофайлы без сжатия в

формате PCM WAV с частотой дискретизации 48 кГц, 16 бит на отсчет.

Сбор многомодальных корпусов является затруднительным процессом, так как он требует наличия большого количества времени, участников, специального оборудования. Также основной проблемой сбора многомодальных корпусов является аннотирование данных, так как важно чтобы метки данных каждой модальности были синхронизированы между собой. К тому же с наступлением коронавирусной пандемии многие люди оказались на самоизоляции, поэтому появилась еще одна проблема сбора корпусов – это сложность в получении однородных данных (с одинаковым освещением, идентичным звуковым качеством записей и тому подобные).

5. Заключение. Возрастающее количество работ, направленное на детектирование масок на лице человека как по видеоинформации, так и по аудиоинформации, показывает актуальность данной задачи в нынешнее время. Существует много исследований на тему обнаружения масок по видеоизображениям, также в открытом доступе можно найти значительное количество корпусов, содержащих изображения лиц как без масок, так и в масках, полученных различными способами. Исследований и разработок, направленных на детектирование средств индивидуальной защиты органов дыхания по акустических характеристиках речи человека, пока достаточно мало, так как это направление начало развиваться только в период пандемии, вызванной коронавирусной инфекцией COVID-19. Ранее проводились исследования, косвенно относящиеся к проблемам обнаружения масок по речи информанта, например, обнаружение различных респираторных заболеваний по голосу, влияние защитных масок на разборчивость и восприятие речи.

Существующие системы позволяют предотвратить распространение коронавирусной инфекции с помощью распознавания наличия/отсутствия масок на лице, также данные системы помогают в дистанционном диагностировании COVID-19 с помощью обнаружения первых симптомов вирусной инфекции по акустическим характеристикам. Однако на сегодняшний день существует ряд нерешенных проблем в области автоматического диагностирования COVID-19 и наличия/отсутствия масок на лицах людей. В первую очередь это низкая точность обнаружения масок и коронавирусной инфекции, что не позволяет осуществлять автоматическую диагностику без присутствия экспертов (медицинского персонала). Многие системы не способны работать в режиме реального времени, из-за чего невозможно производить контроль и мониторинг ношения защитных масок в обществен-

ных местах. Также большинство существующих систем невозможно встроить в смартфон, чтобы пользователи могли в любом месте произвести диагностирование наличия коронавирусной инфекции. Еще одной основной проблемой является сбор данных пациентов, зараженных COVID-19, так как многие люди не согласны распространять конфиденциальную информацию.

На основе этого можно выделить следующие требования, выдвигаемые к разрабатываемым автоматическим системам определения наличия/отсутствия масок на лице человека:

1. Использование видео- и аудио модальностей. В том случае, когда анализ одной модальности невозможен, можно получить информацию о наличии/отсутствии маски с помощью другой модальности. К тому же анализ обеих модальностей позволяет более точно производить детектирование наличие/отсутствие маски на лицах людей.

2. Высокая точность распознавания наличия/отсутствия маски на лицах. В период пандемии особенно важно максимально точно распознавать СИЗ на лицах людей. Разработанные системы должны детектировать наличия/отсутствия масок с точностью более 98%.

3. Системы видеоанализа помимо детектирования наличия/отсутствия масок должны распознавать тип маски и выявлять окклюзии на лицах различными предметами, отличные от СИЗ. А также уметь различать реальные маски от, например, нарисованных на лицах красками.

4. Системы аудиоанализа должны работать со звуками с различных микрофонов, различной дальности от информантов, учитывать телефонные разговоры, работать как в помещениях, так и в общественных местах, учитывать возможный эффект реверберации. Кроме того, такие системы должны быть независимыми от информантов, то есть система должна быть построена при обучении на одних информантах, а проверка системы должна быть выполнена на других информантах.

5. Также системы должны распознавать неправильно надетые маски. Многие люди носят маски, не закрывая носовую и/или ротовую полость, что представляет риск заражения вирусной инфекции.

6. Тестирование систем необходимо производить в реальных условиях, учитывая такие параметры, как различное освещение, дальность объекта от камеры, посторонние шумовые звуки, плохое качество связи. Также при обучении использовать изображения лиц и речевые записи людей в реальных масках, а не синтетических.

В дальнейших исследованиях планируется обучить классификаторы распознавания наличия/отсутствия маски на лице человека по

аудио- и видеоинформации с помощью собранного авторами текущей статьи корпуса BRAVE-MASKS. А также на основе этих данных построить системы распознавания типа маски и верификации информантов с масками на лицах.

