Сергей В. Дворянкин[1], Сергей В. Уленгов[2], Роман А. Устинов[3],
Никита С. Дворянкин[4], Антон О. Антипенко[5]
[1, 2, 3, 5]*Финансовый университет при Правительстве Российской Федерации,*
Ленинградский проспект, 49, Москва, 125993, Россия
[4]*Национальный исследовательский ядерный университет «МИФИ»,*
Каширское шоссе, 31, Москва, 115409, Россия
[1]*e-mail: SVDvoryankin@fa.ru, https://orcid.org/0000-0001-6908-0676*
[2]*e-mail: mesetis@mail.ru, https://orcid.org/0000-0002-6888-4303*
[3]*e-mail: public-ura@yandex.ru, https://orcid.org/0000-0002-8454-9951*
[4]*e-mail: nik.dvrn@gmail.com, https://orcid.org/0000-0002-1580-7179*
[5]*e-mail: an-go-55@yandex.ru, https://orcid.org/0000-0001-8692-6637*

СИСТЕМНОЕ МОДЕЛИРОВАНИЕ РЕЧЕПОДОБНЫХ СИГНАЛОВ И ЕГО ПРИМЕНЕНИЕ В СФЕРЕ БЕЗОПАСНОСТИ, СВЯЗИ И УПРАВЛЕНИЯ

DOI: http://dx.doi.org/10.26583/bit.2019.4.08

Аннотация. Речеподобные сигналы (РПС) наряду с оригинальной речью всё активнее применяются в системах безопасности, связи и управления, всё чаще замещая в них последнюю. **Цель** работы состоит в уточнении инструментария моделирования для формирования речеподобных сигналов с требуемыми характеристиками при построении эффективных систем обработки и защиты акустической (речевой) информации. **Методы.** Рассмотрены методы формирования речеподобных сигналов по заданным изображениям их спектрограмм с раздельной или совместной обработкой их гармонической и формантной структур, соответствующей выбранной для конкретного приложения системной модели РПС. **Результаты.** Предложено в различных приложениях рассматривать РПС в виде комплекса реальных моделей прямого, косвенного и условного подобия системного отображения оригинальной речи. **Выводы**. Рассмотрены основные области применения системных моделей РПС. Оценены возможности и перспективы использования РПС для передачи в каналах речевых коммуникаций графической информации, в том числе о биометрических признаках легитимных удаленных пользователей защищённых информационных ресурсов.

Ключевые слова: речеподобные сигналы, безопасность, модель, биометрия.
Для цитирования: ДВОРЯНКИН, Сергей В. et al. СИСТЕМНОЕ МОДЕЛИРОВАНИЕ РЕЧЕПОДОБНЫХ СИГНАЛОВ И ЕГО ПРИМЕНЕНИЕ В СФЕРЕ БЕЗОПАСНОСТИ, СВЯЗИ И УПРАВЛЕНИЯ. Безопасность информационных технологий, [S.l.], v. 26, n. 4, p. 101–119, 2019. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/1236>. Дата доступа: 05 dec. 2019. doi: http://dx.doi.org/10.26583/bit.2019.4.08.

Sergey V. Dvoryankin[1], Sergei V. Ulengov[2], Roman A. Ustinov[3],
Nikita S. Dvoryankin[4], Anton O. Antipenko[5]
[1, 2, 3, 5]*Financial University under the Government of the Russian Federation,*
Leningradsky Prospekt, 49, Moscow, 125993, Russia
[4]*National Nuclear Research University MEPHI (Moscow Engineering Physics Institute),*
Kashirskoe shosse, 31, Moscow, 115409, Russia
[1]*e-mail: SVDvoryankin@fa.ru, https://orcid.org/0000-0001-6908-0676*
[2]*e-mail: mesetis@mail.ru, https://orcid.org/0000-0002-6888-4303*
[3]*e-mail: public-ura@yandex.ru, https://orcid.org/0000-0002-8454-9951*
[4]*e-mail: nik.dvrn@gmail.com, https://orcid.org/0000-0002-1580-7179*
[5]*e-mail: an-go-55@yandex.ru, https://orcid.org/0000-0001-8692-6637*

Speech-like signal system modeling and its application in the field of security, communication and control access

DOI: http://dx.doi.org/10.26583/bit.2019.4.08

Abstract. Speech-like signals (SLS), along with the original speech, are actively used in security, communication and control systems, increasingly replacing the latter one in them. **The aim** of the work is to refine the modeling tools for the formation of speech-like signals with specified characteristics in the construction of effective systems for processing and protection of acoustic (speech) information. **Methods.** Methods of formation of speech-like signals on the given images of their spectrograms with separate or joint processing of their harmonic and formant structures corresponding to the system model of SLS chosen for a specific application are considered. **Results.** It is proposed in various applications to consider SLS as a complex of real models of direct, indirect and conditional similarity of the system display of the original speech. **Summary.** The main areas of application of such system models of SLS are considered. The possibilities and prospects of using SLS for transmission of graphic information in speech communication channels, including biometric characteristics of legitimate remote users of protected information resources, are estimated.

Keywords: speech-like signals, security, model, biometrics.

For citation: DVORYANKIN, Sergey V. et al. Speech-like signal system modeling and its application in the field of security, communication and control access. IT Security (Russia), [S.l.], v. 26, n. 4, p. 101–119, 2019. ISSN 2074-7136. Available at: <https://bit.mephi.ru/index.php/bit/article/view/1235>. Date accessed: 05 dec. 2019. doi: http://dx.doi.org/10.26583/bit.2019.4.08.

