ЭВОЛЮЦИЯ И НАПРАВЛЕНИЯ РАЗВИТИЯ ТЕХНОЛОГИЙ МАСКИРОВАНИЯ КОНФИДЕНЦИАЛЬНЫХ РЕЧЕВЫХ СООБЩЕНИЙ

Дураковский А. П.[1], Дворянкин С. В.[2], Дворянкин Н. С.[3]

DOI: 10.21681/2311-3456-2024-5-58-66

Цель исследования: *анализ методов и алгоритмов технического закрытия речевой информации в сетях и системах голосовой связи, оценка направлений и перспектив развития технологий речевого маскирования с машинным обучением.*

Методы исследования: *прикладного системного анализа, цифрового спектрально-временного анализа, цифровой обработки сигналов и изображений, образного анализа спектрограмм, машинного обучения.*

Результаты исследования: *обозначены проблемы обеспечения безопасности конфиденциальной голосовой связи в современных условиях. Приведен обзор методов речевой защиты, применяемых на практике в общедоступных каналах голосовой связи. Рассмотрены традиционные и перспективные алгоритмы маскирования речевых сообщений, способы их реализации. Отмечены преимущества последних.*

Научная новизна: *предложены новые способы технического маскирования речи на основе модификации и реконструкции изображений динамических спектрограмм с использованием методов машинного обучения.*

Практическая значимость: *предложены эффективные алгоритмы речевого маскирования. Полученные результаты позволят расширить возможности существующих решений по защите речевой информации в системах и сетях голосовой связи и проектировать более эффективные на основе изложенных подходов.*

Ключевые слова: *информационная безопасность, защита речевой информации, образный анализ-синтез, техническое закрытие речи, речеподобный сигнал, машинное обучение.*

Введение

Современное состояние проблемы защиты речевой информации (РИ) характеризуется постоянным расширением арсенала средств негласного съема и перехвата акустических (речевых) сигналов, технические характеристики и способы применения которых неуклонно совершенствуются[4,5,6].

В связи с этим особый интерес представляют исследования, направленные на выявление принципиально новых подходов к защите речевой информации от НСД, позволяющих существенно усложнить процесс несанкционированного перехвата речевых и попутных полезных акустических фоновых сигналов из каналов голосовой связи (КГС).

Безопасность голосовой связи при передаче конфиденциальных речевых сообщений по каналам коммуникаций основывается на использовании большого количества методов и средств технического закрытия речевого сигнала (РС)[7]. Они преобразуют характеристики речи таким образом, что она становится неразборчивой, непонятной, неузнаваемой для подслушивающего лица, перехватившего обработанное речевое сообщение. Или вообще скрывается факт самой передачи речевого сообщения, которое тем не менее в таком скрытном виде доходит до своего абонента, адресата.

Сегодня внимание исследователей и потребителей обращено на быстрые алгоритмы маскирования, адаптированные под большинство мобильных устройств и приложений, способные в режиме реального времени преобразовывать речевую информацию в защищенный формат[8], прежде всего делая ее неразборчивой или с полным отсутствием в канале передачи признаков исходной защищаемой речи.

1 Дураковский Анатолий Петрович, кандидат технических наук, доцент, доцент кафедры стратегических информационных исследований НИЯУ МИФИ, директор Аттестационно-испытательного центра информационной безопасности и систем защиты информации НИЯУ МИФИ, г. Москва, Россия. E-mail: apdurakovskiy@mephi.ru

2 Дворянкин Сергей Владимирович, доктор технических наук, профессор, профессор кафедры стратегических информационных исследований НИЯУ МИФИ, заведующий лабораторией защиты и обработки аудиовизуальной информации МГЛУ, г. Москва, Россия. E-mail: svdvoryankin@mephi.ru. Https://orcid.org/0000-0001-6908-0676

3 Дворянкин Никита Сергеевич, аспирант НИЯУ МИФИ, г. Москва, Россия. E-mail: nik.dvrn@gmail.com

4 Дворянкин С. В. Маскирование речевой информации: перспективные методы и средства. С. В. Дворянкин, А. А. Мишуков // Спецтехника и связь. – 2009. – № 3.

5 Мишуков, А. А. Образный анализ и маскирование речевой информации / А. А. Мишуков, Р. А. Устинов, Н. С. Дворянкин // Информационные технологии, связь и защита информации МВД России. – 2012. – Вып. 2.

6 Карпов, А. П. Разработка маскиратора аналоговых речевых сигналов / А. П. Карпов // Вестник Пензенского государственного университета. – 2016. – № 1 (13). – С. 62–64.

7 Дворянкин С. В., Девочкин Д. В. Методы закрытия речевых сигналов в телефонных каналах. // Защита информации. Конфидент. – 1995. – №5. – 45–59с.

8 Сперанский В. С., Клинцов О. И. Методы технического закрытия речевых сообщений // T-Comm. 2011. №9. URL: https://cyberleninka.ru/article/n/metody-tehnicheskogo-zakrytiya-rechevyh-soobscheniy (дата обращения: 15.06.2024).

Традиционные способы технического закрытия речи, области применения

Различают два основных класса способов защиты речевого сигнала в каналах коммуникаций от НСД. Первый, аналогово-цифро-аналоговый или просто аналоговый, относится к техническому закрытию и заключается в создании смеси защищаемого РС с помехой (маскировании) и-или в перемешивании (скремблировании) фрагментов исходного РС некоторым образом, делая речь неразборчивой. Это делается путем изменения соотношений между временем, амплитудой и частотой исходного сигнала.

Второй класс способов, криптографический, состоит в преобразовании речевого сигнала в цифровую форму, к которой применимы стандартные методы дискретного шифрования[9].

По некоторым оценкам в последнее время сфера применения маскирующих и скремблирующих алгоритмов технического закрытия, казалось бы, начала сокращаться. Это объяснялось улучшением качества каналов голосовой связи (КГС), ростом производительности и удешевлением привлекаемых вычислительных ресурсов, появлением экономичных «легковесных» криптографических алгоритмов, что существенно продвинуло применение в засекреченной цифровой связи речевых шифраторов.