Литература

1. Habib A. et al. Global Epidemiology of COVID-19 and the Risk of Second Wave. Journal of Drug Delivery and Therapeutics. 2021. vol. 11. no. 1. pP. 1–2.
2. Иванов В.А., Часовская Ю.С. Маски-индивидуальные средства защиты от воздушно-капельных инфекций // Интегративные тенденции в медицине и образовании. 2020. Т. 3. С. 47–53.
3. Boškoski I., Gallo C., Wallace M.B., Costamagna G. COVID-19 pandemic and personal protective equipment shortage: protective efficacy comparing masks and scientific methods for respirator reuse. Gastrointestinal endoscopy. 2020. vol. 92. no. 3. P. 519–523.
4. Macintyre C.R., Chughtai A.A. Facemasks for the prevention of infection in healthcare and community settings. Bmj. 2015. vol. 350.
5. Abdulwhhab M.T. Use of Face-Mask Sampling as a Means of Characterising the Microbiota Exhaled from Human Respiratory Tract in Health and Disease: дис. – University of Leicester. 2020.
6. Нагиев М.Р., Нестерова Н.В. Анализ осведомленности населения об эффективности использования одноразовых медицинских масок в профилактике ОРЗ и ОРВИ, а также перспектива использования лигнина гидролизного в их усовершенствовании // Молодой ученый. 2020. №. 20. С. 207–211.
7. Jiang F. et al. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). Journal of general internal medicine. 2020. vol. 35. no. 5. pp. 1545–1549.
8. Badillo-Goicoechea E., Chang T-H., Kim E., LaRocca S., Morris K., Deng X., Chiu S., Bradford A., Garcia A., Kern C., Cobb C., Kreuter F., Stuart E.A. Global trends and predictors of face mask usage during the COVID-19 pandemic. arXiv preprint arXiv:2012.11678. 2020.
9. Eikenberry S.E. et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infectious Disease Modelling. 2020. vol. 5. pp. 293–308.
10. Гольдштейн Э.М. Факторы, влияющие на смертность от новой коронавирусной инфекции в разных субъектах Российской Федерации // Журнал микробиологии, эпидемиологии и иммунобиологии. 2021. Т. 97. №. 6. С. 604–607.
11. Мусихин И.Г. и другие. Ношение медицинских масок как эффективный способ защиты от covid-19 // Современное общество: опыт, проблемы и перспективы развития. 2021. С. 5–17.
12. Chughtai A.A., Seale H., Macintyre C.R. Effectiveness of cloth masks for protection against severe acute respiratory syndrome coronavirus 2. Emerging infectious diseases. 2020. vol. 26. no. 10.
13. Singh A. et al. Social perception and practices of households regarding mask use in public places during COVID-19 postquarantine period. BLDE University Journal of Health Sciences. 2020. vol. 5. no. 2. P. 209.
14. Rahimi Z. et al. Face mask use among pedestrians during the Covid-19 pandemic in Southwest Iran: an observational study on 10,440 people. BMC Public Health. 2021. vol. 21. no. 1. pp. 1–9.

15. Haischer M.H. et al. Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic. PloS one. 2020. vol. 15. no. 10. P. e0240785.
16. Peretti-Watel P. et al. Attitudes about COVID-19 lockdown among general population, France, March 2020. Emerging infectious diseases. 2021. vol. 27. no. 1. pp. 301–303.
17. Буркова В.Н., Феденок Ю.Н. Медицинская маска как средство индивидуальной и коллективной защиты в условиях пандемии COVID-19 (кросс-культурные аспекты) // Вестник антропологии. (Herald of Anthropology) 2021. Т. 51. №. 3. С. 74–91.
18. Natnael T. et al. Facemask wearing to prevent COVID-19 transmission and associated factors among taxi drivers in Dessie City and Kombolcha Town, Ethiopia. PloS one. 2021. vol. 16. no. 3. P. e0247954.
19. Gunasekaran G.H. et al. Prevalence and acceptance of face mask practice among individuals visiting hospital during COVID-19 pandemic: an observational study. Preprints 2020. 2020.
20. Ge S. et al. Detecting masked faces in the wild with lle-cnns. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 2682–2690.
21. Roy B. et al. MOXA: A Deep Learning Based Unmanned Approach For Real-Time Monitoring of People Wearing Medical Masks. Transactions of the Indian National Academy of Engineering. 2020. vol. 5. no. 3. pp. 509–518.
22. Faisal N., Wasiq K., Salwa Y., Abir H. Face Mask Detection Video Dataset. Mendeley Data. 2020.
23. Wang Z. et al. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093. 2020.
24. Huang B. et al. When Face Recognition Meets Occlusion: A New Benchmark. ICASSP. 2021. pp. 4240–4244.
25. Yi D. et al. Learning face representation from scratch. arXiv preprint arXiv:1411.7923. 2014.
26. Loey M. et al. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement. 2021. vol. 167. P. 108288.
27. Learned-Miller E. et al. Labeled faces in the wild: A survey. Advances in face detection and facial image analysis. 2016. pp. 189–248.
28. Chen Y. et al. Adversarial occlusion-aware face detection. 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). 2018. pp. 1–9.
29. Loey M. et al. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable cities and society. 2021. vol. 65. P. 102600.
30. Ryumina E., Ryumin D., Ivanko D., Karpov A. Novel Method for Protective Face Mask Detection Using Convolutional Neural Networks and Image Histograms. International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences. 2021. vol. XLIV-2/W1-2021. pp. 177–182.
31. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778.
32. Redmon J., Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018.
33. Sandler M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 4510–4520.