Введение

В публикациях по речевым технологиям, их применении в приложениях безопасности, управления и связи всё чаще встречается термин «речеподобные сигналы» (РПС). В англоязычной литературе также распространено довольно устойчивое выражение «speech-like signal» переводимое буквально как речеподобный сигнал [1].

В последние годы РПС получили существенное развитие и во многих областях используются наряду или вместо речевых сигналов. Так в сфере защиты речевой информации (РИ) сегодня активно используются речеподобные помехи, голосовые Интернет-ассистенты общаются с пользователями сети голосами известных дикторов, качество новых систем речевой связи тестируется РПС с заданными характеристиками. Использование РПС повышает эффективность традиционных речевых приложений, позволяя экономить время, инвестиции и другие ресурсы.

Однако до сих пор нет устоявшегося определения содержания термина речеподобного сигнала. Для его корректного определения и применения целесообразно с позиций системного подхода рассмотреть РПС как некую модель или даже комплекс реальных моделей системного отображения оригинала – исходного речевого сигнала (РС).

1. Основные характеристики речи и требования к речеподобным сигналам

Общеизвестно, что речь – это исторически сложившаяся форма общения людей посредством языковых конструкций, создаваемых на основе определенных правил. В общении людей с помощью устной речи физическим посредником между ними является акустический РС, распространяемый посредством акустических волн. Основное его отличие от других видов сигналов состоит в том, что он несет информацию, которая может быть извлечена путем его прямого прослушивания. Важно отметить, что слуховая система человека настроенная на восприятие именно такого сигнала, отличается избирательностью к речи и помехоустойчивостью по отношению к мешающим акустическим воздействиям.

Как правило, речь характеризуется четырьмя группами характеристик:

– *семантическая* или смысловая сторона речи характеризует смысл тех понятий, которые передаются при помощи речевого сообщения;

– *фонетические* характеристики речи: данные, характеризующие речь с точки зрения её звукового состава;

– *интонационные* характеристики, которые в значительной степени определяют стиль речи, ее эмоциональную окраску и дают возможность различать незнакомые языки;

– *физические* характеристики: величины и зависимости, характеризующие речь как акустический сигнал.

В ряде приложений крайне желательно, чтобы все основные характеристики указанных групп также присутствовали и в РПС. Иногда достаточно использовать только частотный диапазон, совпадающий с речевым, и возможность передачи данного вида РПС по существующим каналам речевой связи. Нередко, при генерации и обработке РПС, как и РС, следует учитывать и субъективные свойства слуха: предмаскировку, постмаскировку, нелинейность, пороги слышимости и динамический диапазон [2, 3].

В то же время, в большинстве практических случаев в сфере безопасности и управления необходимо, чтобы РПС с РС были бы только сравнимы по ряду основных акустических характеристик (физическим параметрам) как-то: мощность, звуковое давление, отношение сигнал-шум, полоса частот и форма частотного спектра, период (частота) основного тона, длительность звучания, пауз и др.

Кроме того, необходимо учитывать, что, помимо акустической среды распространения РС, сегодня создана и функционирует информационно-телекоммуникационная (ИТК) среда (инфраструктура), предназначенная для передачи-приема по различным каналам связи вместе с другими видами данных и собственно аудиоданных, с речевой и звуковой информацией, содержащейся в них. Более того, так сложилось, что все существующие ныне каналы телекоммуникаций в своей основе и развитии опирались и отталкивались от телефонных каналов приема-передачи речевой информации, исторически созданными для передачи речевых сообщений в первую очередь.

Первичными источниками акустических сигналов являются механические колебательные системы, например органы речи человека, а вторичными – преобразователи различного типа, например, громкоговорители, озвучивающие сигнал с микрофона, фонограммы. Частоты акустических колебаний в пределах *20 – 20000 Гц* называют звуковыми (их может воспринимать человеческое ухо), ниже *20 Гц* – инфразвуковыми, а выше *20000 Гц* – ультразвуковыми. Их слуховое восприятие человека не воспринимает [4].

Понятно, что полоса частот РПС не должна превышать указанный диапазон *20 – 20000 Гц.* Более того, крайне желательно, чтобы полоса частот РПС была приближена к полосе частот реального РС (*60 – 10000 Гц*), чтобы РПС мог как и речь передаваться по речевым каналам связи, подвергаясь в них всем трансформациям и преобразованиям, как и РС, без потери и искажения своих основных спектрально-временных характеристик.

В зависимости от формы акустических колебаний различают простые (тональные) и сложные сигналы. Тональный – это сигнал, вызываемый колебанием, совершающимся по синусоидальному закону. Сложный сигнал может содержать целый спектр гармонических или узкополосных составляющих. Речь и РПС можно отнести к группе сложных акустических сигналов, представляющих собой, на коротких до *30 мс* интервалах наблюдения, суперпозицию узкополосных кратковременных синусоидальных сигналов – вейвлетов [3].

В процессе произношения голосовые связки, то сжимая, то открывая голосовую щель, пропускают воздух импульсами, частота которых лежит в пределах от *60* до *180 Гц*

у мужчин и от *160* до *300 Гц* у женщин. Средняя частота основного тона (ЧОТ) составляет для мужских голосов *130 Гц* и *250 Гц* для женских [4].

Изменяя положение языка, зубов и губ человек имеет возможность изменять характер звучания и произносить различные гласные, вокализованные звуки, определяющие гармоническую структуру спектра речи через динамику и рисунок движения линий гармоник основного тона (ЛГОТ). Согласные звуки (шумовые и смешанные шумо-вокализованные) в большинстве случаев произносятся без участия голосовых связок.

Проходя через систему резонаторов, одни гармонические составляющие получают усиление, а другие, наоборот, подавляются. Усиленные системой резонаторов области частот называются формантными областями или просто *формантами* [4], а подавленные – антиформантами.