Тем не менее, аналогово-цифро-аналоговое скремблирование до сих пор может и используется там, где применение цифровых систем закрытия речи затруднено из-за наличия возможных ошибок при передаче и сжатии данных в каналах связи с плохой пропускной способностью. Например, наземные линии связи с плохими техническими характеристиками, отечественные каналы связи для телефонов общего пользования, каналы дальней радиосвязи, особенно КВ-диапазона[10].

Таким образом, речевые маскираторы и скремблеры до сих пор применяются там, где невозможно, по ряду причин, использовать шифраторы. Кроме того, концептуальные принципы, понятия и решения, заложенные в скремблирующие и маскирующие алгоритмы, используемые в КГС, также, можно распространить на другие области защиты речевой информации. Например, на шумоподавление и реконструкцию искаженных РС, речеподобные помехи в системах активной акустической защиты помещений для конфиденциальных переговоров [1, 2, 3].

И наконец, с ростом и удешевлением вычислительного ресурса существующий научный задел создает основу разработки нового поколения устройств технического закрытия РС (новые маскираторы), имеющих более высокую степень защищенности близкую к шифраторам, улучшенные качество и разборчивость восстановленной речи близкое к аналоговым скремблерам, достаточную экономичность и простоту реализации, скрытность передачи РС и практическое отсутствие признаков защищаемой речи, по которым в отложенном режиме она могла бы частично быть восстановлена злоумышленником (ЗЛ) [3, 4].

Классификация существующих методов технического закрытия речи

Существующие средства защиты речевой информации в КГС, такие как маскираторы и скремблеры, уменьшают возможности устройств несанкционированного перехвата и прослушивания РС. Они позволяют пользоваться открытыми каналами связи, защищая передаваемую РИ от несанкционированного доступа (НСД) со стороны ЗЛ, работать в асинхронным режиме, обеспечивая при этом хорошее качество звучания РС в каналах с помехами и плохой пропускной способностью. Однако они менее стойкие, чем речевые шифраторы, хотя более экономичны при изготовлении и применению.

Классификация наиболее часто встречаемых и хорошо изученных видов скремблеров представлена в работе[11]. Помимо описанных там традиционных методов технического закрытия (скремблирования) в существующих устройствах речевого закрытия применяются различные методы и алгоритмы цифровой обработки сигналов [4, 5] как известные, на основе цифровых фильтров и преобразования Фурье, так и оригинальные, например, связанные с обработкой изображений узкополосных спектрограмм [3].

Более широкая классификация существующих методов технического закрытия (скремблирования и маскирования) представлена в работе[12] (см. рис. 1).

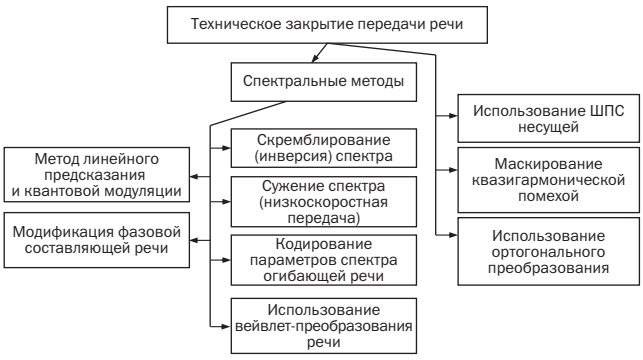

Рис. 1. Классификация современных методов технического закрытия

9 Барсуков В. С., Дворянкин С. В., Шеремет И. А. Безопасность связи в каналах телекоммуникаций. – М.: Электронные знания, 1992–1993. – 122 с.

10 Петраков А. В., Лагутин В. С. Утечка и защита информации в телефонных каналах. – М.: Энергоатомиздат, 1998. 317 с.

11 Петраков А. В., Лагутин В. С. Утечка и защита информации в телефонных каналах. – М.: Энергоатомиздат, 1998. 317 с.

12 Сперанский В. С., Клинцов О. И. Методы технического закрытия речевых сообщений // T-Comm. 2011. №9. URL: https://cyberleninka.ru/article/n/metody-tehnicheskogo-zakrytiya-rechevyh-soobscheniy (дата обращения: 15.06.2024).

Разнообразие представленных на рисунках методов речевого закрытия подтверждает тезис о том, что в настоящее время, в связи с развитием средств вычислительной техники, широким распространением общедоступных каналов передачи аудиоданных в цифровом медиа пространстве, появлением доступных решений в сфере применения методов машинного обучения к аудио обработке, – сформировалась необходимость в разработке эффективных и экономичных средствах технической защиты речевых сообщений, передаваемых в цифровом виде. Эту нишу вполне могли бы занять средства защиты РИ в КГС, построенные на основе новых методов маскирования речевых сообщений и удовлетворяющие современным требованиям.

Требования к перспективным маскираторам речи

Востребованный сегодня тип маскираторов – это асинхронные устройства защиты РИ, имеющие в своей основе такие методы и алгоритмы как: создание полезной смеси защищаемого РС с аддитивной помехой (в том числе речеподобной); изменение спектральной огибающей исходного РС; не требующие обязательной схемы (блока) синхронизации, которая необходима как при скремблировании, так и при дискретизации с шифрованием; практическое отсутствие в маскируемом сигнале признаков исходного РС, по которым методами шумопонижения и реконструкции может быть восстановлена речевая разборчивость [6–17].

В этой связи особый интерес представляют возможности перспективных разрабатываемых маскираторов с использованием решений машинного обучения и технологий синтеза речеподобных сигналов (РПС) с заданными свойствами, объединяющие лучшие характеристики существующих маскираторов и скремблеров и добавляющие новые варианты их использования [6–17].

Методологической основой создания такого рода устройств может послужить уже выше упоминавшийся образный анализ-синтез РС [3, 4].