34. Nagrath P. et al. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable cities and society. 2021. vol. 66. P. 102692.
35. Liu W. et al. Ssd: Single shot multibox detector. Lecture Notes in Computer Science. 2016. vol. 9905. P. 21–37.
36. Anisimov D., Khanova T. Towards lightweight convolutional neural networks for object detection. 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS). 2017. pp. 1–8.
37. Mohan P., Paul A.J., Chirania A.A. Tiny CNN Architecture for Medical Face Mask Detection for Resource-Constrained Endpoints. Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering. 2021. vol. 756.
38. Вашкевич М.И., Азаров И.С. Определение патологии голосового аппарата на основе анализа модуляционного спектра речи в критических полосах. // Труды СПИИРАН. 2020. № 2 (19). С. 249–276.
39. Авдеев В.Б., Трушин В.А., Кунгуров М.А. Унифицированная речеподобная помеха для средств активной защиты речевой информации // Информатика и автоматизация. 2020. № 5 (19). С. 991–1017.
40. Dvoynikova A., Verkholyak O., Karpov A. Emotion Recognition and Sentiment Analysis of Extemporaneous Speech Transcriptions in Russian. Lecture Notes in Computer Science. 2020. vol. 12335 LNAI. pp. 136–144.
41. Deshpande G., Schuller B.W. Audio, Speech, Language, & Signal Processing for COVID-19: A Comprehensive Overview. arXiv preprint arXiv:2011.14445. 2020.
42. Monge-Alvarez J. et al. Robust detection of audio-cough events using local hu moments. IEEE journal of biomedical and health informatics. 2018. vol. 23. vol. 1. pp. 184–196.
43. Schuller B., et al. The Interspeech 2017 computational paralinguistics challenge: Addressee, cold & snoring. INTERSPEECH. 2017. pp. 3442–3446.
44. Sharma N. et al. Coswara A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. INTERSPEECH. 2020. pp. 4811–4815.
45. Brown C., Chauhan J., Grammenos A. et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20). 2020. pp. 3474–3484.
46. Stasak B. et al. Automatic Detection of COVID-19 Based on Short-Duration Acoustic Smartphone Speech Analysis. Journal of Healthcare Informatics Research. 2021. vol. 5. Is. 2. P. 201–207.
47. Saeidi R., Niemi T., Karppelin H., Pohjalainen J., Kinnunen T., Alku P. Speaker recognition for speech under face cover. INTERSPEECH. 2015. pp. 1012–1016.
48. Schuller B., Batliner A., Bergler C., Messner E., Hamilton A., Amiriparian S., Baird A., Rizos G. The INTERSPEECH 2020 Computational paralinguistics challenge: Elderly emotion, Breathing & Masks. INTERSPEECH. 2020. pp. 2042–2046.
49. Montacié C., Caraty M. Phonetic, Frame Clustering and Intelligibility Analyses for the INTERSPEECH 2020 ComParE Challenge. INTERSPEECH. 2020. pp. 2062–2066.
50. Radeck-Arneth S., Milde B. et al. Open source german distant speech recognition: Corpus and acoustic model. International Conference on Text, Speech, and Dialogue. 2015. pp. 480–488.
51. Matos S. et al. Detection of cough signals in continuous audio recordings using hidden Markov models. IEEE Transactions on Biomedical Engineering. 2006. vol. 53. vol. 6. pp. 1078–1083.
52. Monge-Alvarez J. et al. Audio-cough event detection based on moment theory. Applied Acoustics. 2018. vol. 135. pp. 124–135.