Поскольку форманты значительно мощнее других составляющих речи, поэтому они главным образом и воздействуют на ухо слушающего, формируя звучание того или иного звука. Некоторое влияние на этот процесс оказывают и антиформанты. Вместе они формируют формантную структуру спектра речи.

Таким образом, речь содержит в себе форманты, восприятие которых определяет ее разборчивость, и неформантные составляющие, к которым относятся частота (период) основного тона, области частот между формантами (антиформанты) и др., зависящие от индивидуальных особенностей говорящих [4].

В этой связи, для многих приложений желательно, чтобы спектральные составы речи и РПС коррелировались. Большая часть формант звуков речи лежит в пределах от *300* до *3400 Гц* (полоса частот телефонного канала), что и позволяет считать эту полосу частот вполне достаточной для обеспечения хорошей понятности и разборчивости речи (РР). Форманты здесь расположены не только вплотную друг к другу, но даже с некоторым перекрытием.

Спектральный состав сигналов русской речи, характеризуемый статистическими характеристиками ЧОТ ($F_{ОТ}$) и основных формантных частот (F_1–F_4), в основном присутствующими в полосе частот телефонного канала, в соответствии с [1] и др. представлен в табл. 1.

Таблица 1. Спектральный состав речевых сигналов для русского языка

$F_{ОТ}$	F_1	F_2	F_3	F_4
60–300 Гц	150–850 Гц	500–2500 Гц	1500–3500 Гц	2500–4500 Гц

Приведенные в табл. 1 диапазоны формантных частот F_1–F_4 относятся к сонорным неносовым звукам речи при произнесении их средним мужским голосом [1]. Частота основного тона $F_{ОТ}$, указана во всём диапазоне всех возможных мужских, женских и детских голосов.

Различные звуки на наблюдаемых участках спектра имеют разное число формант: гласные – до четырех формант, глухие согласные до *5 – 6* формант [4]. Первые две форманты, определяющие в основном смысловое содержание речевого сообщения, что обусловлено участием в образовании этих звуков основных резонаторов голосового аппарата – полости глотки и носоглотки, называются основными, остальные – вспомогательными.

Можно считать, что характер движения по спектру в координатах частота-время основных формант (формантная структура речевого спектра) определяет семантическую составляющую произносимого звука речи, а рисунок вспомогательных формант, вместе с

изменяемыми по спектру линиями гармоник основного тона (гармоническая структура речевого спектра), характеризуют индивидуальную для каждого человека окраску, тембр речи.

Исключение из передачи или добавление в нее формантных областей вызывает искажение передаваемого звука: либо превращение его в другой звук, либо вообще потерю им признаков звука человеческой речи.

Изменение рисунка гармонической структуры спектра речи связывается с наличием и изменением индивидуальных особенностей звучания речи диктора.

Частоты формант, их число и взаимное расположение для одних и тех же звуков, произносимых разными дикторами, могут сильно различаться. Однако для каждого звука речи характерно определенное положение формантных областей, и при наличии достаточного опыта по изображениям спектрограмм можно читать произнесенный «буштобированный» текст [4].

Перечисленные основные физические характеристики РС иногда желательно, а иногда и крайне необходимо воспроизводить при формировании РПС. Однако, при решении некоторых прикладных задач для упрощения алгоритма формирования РПС допустимо ограничивать перечень требований по степени подобия РПС реальным РС. Во избежание разночтений рекомендуется указывать по каким характеристикам формируемые РПС будут соответствовать РС [1].

Исходя из вышеизложенного, речевой сигнал (РС), как процесс, развивающийся во времени и по частоте, можно рассматривать в виде результата взаимного наложения друг на друга его гармонической (узкополосной) и формантной спектрально-временных структур (см. рис. 1).

На спектрограммах рис. 1 в координатах время-частота уровнем серого цвета указаны мощность сигнала в конкретной точке частотно-временного описания. Самый мощный черный, самый слабый – белый.

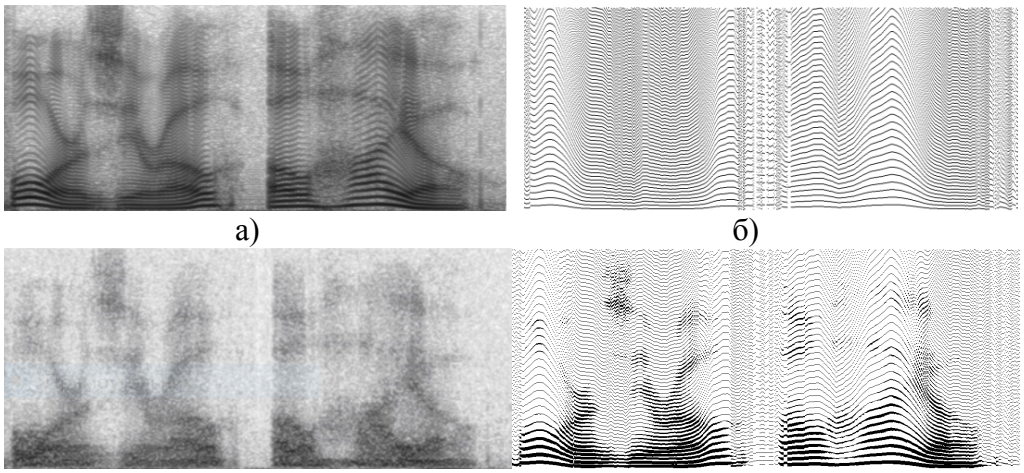


Рис. 1. Спектрограмма исходного РС (а), ее гармоническая (б) и формантная (в) структуры, а также результат совмещения последних – бинарная речевая подпись (г)
(Fig 1. The spectrogram of the original RS (a), its harmonic (б) and formant (в) structures, as well as the result of combining the latter – a binary speech signature (г))

Как видно из рис. 1, расслоение спектрограммы исходного РС (а) на гармоническую (б) и формантную (в) составляющие с последующим их объединением-наложением (г) дает результат спектрального образа практически не отличимый от исходного, что позволяет при формировании РПС для различных приложений вычленять

и обрабатывать отдельные спектральные компоненты РС – оригинала с последующим их агрегированием и синтезом по новому образу спектрограммы, соответствующей ему волновой формы РПС.