Последние методы, основанные на технологии образного анализа-синтеза, заключающегося в переходе от волнового представления РС к изображению динамических узкополосных спектрограмм – графическим образам (ГО), их обработке методами цифровой обработки изображений для решения прикладных задач и обратном переходе (синтезе) от нового изображения к новой волновой форме РС, – неплохо подходят для организации различных новых видов асинхронного маскирования РС, в том числе и ранее не известных.

Выбор показателей качества для объективной оценки алгоритмов асинхронного маскирования речи обусловлен наличием особенностей, возникающих при передаче РС по каналу связи.

Требования к алгоритмам асинхронного маскирования речи можно разбить на две основные группы:

- требования по ограничению доступа злоумышленника к речевой информации;
- требования по обеспечению качественного приема речевой информации получателем при наименьших, аппаратных затратах.

Первая группа требований подразумевает создание наихудших условий для прослушивания линии связи злоумышленником, а вторая – обеспечение хорошего качества восстановленной речи при приемлемых технических характеристиках. Качество восстановленной речи, в свою очередь включающее понятия речевой разборчивости (РР) и ее узнаваемости, определяется устойчивостью алгоритма к различному виду помех, а также к рассогласованиям характеристик маскиратора и демаскиратора.

Очевидно, что злоумышленник будет поставлен в наихудшие условия, если не сможет обнаружить сам факт передачи речи в прослушиваемой линии связи. Тогда основными факторами, характеризующими снижение признаков речи в передаваемом сигнале, являются:

- уровень остаточной разборчивости прослушиваемого РС;
- однородное маскирование всех участков и элементов речи (определяются отсутствием пауз в прослушиваемом РС, а также отсутствием резких скачков громкости и тембра);
- «сглаживание» границ между акустически однородными участками РС.

Для определения речевой разборчивости (РР) в отдельном речевом канале среди используемых при организации сеанса связи целесообразно использовать показатели словесной речевой разборчивости, определенные для защищаемых помещений (ЗП) конфиденциальных переговоров от утечки, представленные в таблице 1[13].

Таблица 1.

Цели и критерии эффективности защиты речевой информации

Цель защиты	Критерий эффективности защиты
Скрытие факта ведения переговоров	$W_n \le 10\%$
Скрытие предмета переговоров	$W_n \le 20\%$
Скрытие содержания переговоров	$W_n \le 30\%$

13 Дворянкин С. В., Макаров Ю. К., Хорев А. А. Обоснование критериев эффективности защиты речевой. информации// Защита информации. Инсайд. – С. Петербург.: 2007. – № 2 – с. 18 – 25.

Для сохранения конфиденциальности переговоров в ЗП считается, что если уровень расчетной словесной разборчивости не превышает *20%* (см. табл. 1), то возможный технический канал утечки речевой информации (ТКУРИ) не требует проведения защитных мероприятий[14]. А если расчетная словесная разборчивость превышает *80%* (что соответствует *100%* фразовой), то перехватываемая ЗЛ по каналу ТКУРИ речевая информация будет полностью понятна нарушителю. Эти же выводы можно отнести и к телекоммуникационным каналам речевой связи, защищаемым от НСД. При необходимости в целях достижения еще большего уровня защиты РИ в каждом из используемых каналов можно дополнительно к известным использовать и новые алгоритмы речевого технического маскирования и наоборот, к новым добавлять старые.

Перспективные методы технического закрытия речевых сообщений

Как уже отмечалось методы цифрового шифрования часто достигают высокого уровня безопасности, но даже в современных системах связи со сжатием речи цифровое шифрование не всегда пригодно для использования. Цифровое шифрование, применяемое перед сжатием, может привести к снижению производительности связи, так как ошибки, вызванные методами сжатия с потерями, могут послужить причиной того, что речевые данные не могут быть правильно расшифрованы. Использование цифрового шифрования после сжатия речи требует серьёзных внутренних аппаратных и программных модификаций. Поэтому имеет смысл обратить внимание на новые методы и подходы, появляющиеся в техническом маскировании.

Современное техническое закрытие может быть удобно для защиты конфиденциальности речи в речевых коммуникациях. Какой-либо новый метод технического закрытия может быть использован перед отправкой речи в сети и системы связи без их модификаций, а на другом, приемном конце речевой коммуникации речь может быть восстановлена даже при наличии ошибок сжатия и ошибок канала.

Если алгоритм технического закрытия хорошо продуман, он будет способен даже обеспечить отсутствие признаков исходной речи, по которым она может быть восстановлена, высокую безопасность систем связи, сохраняя приемлемое качество расшифрованной речи при сравнительно низкой стоимости.

Рассмотрим примеры таких перспективных маскираторов.

14 Дворянкин С. В., Макаров Ю. К., Хорев А. А. Обоснование критериев эффективности защиты речевой. информации// Защита информации. Инсайд. – С. Петербург.: 2007. – № 2 – с. 18 – 25.

1. «Сепараторы»: рассечение-разнесение и расслоение речевой информации

Процесс управления РР в данном случае можно представить в виде некоего преобразования: рассечения и-или расслоения («слайдирования») графического образа (ГО) исходного РС на ряд других, мало похожих или совсем непохожих на исходный ГО, по которым синтезируются неразборчивые речеподобные сигналы (РПС), передаваемые в свои каналы связи на передающем конце, и сшивку или объединение их ГО с последующим синтезом в новый разборчивый сигнал на приемном.

Здесь речевой сигнал (РС) каждого из абонентов конфиденциальных переговоров рассматривается как совокупность и-или как сумма нескольких речеподобных сигналов, каждый из которых имеет свою РР со значением менее заданного уровня (нормы) и может быть передан другому собеседнику по своему отдельному каналу.

$$S(t) = \sum_{K} s_k(t) \qquad W_{s_k} \le W_n \qquad S(t) = \bigcup_{K} s_k(t) \qquad (1)$$

где $S(t)$ – исходный РС, $s_k(t)$ – речеподобные составляющие исходного РС, W_{s_k} – текущая РР для каждой речевой составляющей, а W_n – нормированное значение РР.