53. Gosztolya G., Busa-Fekete R., Grósz T., Tóth L. DNN-based feature extraction and classifier combination for child-directed speech, cold and snoring identification. INTERSPEECH. 2017. pp. 3522–3526.
54. Schuller B., Batliner A., Bergler C., et al. The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates. INTERSPEECH. 2021. P. 5.
55. Schuller B.W., Coppock H., Gaskell A. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. arXiv preprint arXiv:2012.14553. 2020.
56. Klumpp P., et al The Phonetic Footprint of Covid-19?. INTERSPEECH. 2021.
57. Xia T. et al. Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data. arXiv preprint arXiv:2104.02005. 2021.
58. Muguli A. et al. DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics. arXiv preprint arXiv:2103.09148. 2021.
59. Mendel L.L., Gardino J.A., Atcherson S.R. Speech understanding using surgical masks: a problem in health care?. Journal of the American Academy of Audiology. 2008. vol. 19. vol. 9. pp. 686–695.
60. Cohn M., Pycha A., Zellou G. Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. Cognition. 2021. vol. 210. P. 104570.
61. Kalikow D.N., Stevens K.N., Elliott L.L. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. The Journal of the acoustical society of America. 1977. vol. 61. vol. 5. pp. 1337–1351.
62. Pörschmann C., Lübeck T., Arend J.M. Impact of face masks on voice radiation. The Journal of the Acoustical Society of America. 2020. vol. 148. vol. 6. pp. 3663–3670.
63. Saeidi R., Huhtakallio I., Alku P. Analysis of Face Mask Effect on Speaker Recognition. INTERSPEECH. 2016. pp. 1800–1804.
64. Weninger F., Eyben F., Schuller B., Mortillaro M., Scherer K. On the Acoustics of Emotion in Audio: What Speech, Music and Sound have in Common. Frontiers in Emotion Science. 2013. vol. 4. pp. 1–12.
65. Schmitt M., Schuller B. openXBOW – Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit. Journal of Machine Learning Research. 2017. vol. 18. pp. 1–5.
66. Freitag M., Amiriparian S., Pugachevskiy S., Cummins N., Schuller B. AuDeep: Unsupervised Learning of Representations from Audio with Deep Recurrent Neural Networks. Journal of Machine Learning Research. 2018. vol. 18. pp. 1–5.
67. Amiriparian S., Gerczuk M., Ottl S., Cummins N., Freitag M., Pugachevski S., Schuller B. Snore sound classification using image-based deep spectrum features. INTERSPEECH. 2017. pp. 3512–3516.
68. Yang Z., An Z., Fan Z., Jing C., Cao H. Exploration of Acoustic and Lexical Cues for the INTERSPEECH 2020 Computational Paralinguistic Challenge. INTERSPEECH. 2020. pp. 2092–2096.
69. Klumpp P., Arias-Vergara T., Vásquez-Correa J., Pérez-Toro P, Hönig F., Nöth E., Orozco-Arroyave J. Surgical Mask Detection with Deep Recurrent Phonetic Models. INTERSPEECH. 2020. pp. 2057–2061.
70. Illium S., Müller R., Sedlmeier A., Linnhoff-Popien C. Surgical Mask Detection with Convolutional Neural Networks and Data Augmentations on Spectrograms. INTERSPEECH. 2020. pp. 2052–2056.
71. Ristea N., Ionescu R. Are you Wearing a Mask? Improving Mask Detection from Speech Using Augmentation by Cycle-Consistent GANs. INTERSPEECH. 2020. pp. 2102–2106.

72. Koike T., Qian K., Schuller B., Yamamoto Y. Learning Higher Representations from Pre-Trained Deep Models with Data Augmentation for the COMPARE 2020 Challenge Mask Task. INTERSPEECH. 2020. pp. 2047–2051.
73. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014. vol. 2. pp. 2672–2680.
74. Szep J., Hariri S. Paralinguistic Classification of Mask Wearing by Image Classifiers and Fusion. INTERSPEECH. 2020. pp. 2087–2091.
75. Simonyan K., Zisserman A. Very Deep Convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014. P. 14.
76. Huang G., Liu Z., Van Der Maaten L. Weinberger K. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 4700–4708.
77. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017. vol. 60. vol. 6. pp. 84–90.
78. Markitantov M. et al. Ensembling end-to-end deep models for computational paralinguistics tasks: ComParE 2020 Mask and Breathing Sub-challenges. INTERSPEECH. 2020. P. 2666.
79. Schuller B.W. et al. Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. arXiv preprint arXiv:2003.11117. 2020.
80. Fecher N. The "audio-visual face cover corpus": investigations into audio-visual speech and speaker recognition when the speaker's face is occluded by facewear. INTERSPEECH. 2012. pp. 2250–2253.
81. Корпус аудиовизуальных русскоязычных данных людей в защитных масках (BRAVE-MASKS - Biometric Russian Audio-Visual Extended MASKS corpus). Свидетельство о государственной регистрации Базы данных № 2021621094 от 26.05.2021, авторы: Маркитантов М.В., Рюмин Д.А., Рюмина Е.В., Карпов А.А., правообладатель: СПб ФИЦ РАН.

Двойникова Анастасия Александровна — младший научный сотрудник, лаборатория речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: искусственный интеллект, машинное обучение, нейронные сети, распознавание защитных масок по аудиоинформации, сентимент-анализ текстовых данных. Число научных публикаций — 10. dvoynikova.a@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812) 328 04 21.

Маркитантов Максим Викторович — младший научный сотрудник, лаборатория речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: искусственный интеллект, машинное обучение, речевые технологии, компьютерная паралингвистика, распознавание характеристик диктора, распознавание пола и возраста диктора, обнаружение защитных масок по аудиоинформации. Число научных публикаций — 10. m.markitantov@yandex.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328 04 21.

Рюмина Елена Витальевна — младший научный сотрудник, лаборатория речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: аффективные вычисления, цифровая обработка изображений, распознавание визуальных сигналов, автоматическое распознавание паралингвистических явлений, машинное обучение, нейронные сети, биометрические системы, человеко-машинные интерфейсы.