Как уже отмечалось, интонационные и психоэмоциональные особенности звучащей речи связаны с гармонической (узкополосной) структурой спектра речи (см рис. 1. б), на которой отражаются динамика и рисунок ЛГОТ. Смысловое содержание речевого сообщения, как основной объект защиты от несанкционированного доступа (НСД) в системах безопасности, управления и связи, определяется динамикой перестройки формантной структуры или огибающей спектра (см. рис. 1. в). Поэтому процессы семантического и интонационного речеобразования и речепреобразования, связанные с компонентами и динамикой спектральной огибающей, называемой фонетической функцией Пирогова, удобно исследовать посредством цифрового спектрально-временного анализа изображений спектрограмм [3, 4].

2. Определение системных моделей РПС в рамках культуры речевых коммуникаций и существующей инфокоммуникационной среды

С учетом изложенного, от речеподобного сигнала в указанных в теме статьи приложениях целесообразно потребовать, не столько акустического сходства или в какой-то мере похожести звучания с оригинальным РС, а сколько – ингерентности или встраиваемости характеристик сформированного РПС, как отдельного, особого вида акустических сигналов, в существующую ИТК инфраструктуру, предназначенную для передачи-приема речевых сообщений в различных видах их представления (аналог-цифра), с учетом всех возможных различных промежуточных преобразований.

В этом смысле с позиций системного подхода [5] предлагается различать следующие основные группы *РПС*, рассматриваемых *в качестве моделей* исходной речи – оригинала:

– *модель прямого подобия* источнику речи: РПС как речевое сообщение, воспринятое и запомненное посредством органов слуха или как сигнал, зафиксированный на носителе информации в виде аудиозаписи, фонограммы;

– *модель косвенного подобия* или *аналогия* по большинству основных характеристик оригинальной речи, прежде всего по гармонической и формантой структуре: линиям гармоник основного тона (ЛГОТ), спектральной огибающей с треками движения формант и др.;

– *модель условного подобия* подразумевает соответствие такой модели речевому оригиналу в плане реализации определенных для него возможностей *в результате соглашения* между пользователями и носит условный, информационный характер.

Прямое подобие между моделью и оригиналом устанавливается вследствие их непосредственного взаимодействия как результат слухового восприятия – образ звука речи, либо вследствие цепочки таких взаимодействий, например, как результат фиксации речи на носителе информации – микрофон, аудиозапись, фонограмма. Нередко модель прямого подобия – фонограмму, из-за достаточной полноты присутствующих в ней характеристик исходной речи рассматривают в качестве оригинала, чтобы на практике уже использовать производные от нее модели косвенного и условного подобия записанным на ней сигналам.

Для косвенного подобия, или аналогии похожесть, аналогичность двух явлений объясняется совпадением закономерностей, которым они подчиняются [5]. К сигналам этой группы можно отнести такие РПС как-то: синтезированная компьютером речь; непреднамеренно искаженный техническими устройствами и помехами голос, утративший натуральность и естественность; намеренно измененный голос по признакам

возраста, пола и т.п.; клонированный программно-техническими средствами голос, подражающий конкретному диктору; «речеподобные помехи» (РПП), сложенные из фрагментов РС диктора (ов), для использования в системах активной акустозащиты каналов речевой связи и помещений конфиденциальных переговоров и др.

Однако следует осторожно пользоваться аналогиями, поскольку, кроме совпадающих закономерностей и характеристик, у в принципе разных сигналов (РС и РПС) есть и несовпадающие. Поэтому не все заключения о РПС как модели можно переносить на оригинал и не все особенности оригинала содержатся в модели – аналоге. В ряде приложений для РПС целесообразно даже использовать понятие «*сила аналогии*», связанное со степенью «перекрытия» сравниваемых характеристик РС и РПС [5].

Третий тип реальных моделей основан на подобии, которое не является ни прямым, ни косвенным. Они успешно работают, но лишь до тех пор, пока известны и соблюдаются договоренности, соглашения о их значении [5].

Например, подбираются такие параметры РПС, которые реализуют возможность приема-передачи речеподобного сигнала в существующей речевой ИТК инфраструктуре для передачи в нем иной, отличной от речевой информации: психоэмоциональной, графической, биометрической, стеганографической и др. [3, 6, 7, 8].

Важная особенность РПС как модели состоит в том, что такая модель никогда не будет тождественна оригиналу (даже, когда стараются всеми силами этого достичь, например, клонированный голос с заданными параметрами диктора). Часто в этом просто нет необходимости: каждая модель нужна для определенной цели, а для этого потребуется лишь некоторая и далеко не вся информация об оригинале [5].

Целевая предназначенность моделей имеет ряд важных следствий.

Первое состоит в том, что цель моделирования и сам выбор модели РПС определяется пользователем, в зависимости от задачи, которую он собирается решить.

Кроме того, разнообразие целей ведет к множественности различных моделей для одного и того же оригинала. В ряде приложений различные модели явления, процесса могут даже противоречить друг другу [5, 8] или совместно комплексно использоваться [3, 5].