Такой отдельный РС, будучи возможно перехваченным в одном из контролируемых ЗЛ каналов связи уже не будет понятен нарушителю. У легального же пользователя на приемном конце все полученные по разным путям элементарные сигналы снова объединяются (сшиваются, склеиваются) по определенным правилам в один, теперь уже разборчивый сигнал.

Общая схема такой защищенной голосовой связи для двух абонентов и одно¬временно используемых ими 4-х каналов (3-и сотовых операторов «большой тройки» плюс канал VoIP) показана на рис. 2. Понятно, что эта модель может быть расширена на большее число используемых каналов и участников переговоров.

Класс методов разделения исходного РС на неразборчивые речеподобные составляющие весьма широк: от полосовой фильтрации по группам равно артикуляционных полос до спектрально-временной обработки фонетической функции (динамической огибающей спектра), определяющей РР:

$$P(\omega, t) = \log\left[\frac{S(\omega, t)}{S(\omega, t - \tau)}\right] \qquad (2)$$

где $P(\omega, t)$ – фонетическая функция Пирогова, а $S(\omega, t)$ и $S(\omega, t - \tau)$ – модули кратковременных спектров в соответствующие моменты времени.

Плюс от сеанса к сеансу можно организационно изменять набор участвующих в модели сплиттера

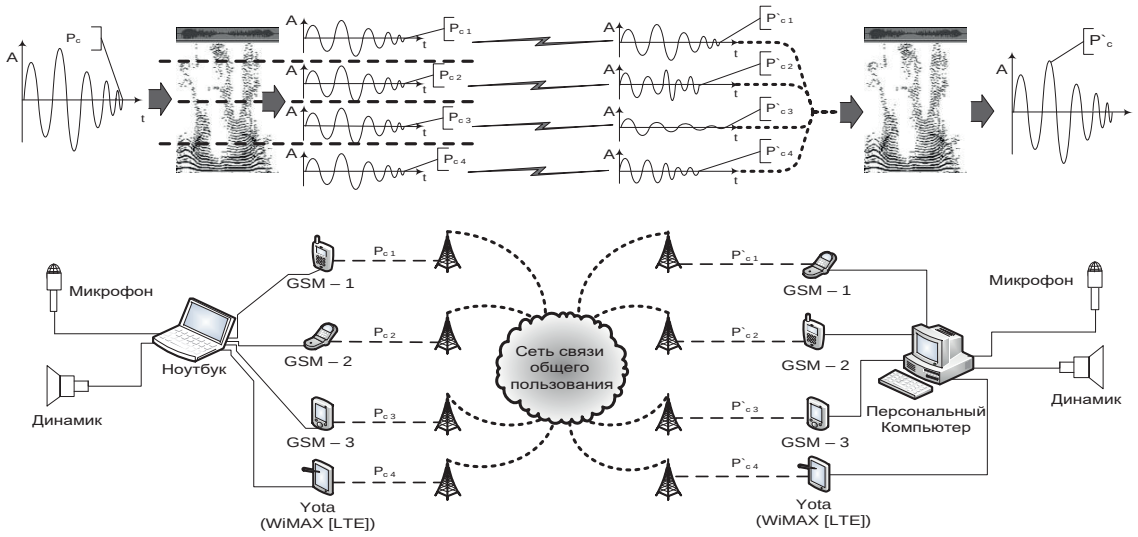

Рис. 2. Общая схема многоканальной системы защищенной речевой связи, маскированной разделением и расслоением

каналов, добавляя каналы новых операторов связи (например, фиксированную телефонную связь, другие сервисы VoIP и сотовой связи) и исключая «старых», предыдущих.

2. «Реконструкторы», восстанавливающие РР из принятой части РИ

Эти маскираторы используют только отделенную для передачи неразборчивую часть исходной речевой информации, по которой на приемном конце КГС с использованием заранее сформированного речевого корпуса диктора может быть реконструирована вся спектрограмма защищаемого сигнала и произведена ее инверсия по восстановлению его волновой формы и РР. Остаточная неразборчивая часть исходного РС, подлежащая передаче, может быть получена путем фильтрации РС или микшированием его с шумом или каким-то другим способом. Важно, чтобы осталось не менее 3-х гармоник на принятом материале, по которым потом будет возможна реконструкция спектрограммы и синтез клона исходного РС.

То есть после нахождения основного тона на полученном остаточном речевом материале необходимо восстановить гармоническую структуру речи, найти все гармоники с частотой, кратной частоте основного тона, с использованием следующего соотношения:

$$\omega_i = i \cdot \omega_{\text{осн}},\ i \in \left[\frac{\nu}{2 \cdot \omega_{\text{осн}}}\right], \qquad (3)$$

где ω_i – круговая частота i-ой гармоники.

Суть используемого алгоритма восстановления гармонической структуры РС[15] посредством поиска кратных основному тону гармоник на одном временном срезе сонограммы заключается в следующем:

а) каждый частотно-временной срез рассчитывается и анализируется независимо от других посредством кратковременного анализа Фурье;
б) на каждом временном слое определяется частота основного тона параболическим способом (частоту основного тона измеряем как количество (ν/N) Гц, ν – частота дискретизации звукового сигнала);
в) на временном срезе находится точка с частотой, наиболее близкой к частоте основного тона, которая помечается условным красным маркером;
г) на этом же временном слое находятся точки с частотой, наиболее близкой к удвоенной, утроенной,... и т.д. частоте основного тона; выбранные точки также помечаются;
д) амплитуда всех непомеченных точек полагается равной нулю.

Результат работы описанного выше алгоритма изображен на рис. 3, где представлены результаты параболической коррекции линий гармоник по вершинам парабол спектральных разверток. Как видно, после проведенной коррекции треки даже верхних гармоник (красные линии наверху) являются непрерывными и совпадают с точными исходными значениями на изображении узкополосной спектрограммы.