Число научных публикаций — 12. ryumina.e@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328 04 21.

Рюмин Дмитрий Александрович — канд. техн. наук, старший научный сотрудник, лаборатория речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: цифровая обработка изображений, распознавание образов, автоматическое распознавание визуальной речи, многомодальные интерфейсы, машинное обучение, нейронные сети, биометрия, человеко-машинные интерфейсы. Число научных публикаций — 41. ryumin.d@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328 04 21.

Карпов Алексей Анатольевич — заведующий лабораторией, лаборатория речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика и другие. Число научных публикаций — 300+. karpov@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328 04 21.

Поддержка исследований. Исследование выполнено при поддержке Российского Фонда Фундаментальных Исследований № 20-04-60529, а также частично в рамках бюджетной темы № 0073-2019-0005.

DOI 10.15622/ia.20.5.5

A. DVOYNIKOVA, M. MARKITANTOV, E. RYUMINA, D. RYUMIN, A. KARPOV

ANALYTICAL REVIEW OF AUDIOVISUAL SYSTEMS FOR DETERMINING PERSONAL PROTECTIVE EQUIPMENT ON A PERSON'S FACE

Dvoynikova A., Markitantov M., Ryumina E., Ryumin D., Karpov A. **Analytical Review of Audiovisual Systems for Determining Personal Protective Equipment on a Person's Face.**

Abstract. Since 2019 all countries of the world have faced the rapid spread of the pandemic caused by the COVID-19 coronavirus infection, the fight against which continues to the present day by the world community. Despite the obvious effectiveness of personal respiratory protection equipment against coronavirus infection, many people neglect the use of protective face masks in public places. Therefore, to control and timely identify violators of public health regulations, it is necessary to apply modern information technologies that will detect protective masks on people's faces using video and audio information. The article presents an analytical review of existing and developing intelligent information technologies for bimodal analysis of the voice and facial characteristics of a masked person. There are many studies on the topic of detecting masks from video images, and a significant number of cases containing images of faces both in and without masks obtained by various methods can also be found in the public access. Research and development aimed at detecting personal respiratory protection equipment by the acoustic characteristics of human speech is still quite small, since this direction began to develop only during the pandemic caused by the COVID-19 coronavirus infection. Existing systems allow to prevent the spread of coronavirus infection by recognizing the presence/absence of masks on the face, and these systems also help in remote diagnosis of COVID-19 by detecting the first symptoms of a viral infection by acoustic characteristics. However, to date, there is a number of unresolved problems in the field of automatic diagnosis of COVID-19 and the presence/absence of masks on people's faces. First of all, this is the low accuracy of detecting masks and coronavirus infection, which does not allow for performing automatic diagnosis without the presence of experts (medical personnel). Many systems are not able to operate in real time, which makes it impossible to control and monitor the wearing of protective masks in public places. Also, most of the existing systems cannot be built into a smartphone, so that users be able to diagnose the presence of coronavirus infection anywhere. Another major problem is the collection of data from patients infected with COVID-19, as many people do not agree to distribute confidential information.

Keywords: identification of protective masks, voice characteristics of speakers, facial characteristics, COVID-19, personal protective equipment, cough detection.

Dvoynikova Anastasia — Junior researcher, Laboratory of speech and multimodal interfaces, SPC RAS. Research interests: artificial intelligence, machine learning, neural networks, recognition of protective masks by audio information, sentiment analysis. The number of publications — 10. dvoynikova.a@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812) 328 04 21.

Markitantov Maxim — Junior researcher, Laboratory of speech and multimodal interfaces, SPC RAS. Research interests: artificial intelligence, machine learning, speech technologies, computational paralinguistics, recognition of the speaker's characteristics, speaker's age and gender recognition, detection of protective masks by audio information. The number of publications — 10. m.markitantov@yandex.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328 04 21.

Ryumina Elena — Junior researcher, Laboratory of speech and multimodal interfaces, SPC RAS. Research interests: Affective computing, digital image processing, visual signal recognition, automatic recognition of paralinguistic phenomena, machine learning, neural networks, biometric systems, human-machine interfaces. The number of publications — 12. ryumina.e@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328 04 21.

Ryumin Dmitry — Ph.D., Senior researcher, Laboratory of speech and multimodal interfaces, SPC RAS. Research interests: digital image processing, pattern recognition, automatic visual speech recognition, multimodal interfaces, machine learning, neural networks, biometrics, human-machine interfaces. The number of publications — 41. ryumin.d@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328 04 21.

Karpov Alexey — Head of laboratory, Laboratory of speech and multimodal interfaces, SPC RAS. Research interests: speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics. The number of publications — 300+. karpov@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328 04 21.

Acknowledgements. This research was supported by the Russian Foundation for Basic Research № 20-04-60529, as well as partly within the framework of the budget theme No. 0073-2019-0005.