Это в полной мере касается РС, где для различных аудиокодеков разработано и исследовано большое количество разнообразных моделей процессов речеобразования и речепреобразования [8].

3. Формирование термина «речеподобный сигнал» через свойства его системных моделей

Сформулированные ранее требования к РПС как системной модели, по наличию схожести ряда физических характеристик с оригинальным РС и свойству встраиваемости в инфраструктуру речевых коммуникаций, не дают непосредственной возможности установить метрику их подобия речевым сигналам и соответствия их характеристик особенностям ИТК инфраструктуры. Как верно отмечается в работах [1, 9] «отсутствие такой метрики является в настоящее время основным препятствием для становления единого и непротиворечивого понимания термина "речеподобный сигнал"».

Трудности формирования этого термина связаны ещё и со следующими обстоятельствами.

Во-первых, на практике конкретные приложения зачастую требуют сходства РПС с РС лишь по отдельным, каждый раз по-новому выбираемым характеристикам РС [1] и моделям процессов речеобразования, нередко с учетом особенностей слухового восприятия [3, 8].

Во-вторых, отсутствует понимание того, что в полосе частот РС помимо или вместо речевой в ИТК среде может передаваться информация иного рода, внедренная в параметры такого РПС [3].

В-третьих, отсутствует должное методологическое разграничение подходов к различным видам речевой пост- и предобработки, в ходе которых при достижении поставленной цели возможны модификации конструкций и характеристик оригинального РС (речепреобразование), которые позволяют воспринимать результат таких преобразований как новый РПС.

Так, для последнего случая существенно различными могут оказаться результаты обработки сигналов, принимаемых в условиях шумов и помех. Так, при шумоочистке искаженного помехами РС восстановленный сигнал может полностью утратить некоторые свойства исходного РС и-или приобрети новые реконструированные признаки речи в процессе восстановления разборчивости. Это же можно отнести к РС, претерпевшему вокодерные преобразования в каналах речевой связи, а также в ходе выполнения процедур сжатия-восстановления в базах хранения голосовых данных и т.п.

В последние годы наблюдаются значительные успехи, связанные с развитием и применением технологий распознавания и формирования речи в основе которых лежит построение и обработка ее амплитудных (мощностных) спектрально-временных описаний или спектрограмм, рассматриваемых в качестве изображений. Так, рекурсивная архитектура нейронной сети WaveNet, построенная исследователями из DeepMind, позволяет преобразовывать звук или текст в другой звук напрямую, без привлечения заранее записанных фрагментов речи. Для перехода от сконструированных образов спектрально-временных описаний к волновой форме нового сигнала на основе WaveNet разработан и используется программный продукт вокодерного синтеза (генерации) звука по его мел-спектрограмме – Tacotron 2.

Помимо этого, для формирования РПС по заданным изображением линейных и мэл-спектрограмм, используются и другие методы спектральной инверсии, основанные на использовании глубинных нейронных сетей для голоса (DeepVoice) и алгоритмов вокодерного синтеза волновой формы РПС по спектрограмме с искусственной фазой типа WaveGlow, WaveRNN, Griffin-Lim, «речевая подпись» (РП) и др. [10–13].

В работе [3] на базе спектральной инверсии и узкополосной модели РС предложены методы формирования РПС на вокализованных и невокализованных участках с заданной гармонической и формантной структурой, позволяющие переводить изображение любого содержания в спектрограмму конкретного РПС.

Технология спектральной инверсии позволила создавать и применять такие перспективные речевые технологии, как: клонирование речи голосом известного диктора [14]; преобразование в звук речевой, биометрической, графической и иной информации; постановка, скрытная передача и обнаружение в РС цифровых графических аудиомаркеров в виде водяных знаков (ЦВЗ) или биопризнаков для подтверждения подлинности самого РС или его авторства.

Всё это теперь дает возможность уточнить термин и модели РПС с позиций содержащихся в нем спектрально-временных характеристик и их соотнесения с подобными характеристиками речи. При этом РПС, также, как и оригинальный РС (речь), снова будем рассматривать как сложный акустический сигнал, представляемый в виде суперпозиции узкополосных элементарных синусоподобных сигналов (вейвлетов) на коротких участках (фреймах, сегментах) наблюдения или анализа [3].

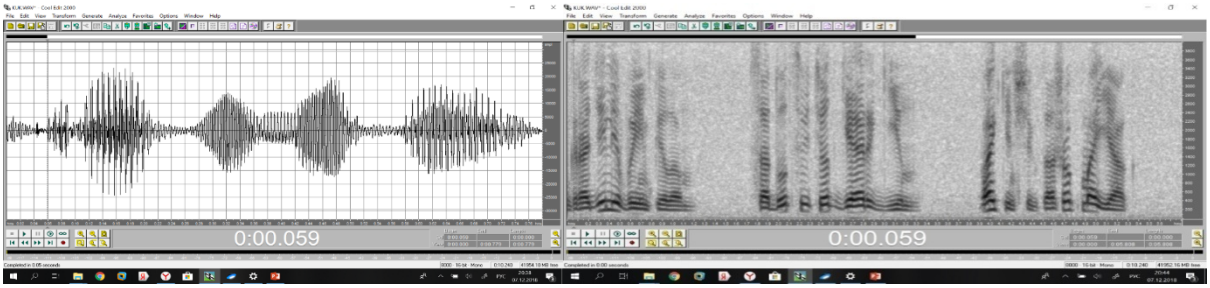
Тогда, в самом широком смысле, под *РПС,* как моделями *прямого подобия* оригинальной речи, здесь и далее будем понимать зарегистрированные в качестве

аналоговых и-или цифровых аудиозаписей, фонограмм сигналы, имеющие большинство сходных с оригинальной речью количественных и качественных характеристик в частотно-временной области. *Такие компоненты спектра этой модели РПС как формантная и гармоническая структура в приложениях не разделяются и не изменяются. Предусматривается абсолютное сходство звучания таких сигналов с речевыми оригиналами.*

Таким образом, по таким важным характеристикам, опознаваемым в ходе слухового восприятия как-то: громкость, разборчивость, естественность, узнаваемость и др., электроакустические модели РПС прямого подобия и оригинальная речь в виде прямого акустического сигнала практически полностью совпадают. Будут наблюдаться совпадения по частоте основного тона, форме спектральной огибающей, расположению формант и другим важным характеристикам.