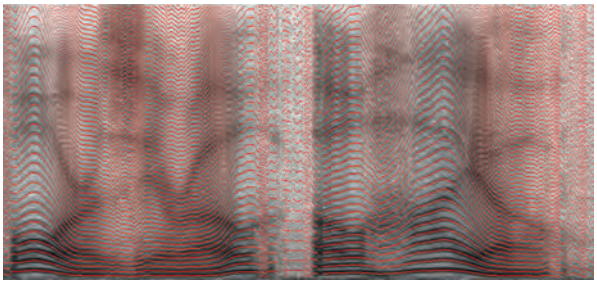

Рис. 3. Сонограмма и линии гармоник (красные треки) после коррекции частоты основного тона по вершинам парабол спектральных разверток.

15 Дворянкин С. В., Алюшин В. М. Метод реконструкции гармонической структуры спектральных описаний искаженной шумами и помехами речи. // Известия Института инженерной физики. 2013. № 2 (28). С. 57–62.

Дополнительным критерием проверки правильности нахождения основного тона являлась максимизация суммы амплитуд первых 7 кратных гармоник:

$$\sum_{K}^{7} |X[i \cdot x_в]| \longrightarrow max,$$

где $[a]$ – целая часть числа a.

Предложенный метод работает в случае зашумленного сигнала при условии, что треки только некоторых первых гармоник «видны» на фоне шумов (рис. 4).

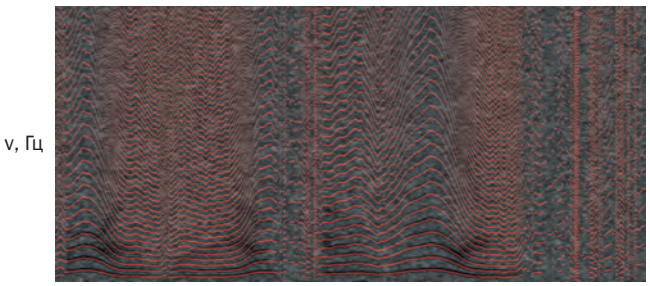

Рис. 4. Сонограмма и линии гармоник зашумленного речевого сообщения

Как показали эксперименты, при зашумлении речевого сообщения белым шумом треки гармоник и восстановленная по ним гармоническая структура находятся корректно даже при отношении сигнал/шум до –12 Дб. При зашумлении речевого сообщения естественными помехами, гармоническая структура может корректно восстанавливаться при отношении сигнал/шум до –8 ... –5 Дб.

Оставшаяся в шумах часть РС неразборчива, но РР может быть восстановлена по реконструированной гармонической структуре и соответствующей ей формантной из речевого корпуса диктора.

Разработанный метод и алгоритм нахождения основного тона вокализованных участков РС позволяет восстанавливать гармоническую структуру сигнала (рис. 6) даже в случае, если часть спектральных описаний сигнала, содержащих линию основного тона и несколько верхних и нижних гармоник (рис. 4, 5), были «утеряны», зашумлены.

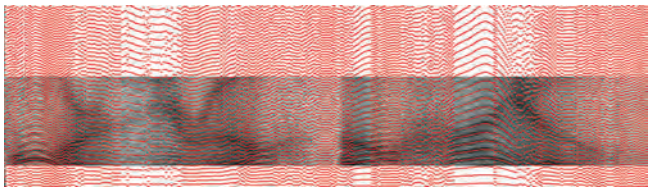

Рис. 5. Гармоническая структуры РС, восстановленная по спектральным описаниям с частичной потерей информации.

3. «Заместители» с заменой (подменой) опорных элементов речи

Заместители это маскираторы с подменой исходных автоматически распознающихся фонем (или других элементов речевого потока) на фонемы иного несуществующего языка из речевой базы данных виртуального диктора и обратно.

Рассмотрим один из таких методов обеспечения конфиденциальности речи на основе маскировки звука и материала известного речевого корпуса [6]. Структурная схема предлагаемого метода приведена на рис. 6.

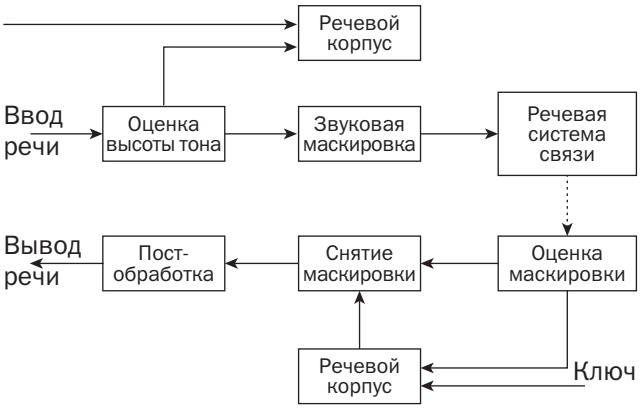

Рис. 6. Структурная схема метода обеспечения конфиденциальности речи на основе маскировки звука и ресурса корпуса речи

Здесь входная речь сначала сегментируется на кадры, и для каждого кадра выполняется оценка высоты тона по описанному выше алгоритму[16], которая выступает как часть идентификатора для нахождения нужных кадров из речевого корпуса.

Другой частью идентификатора является секретный ключ, который раздается каждому пользователю. В целях безопасности секретные ключи могут меняться через определенные промежутки времени или каждый раз. Помимо ключей, сам корпус также может быть заменен по расписанию для обеспечения безопасности связи.

Каждый кадр входной речи маскируется соответствующими кадрами в корпусе. Для каждого целевого кадра речи выбирается один или несколько кадров речи из корпуса, причем выбранные кадры должны иметь ту же высоту тона, что и входной кадр. Чем больше кадров корпуса используется для маскировки каждого целевого речевого кадра, тем лучшего эффекта маскировки можно достичь несмотря на то, что сам алгоритм маскировки будет сложнее[17]. Чтобы высота тона не менялась, маскирование ограничивается линейными аддитивными операциями. Затем маскированная речь вводится в коммуникационные системы вместо оригинальной речи.