References

1. Habib A. et al. Global Epidemiology of COVID-19 and the Risk of Second Wave. Journal of Drug Delivery and Therapeutics. 2021. vol. 11. no. 1. pP. 1–2.
2. Ivanov V.A., Chasovskaya Y.S. [Masks - personal protective equipment against airborne infections]. Integrativnye tendencii v medicine i obrazovanii [Integrative trends in medicine and education]. 2020. vol. 3. pP. 47–53. (In Russ).
3. Boškoski I., Gallo C., Wallace M.B., Costamagna G. COVID-19 pandemic and personal protective equipment shortage: protective efficacy comparing masks and scientific methods for respirator reuse. Gastrointestinal endoscopy. 2020. vol. 92. no. 3. P. 519–523.
4. Macintyre C.R., Chughtai A.A. Facemasks for the prevention of infection in healthcare and community settings. Bmj. 2015. vol. 350.
5. Abdulwhhab M.T. Use of Face-Mask Sampling as a Means of Characterising the Microbiota Exhaled from Human Respiratory Tract in Health and Disease: дис. – University of Leicester. 2020.
6. Nagiev M.R., Nesterova N.V. [Analysis of public awareness about the effectiveness of using disposable medical masks in the prevention of acute respiratory infections and acute respiratory viral infections, as well as the prospect of using hydrolytic lignin in their improvement]. Molodoj uchenyj [Young scientist]. 2020. no. 20. pP. 207–211. (In Russ).
7. Jiang F. et al. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). Journal of general internal medicine. 2020. vol. 35. no. 5. pp. 1545–1549.
8. Badillo-Goicoechea E., Chang T-H., Kim E., LaRocca S., Morris K., Deng X., Chiu S., Bradford A., Garcia A., Kern C., Cobb C., Kreuter F., Stuart E.A. Global trends and predictors of face mask usage during the COVID-19 pandemic. arXiv preprint arXiv:2012.11678. 2020.

9. Eikenberry S.E. et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infectious Disease Modelling. 2020. vol. 5. pp. 293–308.
10. Goldstein E.M. [Factors affecting mortality for the novel coronavirus infection in different regions of the Russian Federation]. Zhurnal mikrobiologii, epidemiologii I immunobiologii [Journal of Microbiology, Epidemiology and Immunobiology]. 2021. vol. 97. no. 6. pp. 604–607. (In Russ).
11. Musikhin I.G. et al. [Wearing medical masks as an effective way of protection against covid-19]. Sovremennoye obshchestvo: opyt. problemy i perspektivy razvitiya [Modern society: experience, problems and development prospects]. 2021. pp. 5–17. (In Russ).
12. Chughtai A.A., Seale H., Macintyre C.R. Effectiveness of cloth masks for protection against severe acute respiratory syndrome coronavirus 2. Emerging infectious diseases. 2020. vol. 26. no. 10.
13. Singh A. et al. Social perception and practices of households regarding mask use in public places during COVID-19 postquarantine period. BLDE University Journal of Health Sciences. 2020. vol. 5. no. 2. P. 209.
14. Rahimi Z. et al. Face mask use among pedestrians during the Covid-19 pandemic in Southwest Iran: an observational study on 10,440 people. BMC Public Health. 2021. vol. 21. no. 1. pp. 1–9.
15. Haischer M.H. et al. Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic. PloS one. 2020. vol. 15. no. 10. P. e0240785.
16. Peretti-Watel P. et al. Attitudes about COVID-19 lockdown among general population, France, March 2020. Emerging infectious diseases. 2021. vol. 27. no. 1. pp. 301–303.
17. Burkova V.N., Fedenok J.N. [Medical mask as a means of personal and collective protection in the context of the COVID-19 pandemic (cross-cultural aspects)]. Vestnik antropologii [Herald of Anthropology]. 2021. vol. 51. no. 3. pp. 74–91. (In Russ).
18. Natnael T. et al. Facemask wearing to prevent COVID-19 transmission and associated factors among taxi drivers in Dessie City and Kombolcha Town, Ethiopia. PloS one. 2021. vol. 16. no. 3. P. e0247954.
19. Gunasekaran G.H. et al. Prevalence and acceptance of face mask practice among individuals visiting hospital during COVID-19 pandemic: an observational study. Preprints 2020. 2020.
20. Ge S. et al. Detecting masked faces in the wild with lle-cnns. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 2682–2690.
21. Roy B. et al. MOXA: A Deep Learning Based Unmanned Approach For Real-Time Monitoring of People Wearing Medical Masks. Transactions of the Indian National Academy of Engineering. 2020. vol. 5. no. 3. pp. 509–518.
22. Faisal N., Wasiq K., Salwa Y., Abir H. Face Mask Detection Video Dataset. Mendeley Data. 2020.
23. Wang Z. et al. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093. 2020.
24. Huang B. et al. When Face Recognition Meets Occlusion: A New Benchmark. ICASSP. 2021. pp. 4240–4244.
25. Yi D. et al. Learning face representation from scratch. arXiv preprint arXiv:1411.7923. 2014.
26. Loey M. et al. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement. 2021. vol. 167. P. 108288.