Еще раз отметим, что РС и РПС прямого подобия в литературе часто отождествляются, хотя это и не совсем верно. Такой модели РПС всегда соответствует конкретный, порождающий её акустический РС, с большинством основных, прежде всего, физических характеристик которого наблюдается практически полное совпадение.

Пример осциллограммы и спектрограммы РПС прямого подобия представлен на рис. 2.



а) б)

Рис. 2. РПС прямого подобия как «след» оригинального (исходного) речевого сигнала, его осциллограмма а) и спектрограмма б)
(Fig. 2. Speech-like signal of direct similarity as a «trace» of the original (original) speech signal, its oscillogram a) and spectrogram б))

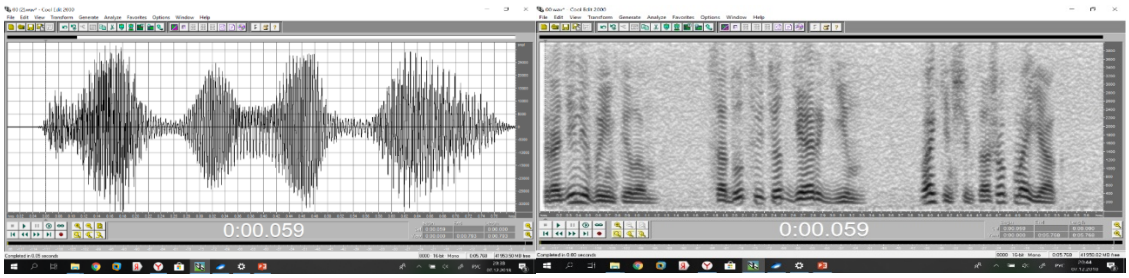
Здесь и далее на осциллограммах по оси абсцисс отложено время, по оси ординат – амплитуда. На спектрограммах время и частота – соответственно. Мощность сигнала на данной частоте в текущий момент времени указана в уровнях серого цвета. Черный цвет соответствует самому мощному пикселю спектра, белый – самому слабому. Спектрограммы строились посредством звуковых редакторов Adobe Audition v.1.5. и «Лазурь». Частота дискретизации составляла *8 КГц*, интервал анализа-синтеза *50* отсчетов или *6,25 мс*.

Под *РПС,* как моделями *косвенного подобия* оригинальной речи будем понимать сигналы, имеющие статистически схожие с оригинальной речью характеристики амплитудных спектрально-временных разверток, учитывающих специфику динамики спектральной огибающей усреднённой речи. *Формантная и гармоническая структура спектра такой модели РПС в приложениях могут разделяться, формироваться и обрабатываться как совместно, так и по отдельности. В последнем случае на конечном этапе все равно требуется их обязательное слияние-совмещение в общее изображение финишной спектрограммы, по которой будет синтезироваться РПС с заданными характеристиками. Предусматривается сходство звучания таких сигналов с речевыми.*

Хорошим и распространённым примером РПС косвенного подобия могут служить речеподобные помехи, а также результаты работы многочисленных программ компьютерных синтезаторов «текст в речь», и изменителей голоса. В последнем случае модели РПС косвенного подобия можно рассматривать как производные от моделей прямого подобия.

Задаваемая частота основного тона и рисунок формант частотного спектра являются важными параметрами, которые должны быть обязательно учтены при формировании РПС косвенного подобия для решения конкретной задачи обработки и/или защиты речевой информации. Такой модели РПС не всегда ставится в соответствие конкретный, оригинальный РС. Нередко для конкретного приложения модель косвенного подобия формируется своим определенным образом, без «равнения» на оригинал-аналог. Для получения необходимых частотно-временных характеристик, РПС схожих с оригинальной речью, используются среднестатистические данные о ее основных характеристиках, как, например, в табл. 1.

Примеры такого рода РПС косвенного подобия показаны на рис. 3 и 4.



а) б)

Рис. 3. Речеподобный сигнал (осциллограмма – а, спектрограмма – б)
(Fig. 3. Speech-like signals (oscillogram – a, spectrogram – б))

На рис. 3 приведен РПС, сгенерированный по спектрограмме рис. 2 исходного оригинального сигнала по улучшенному алгоритму синтеза Griffin-Lim+«Лазурь РП» [10] с подстановкой и протяжкой новой искусственной фазы для спектральных составляющих.