16 Дворянкин С. В., Алюшин В. М. Метод реконструкции гармонической структуры спектральных описаний искаженной шумами и помехами речи. // Известия Института инженерной физики. 2013. № 2 (28). С. 57–62.

17 Алдошина И. А. Лекции по психоакустике. / Архив журнала «Звукорежиссер» : 1999-2002. URL: https://prazdnikson.ru/i-aldoshina-lektsii-po-psihoakustike (дата обращения: 16.06.2024).

На принимающей стороне эффект маскировки снимается с принимаемой замаскированной речи также в соответствии с индексом высоты тона. Принимающая сторона должна иметь тот же корпус и секретный ключ, как и передающая сторона, чтобы из корпуса можно было выбрать те же маскирующие кадры и затем правильно восстановить речь в кадре с помощью обратного алгоритма демаскировки звука.

Для успешного восстановления речи алгоритм должен быть спроектирован с определенной допуском к погрешности высоты тона на случай, если из-за ошибок канала связи высота тона может несколько отличаться до и после передачи.

Целью постобработки является снижение влияния погрешности высоты тона и улучшение качества восстановленной речи. Так, простое удаление неправильно принятого кадра не принесет очевидного ущерба качеству речи. Для лучшего качества речи удаленный речевой кадр также может быть скомпонован из данных соседних кадров и восстановлен.

В этом методе неразборчивость замаскированной речи поможет сохранить конфиденциальность коммуникации даже в том случае, если передаваемая речь была перехвачена ЗЛ, а секретность речевого корпуса и секретные ключи определяют стойкость к взлому системы. Таким образом, можно обеспечить высокий уровень безопасности речевой связи.

В данном методе, как и в предыдущих, звуковая маскировка выполняется перед отправкой речи в коммуникационную систему, а восстановление речи выполняется на ее выходе. Никаких модификаций коммуникационной системы не требуется. Тогда сквозная защита конфиденциальности речи может быть реализована с небольшими затратами.

Даже если в системах речевой связи используется сжатие с потерями, предложенный метод работоспособен. Результаты экспериментов показали, что предложенный метод достигает хорошей производительности, когда в системах связи используются алгоритмы кодирования формы речевой волны[18].

4. «Трансляторы» – переводчики речи на незнакомый язык

Частный случай маскираторов заместителей, активно развивающийся. В работе [7] представлена нейронная сеть, которая может напрямую переводить речь с одного языка на другой, не опираясь на промежуточное текстовое представление. Сеть проходит сквозное обучение и учится сопоставлять спектрограммы речи с целевыми спектрограммами на другом языке, соответствующими переведенному контенту в другом каноническом голосе. Далее так переведенная речь синтезируется по новой спектрограмме, используя голос диктора-источника, носителя неизвестного языка.

Эксперименты на двух наборах данных для перевода речи с испанского на английский напрямую, без опоры на промежуточное текстовое представление. показали, что предложенная модель незначительно уступает базовому каскаду из модели перевода речи в текст и модели синтеза текста в речь, что свидетельствует о применимости подхода для решения сложной задачи защиты конфиденциальности речевых коммуникаций в реальном времени.

5. «Смесители» – устройства информационного маскирования

Маскирование, возникающее, когда целевая речь сопровождается одним мешающим фактором – голосом или несколькими голосами и звуками, часто называют информационным (ИМ). Как правило, ИМ больше, когда интерферирующий голос громок и разборчив, чем когда он тих и неразборчив (например, речь на незнакомом языке). Но относительный вклад акустико-фонетической и лингвистической интерференции часто трудно оценить из-за акустических различий между интерферирующими сторонами (например, разными собеседниками) [8].

Тем не менее, ИМ находит свое распространение в системах активной защиты ЗП, при формировании речеподобных помех, адаптивными защищаемому РС [2]. Правда здесь не нужен процесс демаскирования.

В КГС удобно использовать независимые голосовые помехи, в качестве которых может использоваться звуковое вещание нескольких (не менее трех) новостных программ ведущимися разными дикторами.

Получаемая таким образом помеховая смесь формируется одновременно на всех концах защищаемой системы голосовой связи. Количество радиостанций, частоты их вещания, тип дикторов и др. независимые параметры определяют секретные ключи для организации процессов маскирования и демаскирования без их распределения. Содержимое таких ключей может неоднократно меняться в процессе речевого обмена.

Важно, чтобы помехи были подобраны таким образом, чтобы процесс демаскирования на приёмном конце КГС сводился к упрощенным компенсации и коррекции её следов на основе образного анализа изображения зашумленной спектрограммы [1].

6. «Маркеры и подложки» – встроенные в сигналы и сообщения спектрограммы

Могут выступать в качестве речевой подписи, цифровых водяных знаков для подтверждения подлинности конфиденциальных данных и-или скрытной передачи речи. Признаки исходной защищаемой речи в явном виде отсутствуют.

Используются в популярных носителях информации разных видов как встроенные в них изображения

18 Ding Qi, Nan Longmei, Xu jinfu. A Speech Privacy Protection Method Based on Sound Masking and Speech Corpus. 8th International Congress of Information and Communication Technology (ICICT – 2018).

полутоновых и бинарных сонограмм (речевой подписи), не влияющими на качество передаваемых–сохраняемых носителями данных, с последующей инверсией спектрограмм и реконструкцией по ним РС по запросу [3].

С помощью скрытно и–или открыто представляемой речевой подписи на информносителях можно не только передавать–сохранять конфиденциальную информацию, но и обеспечивать ее аутентичность, формируя сигналы со спектром идентичным биопризнаку пользователя.

Заключение

Рассмотрена классификация существующих маскираторов и устройств технического закрытия речи нового типа. Выработаны требования к перспективным устройствам маскирования, из которых особо отмечены асинхронность, гибридность, скорость и гибкость работы, отсутствие признаков открытой исходной речи.

Для оценки уровня защищенности маскируемого канала голосовой связи предлагается использовать показатель словесной разборчивости с нормами и критериями, принятыми для защищаемых помещений конфиденциальных переговоров.