27. Learned-Miller E. et al. Labeled faces in the wild: A survey. Advances in face detection and facial image analysis. 2016. pp. 189–248.
28. Chen Y. et al. Adversarial occlusion-aware face detection. 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). 2018. pp. 1–9.
29. Loey M. et al. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable cities and society. 2021. vol. 65. P. 102600.
30. Ryumina E., Ryumin D., Ivanko D., Karpov A. Novel Method for Protective Face Mask Detection Using Convolutional Neural Networks and Image Histograms. International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences. 2021. vol. XLIV-2/W1-2021. pp. 177–182.
31. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770–778.
32. Redmon J., Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018.
33. Sandler M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 4510–4520.
34. Nagrath P. et al. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable cities and society. 2021. vol. 66. P. 102692.
35. Liu W. et al. Ssd: Single shot multibox detector. Lecture Notes in Computer Science. 2016. vol. 9905. P. 21–37.
36. Anisimov D., Khanova T. Towards lightweight convolutional neural networks for object detection. 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS). 2017. pp. 1–8.
37. Mohan P., Paul A.J., Chirania A.A. Tiny CNN Architecture for Medical Face Mask Detection for Resource-Constrained Endpoints. Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering. 2021. vol. 756.
38. Vashkevich M., Azarov I.. [Voice Pathology Detection based on Analysis of Modulation Spectrum in Critical Bands]. Trudy SPIIRAN [SPIIRAS Proceedings]. 2020. no 2 (19). pp. 249–276. (In Russ).
39. Avdeev V., Trushin V., Kungurov M. [Unified Speech-Like Interference for Active Protection of Speech Information]. Informatika i avtomatizacija [Informatics and Automation]. 2020. no 5 (19). pP. 991–1017. doi: 10.15622/ia.2020.19.5.4. (In Russ).
40. Dvoynikova A., Verkholyak O., Karpov A. Emotion Recognition and Sentiment Analysis of Extemporaneous Speech Transcriptions in Russian. Lecture Notes in Computer Science. 2020. vol. 12335 LNAI. pp. 136–144.
41. Deshpande G., Schuller B.W. Audio, Speech, Language, & Signal Processing for COVID-19: A Comprehensive Overview. arXiv preprint arXiv:2011.14445. 2020.
42. Monge-Alvarez J. et al. Robust detection of audio-cough events using local hu moments. IEEE journal of biomedical and health informatics. 2018. vol. 23. vol. 1. pp. 184–196.
43. Schuller B., et al. The Interspeech 2017 computational paralinguistics challenge: Addressee, cold & snoring. INTERSPEECH. 2017. pp. 3442–3446.
44. Sharma N. et al. Coswara A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. INTERSPEECH. 2020. pp. 4811–4815.
45. Brown C., Chauhan J., Grammenos A. et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. In Proceedings of the 26th

ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20). 2020. pp. 3474–3484.
46. Stasak B. et al. Automatic Detection of COVID-19 Based on Short-Duration Acoustic Smartphone Speech Analysis. Journal of Healthcare Informatics Research. 2021. vol. 5. Is. 2. P. 201–207.
47. Saeidi R., Niemi T., Karppelin H., Pohjalainen J., Kinnunen T., Alku P. Speaker recognition for speech under face cover. INTERSPEECH. 2015. pp. 1012–1016.
48. Schuller B., Batliner A., Bergler C., Messner E., Hamilton A., Amiriparian S., Baird A., Rizos G. The INTERSPEECH 2020 Computational paralinguistics challenge: Elderly emotion, Breathing & Masks. INTERSPEECH. 2020. pp. 2042–2046.
49. Montacié C., Caraty M. Phonetic, Frame Clustering and Intelligibility Analyses for the INTERSPEECH 2020 ComParE Challenge. INTERSPEECH. 2020. pp. 2062–2066.
50. Radeck-Arneth S., Milde B. et al. Open source german distant speech recognition: Corpus and acoustic model. International Conference on Text, Speech, and Dialogue. 2015. pp. 480–488.
51. Matos S. et al. Detection of cough signals in continuous audio recordings using hidden Markov models. IEEE Transactions on Biomedical Engineering. 2006. vol. 53. vol. 6. pp. 1078–1083.
52. Monge-Alvarez J. et al. Audio-cough event detection based on moment theory. Applied Acoustics. 2018. vol. 135. pp. 124–135.
53. Gosztolya G., Busa-Fekete R., Grósz T., Tóth L. DNN-based feature extraction and classifier combination for child-directed speech, cold and snoring identification. INTERSPEECH. 2017. pp. 3522–3526.
54. Schuller B., Batliner A., Bergler C., et al. The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates. INTERSPEECH. 2021. P. 5.
55. Schuller B.W., Coppock H., Gaskell A. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. arXiv preprint arXiv:2012.14553. 2020.
56. Klumpp P., et al The Phonetic Footprint of Covid-19?. INTERSPEECH. 2021.
57. Xia T. et al. Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data. arXiv preprint arXiv:2104.02005. 2021.
58. Muguli A. et al. DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics. arXiv preprint arXiv:2103.09148. 2021.
59. Mendel L.L., Gardino J.A., Atcherson S.R. Speech understanding using surgical masks: a problem in health care?. Journal of the American Academy of Audiology. 2008. vol. 19. vol. 9. pp. 686–695.
60. Cohn M., Pycha A., Zellou G. Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. Cognition. 2021. vol. 210. P. 104570.
61. Kalikow D.N., Stevens K.N., Elliott L.L. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. The Journal of the acoustical society of America. 1977. vol. 61. vol. 5. pp. 1337–1351.
62. Pörschmann C., Lübeck T., Arend J.M. Impact of face masks on voice radiation. The Journal of the Acoustical Society of America. 2020. vol. 148. vol. 6. pp. 3663–3670.
63. Saeidi R., Huhtakallio I., Alku P. Analysis of Face Mask Effect on Speaker Recognition. INTERSPEECH. 2016. pp. 1800–1804.
64. Weninger F., Eyben F., Schuller B., Mortillaro M., Scherer K. On the Acoustics of Emotion in Audio: What Speech, Music and Sound have in Common. Frontiers in Emotion Science. 2013. vol. 4. pp. 1–12.