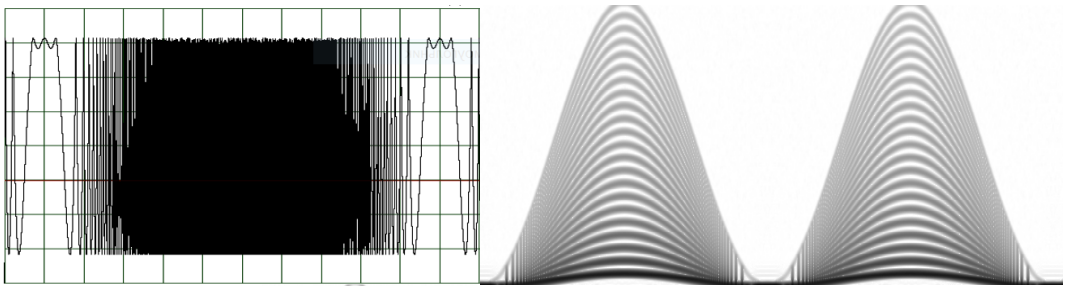
Для построения РПС рис. 3 использовалась фазовая спектральная информация о речевом оригинале – аналоге, получаемая из его амплитудного спектра. На рис. 4 информация о конкретном аналоге отсутствует, но используется информация о среднестатистических параметрах русской речи из табл. 1.

Осциллограммы на рис. 2 и рис. 3 не совпадают, хотя спектрограммы и звучание обоих сигналов практически неразличимы, даже, несмотря на то, что, РПС рис. 3 синтезировался по спектрограмме рис. 2 с учетом и совместной обработкой ее гармонической и формантной структур, используя математическую модель РС как суммы узкополосных сигналов [3], а протяжка фазы осуществлялась с искусственно заданными начальными значениями.

На рис. 4 показаны РПС, синтезированные без равнения на речевой оригинал, и по своей структуре близкие к гласным звукам речи, по сути являющиеся их идеализированным представлением.

В частотном спектре РПС на рис. 4, как и в обычной речи, на участках речевых вокализмов, содержится большое число гармонических составляющих, амплитуды которых уменьшаются с ростом частоты. В процессе генерации такой сигнал может быть подвергнут обработке эквалайзером с целью коррекции спектра для еще большего

соответствия усреднённому спектру русской речи. Также программным способом может меняться частота основного тона и длительность РПС.



а) б)

Рис. 4. Тестовый речеподобный сигнал (осциллограмма – а), спектрограмма – б)) косвенного подобия без наличия речевого аналога-оригинала
(Fig. 4. Test speech-like signal (oscillogram-a), spectrogram-б)) of indirect similarity without the presence of a speech analogue-the original)

Полученные на рис.4 изображения спектрограмм квазивокализованных речеподобных сигналов микшировались с разными типами формант. По полученному изображению спектрограммы как полезной смеси его гармонической и формантной структур производился синтез нового РПС и строилась его динамическая узкополосная спектрограмма, на основании визуального анализа изображения которой делался вывод о возможном использовании данного РПС в качестве своеобразного тестового сигнала в приложениях обработки и защиты РИ.

Результат одного из таких экспериментов показан на рис. 5.

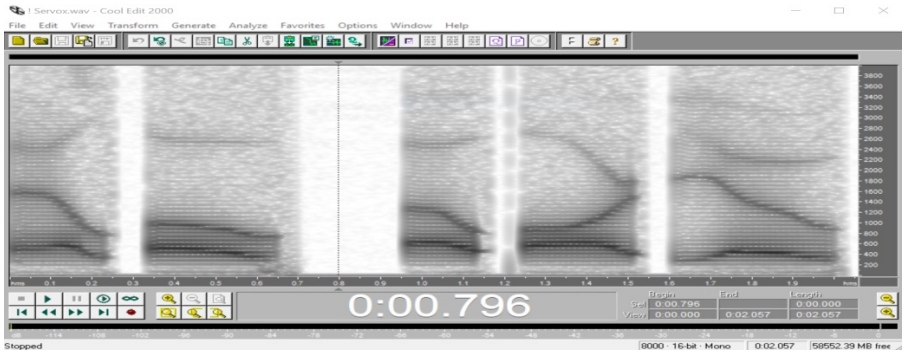


Рис. 5. Спектрограмма РПС как модели косвенного подобия в виде смеси гармонической и формантной структур
(Fig. 5. Spectrogram of speech-like signal as a model of indirect similarity in the form of a mixture of harmonic and formant structure)

Здесь использовались три форманты для формирования каждого отдельного нового слова синтезируемой речи. Каждая из формант двигалась на изображении спектрограммы по своей собственной траектории. Для корректировки изображения спектрограммы с последующим синтезом по ней нового РПС, также использовались инструменты отрисовки антиформант в зонах их действия. Частота основного тона – *130 Гц* (средняя для мужских голосов) была жестко задана и не менялась на всем протяжении РПС, поэтому ЛГОТ параллельны оси времени.

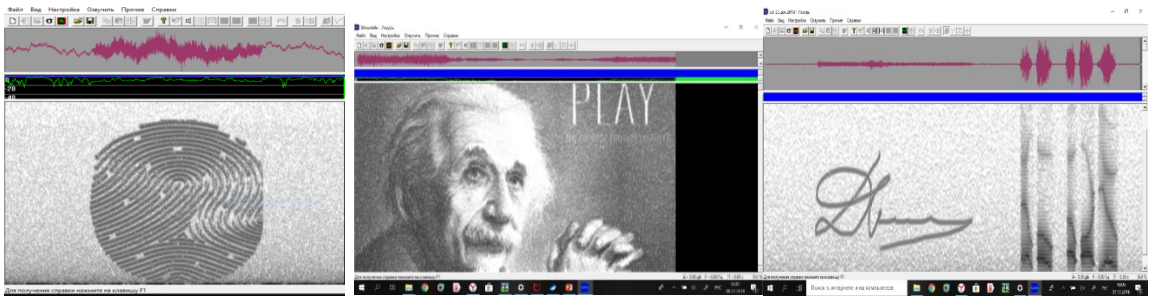
Примерно таким же образом формируют синтезированную речь голосообразующие аппараты для людей с нарушениями речи. В них правда происходит конвергенция искаженной оригинальной речи больного с прооперированными голосовыми связками в искусственно созданную разборчивую речь не лучшего качества звучания. Для его улучшения необходима коррекция движения на спектрограммах ЛГОТ с последующим синтезом и переходом к волновой форме нового РПС.