За основу разрабатываемых компьютерных технологий обеспечения безопасности (приема, передачи и хранения) конфиденциальных речевых сообщений через управление их разборчивостью можно принять технологию образного анализа РС, заключающаяся в переходе, посредством кратковременного преобразования Фурье, от волнового представления РС к изображению динамических узкополосных спектрограмм – графических образов (ГО), их обработке методами цифровой обработки изображений, распознавания образов и методов машинного обученияИ для решения поставленных прикладных задач и обратном переходе (синтезе) от нового изображения, посредством инверсии спектрограмм, к новой волновой форме.

Новые виды речевых маскираторов, разрабатываемых на основе предложенной технологии образного анализа, будут обладать рядом неоспоримых преимуществ по сравнению с аналогами: достаточно невысокой стоимостью, относительно высокой стойкостью, максимальной оперативностью, отсутствием остаточной разборчивости, повышенным качеством восстановленного сигнала, устойчивой работой на каналах среднего и низкого качества.

Литература

1. *Хорев А. А., Дворянкин С. В., Козлачков С. Б., Василевская Н. В. Анализ предельных возможностей методов шумопонижения и реконструкции речевых сигналов, маскируемых различными типами помех //Вопросы кибербезопасности. 2024. № 1 (59). С. 89–100.*
2. *Дворянкин С. В., Дворянкин Н. С., Устинов Р. А. Речеподобная помеха, стойкая к шумоочистке, как результат скремблирования защищаемой речи // Вопросы кибербезопасности. 2022. № 5 (51). С. 14–27.*
3. *Дворянкин С. В., Дворянкин Н. С., Устинов Р. А. Развитие технологий образного анализа-синтеза акустической (речевой) информации в системах управления, безопасности и связи // Безопасность информационных технологий =IT Security. Том 26, № 1. 2019. С. 64–76. DOI: http://dx.doi.org/10.26583/bit.2019.1.07*
4. *Голиков А. М. Исследование методов аналогового скремблирования: Учебно-методическое пособие по лабораторной работе [Электронный ресурс] / А. М. Голиков. – Томск: ТУСУР, 2019. – 25 с.*
5. *Столбов М. Б. Основы анализа и обработки речевых сигналов / М. Б. Столбов – СПб.: НИУ ИТМО, 2021. – 101 с.*
6. *Tom Backstrom. Privacy in Speech Technology. arXiv:2305.05227v1 [eess.AS] 9 May 2023/*
7. *Ye Jia, Ron J. Weiss, Fadi Biadsy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu Direct speech-to-speech translation with a sequence-to-sequence model. arXiv:1904.06037v1 [cs.CL] 12 Apr 2019.*
8. *Robert J. Summers, Brian Roberts. Informational masking of speech by acoustically similar intelligible and unintelligible interferers. The Journal of the Acoustical Society of America 147(2):1113-1125. February 2020. DOI:10.1121/10.0000688*
9. *Jennifer Williams, Karla Pizzi, Paul-Gauthier Noé , Sneha Das. Exploratory Evaluation of Speech Content Masking. arXiv:2401.03936v1 [eess.AS] 8 Jan 2024.*
10. *Sonia Yasmin, Vanessa C. Irsik, Ingrid S. Johnsrude, Björn Herrmann. The Effects of Speech Masking on Neural Tracking of Acoustic and Semantic Features of Natural Speech. Neuropsychologia doi: 10.1016/j.neuropsychologia.2023.108584. doi:https://doi.org/10.1101/2023.02.10.527537.*
11. *Y. Chen, Y. Assael, B. Shillingford, D. Budden, S. Reed, H. Zen, Q. Wang, L. C. Cobo, A. Trask, B. Laurie et al., «Sample efficient adaptive text-to-speech», in Proc. ICLR, 2019.*
12. *Y. Jia, M. Johnson, W. Macherey, R. J. Weiss, Y. Cao, C. -C. Chiu, N. Ari et al., «Leveraging weakly supervised data to improve endto-end speech-to-text translation», in Proc. ICASSP, 2019.*
13. *A. Haque, M. Guo, and P. Verma, «Conditional end-to-end audio transforms», in Proc. Interspeech, 2018. [23] J. Zhang, Z. Ling, L. -J. Liu, Y. Jiang, and L. -R. Dai, «Sequenceto-sequence acoustic modeling for voice conversion», IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019.*
14. *F. Biadsy, R. J. Weiss, P. J. Moreno, D. Kanevsky, and Y. Jia, «Parrotron: An end-to-end speech-to-speech conversion model and its applications to hearing-impaired speech and speech separation», arXiv:1904.04169, 2019.*
15. *J. Shen, P. Nguyen, Y. Wu, Z. Chen et al., «Lingvo: a modular and scalable framework for sequence-to-sequence modeling», 2019.*
16. *K. Irie, R. Prabhavalkar, A. Kannan, A. Bruguier, D. Rybach, and P. Nguyen, «Model unit exploration for sequence-to-sequence speech recognition», arXiv:1902.01955, 2019.*
17. *W. -N. Hsu, Y. Zhang, R. J. Weiss, H. Zen, Y. Wu, Y. Wang, Y. Cao, Y. Jia, Z. Chen, J. Shen et al., «Hierarchical generative modeling for controllable speech synthesis», in Proc. ICLR, 2019.*

EVOLUTION AND DIRECTIONS OF DEVELOPMENT OF TECHNOLOGIES FOR MASKING CONFIDENTIAL SPEECH MESSAGES

Durakovskiy A. P.[19], Dvoryankin S. V.[20], Dvoryankin N. S.[21]

Purpose of the research: *analysis of methods and algorithms of technical closure of speech information in networks and systems of voice communication, evaluation of directions and prospects of development of speech masking technologies with machine learning.*

Research methods: *applied systems analysis, digital spectral-time analysis, digital signal and image processing, image analysis of spectrograms, machine learning*

Research results: *the problems of ensuring the security of confidential voice communication in modern conditions are outlined. The review of speech protection methods used in practice in public voice communication channels is given. Traditional and perspective algorithms of masking of speech messages, methods of their realization are considered. The advantages of the latter over the ones are noted.*