65. Schmitt M., Schuller B. openXBOW – Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit. Journal of Machine Learning Research. 2017. vol. 18. pp. 1–5.
66. Freitag M., Amiriparian S., Pugachevskiy S., Cummins N., Schuller B. AuDeep: Unsupervised Learning of Representations from Audio with Deep Recurrent Neural Networks. Journal of Machine Learning Research. 2018. vol. 18. pp. 1–5.
67. Amiriparian S., Gerczuk M., Ottl S., Cummins N., Freitag M., Pugachevski S., Schuller B. Snore sound classification using image-based deep spectrum features. INTERSPEECH. 2017. pp. 3512–3516.
68. Yang Z., An Z., Fan Z., Jing C., Cao H. Exploration of Acoustic and Lexical Cues for the INTERSPEECH 2020 Computational Paralinguistic Challenge. INTERSPEECH. 2020. pp. 2092–2096.
69. Klumpp P., Arias-Vergara T., Vásquez-Correa J., Pérez-Toro P, Hönig F., Nöth E., Orozco-Arroyave J. Surgical Mask Detection with Deep Recurrent Phonetic Models. INTERSPEECH. 2020. pp. 2057–2061.
70. Illium S., Müller R., Sedlmeier A., Linnhoff-Popien C. Surgical Mask Detection with Convolutional Neural Networks and Data Augmentations on Spectrograms. INTERSPEECH. 2020. pp. 2052–2056.
71. Ristea N., Ionescu R. Are you Wearing a Mask? Improving Mask Detection from Speech Using Augmentation by Cycle-Consistent GANs. INTERSPEECH. 2020. pp. 2102–2106.
72. Koike T., Qian K., Schuller B., Yamamoto Y. Learning Higher Representations from Pre-Trained Deep Models with Data Augmentation for the COMPARE 2020 Challenge Mask Task. INTERSPEECH. 2020. pp. 2047–2051.
73. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014. vol. 2. pp. 2672–2680.
74. Szep J., Hariri S. Paralinguistic Classification of Mask Wearing by Image Classifiers and Fusion. INTERSPEECH. 2020. pp. 2087–2091.
75. Simonyan K., Zisserman A. Very Deep Convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014. P. 14.
76. Huang G., Liu Z., Van Der Maaten L. Weinberger K. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 4700–4708.
77. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017. vol. 60. vol. 6. pp. 84–90.
78. Markitantov M. et al. Ensembling end-to-end deep models for computational paralinguistics tasks: ComParE 2020 Mask and Breathing Sub-challenges. INTERSPEECH. 2020. P. 2666.
79. Schuller B.W. et al. Covid-19 and computer audition: An overview on what speech & sound analysis could contribute in the sars-cov-2 corona crisis. arXiv preprint arXiv:2003.11117. 2020.
80. Fecher N. The "audio-visual face cover corpus": investigations into audio-visual speech and speaker recognition when the speaker's face is occluded by facewear. INTERSPEECH. 2012. pp. 2250–2253.
81. Korpus audiovizualnykh russkoyazychnykh dannykh lyudey v zashchitnykh maskakh (BRAVE-MASKS - Biometric Russian Audio-Visual Extended MASKS corpus). Svidetelstvo o gosudarstvennoy registratsii Bazy dannykh № 2021621094. 26.05.2021. Authors: Markitantov M.V., Ryumin D.A., Ryumina E.V., Karpov A.A., pravoobladatel: SPC RAS. (In Russ).