Science significance: *New methods of technical speech masking based on modification and reconstruction of dynamic spectrogram images using artificial intelligence are proposed*

Practical: *effective speech masking algorithms are proposed. The obtained results will allow to expand the possibilities of existing solutions for protection of speech information in voice communication systems and networks and to design more effective ones based on the described approaches.*

Keywords: *information security, speech information protection, image analysis-synthesis, technical speech closure, speech-like signal, machine learning.*

References

1. *Horev A. A., Dvoryankin S. V., Kozlachkov S. B., Vasilevskaya N. V. Analiz predel'nyh vozmozhnostej metodov shumoponizheniya i rekonstrukcii rechevyh signalov, maskiruemyh razlichnymi tipami pomekh. //Voprosy kiberbezopasnosti. 2024. № 1 (59). S. 89–100.*
2. *Dvoryankin S. V., Dvoryankin N. S., Ustinov R. A. Rechepodobnaya pomekha, stojkaya k shumoochistke, kak rezul'tat skremblirovaniya zashchishchaemoj rechi. // Voprosy kiberbezopasnosti. 2022. № 5 (51). S. 14–27.*
3. *Dvoryankin S. V., Dvoryankin N. S., Ustinov R. A. Razvitie tekhnologij obraznogo analiza-sinteza akusticheskoj (rechevoj) informacii v sistemah upravleniya, bezopasnosti i svyazi // Bezopasnost' informacionnyh tekhnologij =IT Security. Tom 26, № 1. 2019. C. 64–76. DOI: http://dx.doi.org/10.26583/bit.2019.1.07*
4. *Golikov A. M. Issledovanie metodov analogovogo skremblirovaniya: Uchebno-metodicheskoe posobie po laboratornoj rabote [Elektronnyj resurs] / A. M. Golikov. – Tomsk: TUSUR, 2019. – 25 s.*
5. *Stolbov M. B. Osnovy analiza i obrabotki rechevyh signalov / M. B. Stolbov – SPb.: NIU ITMO, 2021. – 101 s.*
6. *Tom Backstrom. Privacy in Speech Technology. arXiv:2305.05227v1 [eess.AS] 9 May 2023/*
7. *Ye Jia, Ron J. Weiss, Fadi Biadsy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu Direct speech-to-speech translation with a sequence-to-sequence model. arXiv:1904.06037v1 [cs.CL] 12 Apr 2019.*
8. *Robert J. Summers, Brian Roberts. Informational masking of speech by acoustically similar intelligible and unintelligible interferers. The Journal of the Acoustical Society of America 147(2):1113-1125. February 2020. DOI:10.1121/10.0000688*
9. *Jennifer Williams, Karla Pizzi, Paul-Gauthier Noé , Sneha Das. Exploratory Evaluation of Speech Content Masking. arXiv:2401.03936v1 [eess.AS] 8 Jan 2024.*
10. *Sonia Yasmin, Vanessa C. Irsik, Ingrid S. Johnsrude, Björn Herrmann. The Effects of Speech Masking on Neural Tracking of Acoustic and Semantic Features of Natural Speech. Neuropsychologia doi: 10.1016/j.neuropsychologia.2023.108584. doi:https://doi.org/10.1101/2023.02.10.527537.*
11. *Y. Chen, Y. Assael, B. Shillingford, D. Budden, S. Reed, H. Zen, Q. Wang, L. C. Cobo, A. Trask, B. Laurie et al., «Sample efficient adaptive text-to-speech», in Proc. ICLR, 2019.*
12. *Y. Jia, M. Johnson, W. Macherey, R. J. Weiss, Y. Cao, C. -C. Chiu, N. Ari et al., «Leveraging weakly supervised data to improve endto-end speech-to-text translation», in Proc. ICASSP, 2019.*
13. *A. Haque, M. Guo, and P. Verma, «Conditional end-to-end audio transforms», in Proc. Interspeech, 2018. [23] J. Zhang, Z. Ling, L. -J. Liu, Y. Jiang, and L. -R. Dai, «Sequenceto-sequence acoustic modeling for voice conversion», IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019.*
14. *F. Biadsy, R. J. Weiss, P. J. Moreno, D. Kanevsky, and Y. Jia, «Parrotron: An end-to-end speech-to-speech conversion model and its applications to hearing-impaired speech and speech separation», arXiv:1904.04169, 2019.*
15. *J. Shen, P. Nguyen, Y. Wu, Z. Chen et al., «Lingvo: a modular and scalable framework for sequence-to-sequence modeling», 2019.*
16. *K. Irie, R. Prabhavalkar, A. Kannan, A. Bruguier, D. Rybach, and P. Nguyen, «Model unit exploration for sequence-to-sequence speech recognition», arXiv:1902.01955, 2019.*
17. *W. -N. Hsu, Y. Zhang, R. J. Weiss, H. Zen, Y. Wu, Y. Wang, Y. Cao, Y. Jia, Z. Chen, J. Shen et al., «Hierarchical generative modeling for controllable speech synthesis», in Proc. ICLR, 2019.*

19 Anatoly P. Durakovskiy, Ph.D. (in Tech.), Associate Professor of the Department of Strategic Information Studies of MEPhI, Director of the Attestation and Testing Centre for Information Security and Information Protection Systems of MEPhI. Moscow. Russia. E-mail: apdurakovskiy@mephi.ru

20 Sergey V. Dvoryankin, Dr. Sc. (of Tech.), Professor, Professor of the Department of Strategic Information Studies, National Research Nuclear University MEPhI, Head of the Laboratory for the Protection and Processing of Audiovisual Information, Moscow State Linguistic University. Moscow. Russia. E-mail: svdvoryankin@mephi.ru, https://orcid.org/0000-0001-6908-0676

21 Nikita S. Dvoryankin, postgraduate student, National Research Nuclear University MEPhI. Moscow. Russia. E-mail: nik.dvrn@gmail.com