Кроме того, такие модели РПС могут быть использованы для создания специальных типов аудиомаркеров для защиты бумажных и электронных документов от фальсификации и подделки [10], например, в виде бинарной речевой подписи (РП), показанной как пример на рис. 1 (г) и проставляемой на защищаемом документе вместе с печатью и рукописной подписью. В такой РП могут содержаться основные, ключевые моменты защищаемого документа, которые не должны быть сфальсифицированы. При считывании-фотографировании такой РП с последующим синтезом и озвучиванием по так получаемой спектрограмме речевого сообщения происходит аудитивное сравнение его смыслового содержания с основными местами текста документа с одновременным узнаванием голоса его автора [10].

Под *РПС*, как моделями *условного подобия* оригинальной речи, предлагается понимать сигналы, имеющие заданную мощность, длительность звучания и форму амплитудной спектральной огибающей в речевой полосе частот, несущие в своих параметрах иную, отличную от речевой информацию.

Для такой модели РПС в приложениях формируется, обрабатывается и синтезируется в звук в речевом диапазоне частот только формантная или только гармоническая структура спектра. Сходства звучания таких сигналов с речевыми не предусматривается.

Примеры осциллограмм и спектрограмм РПС условного подобия, несущие в своих спектральных характеристиках биометрическую информацию легитимного пользователя удаленного защищенного информационного ресурса представлены на рис. 6.



а) б) в)

Рис. 6. Осциллограммы и спектрограммы РПС со встроенными в спектральной развертке биометрическими данными легитимного пользователя удаленного защищенного ресурса: отпечатка пальца – а); фотопортрета – б); рукописной подписи и параметров процесса ее написания – в)
(Fig. 6. Oscillograms and spectrograms of speech-like with built-in spectral scan biometric data of a legitimate user of a remote protected resource: fingerprint-a); photo-portrait-б); handwritten signature and parameters of the process of its writing-в))

На приемном конце канала речевой связи такие биометрические признаки можно выделить из РПС и сравнить с эталонными в базе данных, для принятия решения о правомочности пользователя.

Отдельно сформированная гармоническая структура РПС условного подобия позволяет синтезировать и предавать по каналам голосовой связи сигналы, в полосе частот которых рисунок треков узкополосных процессов (гармоник) спектрограммы повторяет рисунок отпечатка пальца легитимного пользователя (рис. 6 а).

Отдельно сформированная формантная структура РПС условного подобия позволяет синтезировать и предавать по каналам голосовой связи сигналы, в полосу частот которых встроен, например, фотопортрет легитимного пользователя (рис. 6 б).

По мнению авторов, особенно интересен, удобен и перспективен комплексный способ подтверждения подлинности голоса удаленного пользователя (см. рис. 6 в), посредством включения в подтверждающий РПС статического образа его рукописной подписи (условное подобие) и динамических параметров процесса ее написания (косвенное подобие).

Таким образом, совместно сформированные модели РПС условного и косвенного подобия могут быть использованы для кодирования статических и динамических характеристик процесса написания рукописной подписи (образ сигнатуры, нажим и направление движения пера при ее написании), «подписывающей» или подтверждающей подлинность значимого речевого сообщения или произносимой секретной фразы в виде динамически изменяемых гармонических и формантных параметров нового РПС синтезированного по соответствующим образом подготовленной спектрограмме.

Такие РПС со встроенной в их спектр данных о биопризнаках или цифровом коде удаленного пользователя можно передавать по каналам речевой связи подтверждая легитимность абонента и не прерывая при этом процесса его голосового взаимодействия с оператором call-центра (см. рис. 6).

Рассмотрим подробнее различные области применения предложенного комплекса системных моделей РПС и их возможную классификацию.

4. Области применения РПС и их классификация

Как уже отмечалось РПС, наряду с традиционными РС, находят всё большее применение в областях безопасности управления и связи, всё чаще замещая в них последние [12, 13].

Поэтому первичную классификацию РПС можно проводить по указанным областям их применения, а вторичную – по методам их моделирования.

В области безопасности РПС активно применяются при реализации таких направлений, как-то:

– голосовая идентификация и аутентификация (РПС используется как модель прямого подобия);

– постановка активных маскировочных речеподобных помех в целях активной акустозащиты выделенных помещений (ВП) конфиденциальных переговоров и речевых каналов связи от несанкционированного доступа (РПС как комплекс моделей косвенного и условного подобия);

– оценка остаточной речевой разборчивости (РР) в выделенных помещениях после включения средств активной и пассивной акустозащиты (РПС как модель косвенного подобия);

– скрытная передача данных, «зашитых» в параметрах РПС в каналах голосового трафика сотовых систем связи, используя преимущества приоритетной передачи голоса в них (РПС как комплекс моделей косвенного и условного подобия);

– скрытная передача речевой информации (РИ) в сетях и каналах передачи данных (РПС как модель косвенного подобия);

– защита авторских и смежных прав на культурный продукт и произведения, представленные и передаваемые в форматах цифрового контента (РПС как комплекс моделей косвенного и условного подобия);

– аудиомаркирование носителей информации и ее верификации посредством постановки на них и считывания изображений сонограмм (аудимаркеров) – речевой подписи, с последующим восстановлением по ним РПС с включенными в его смысловое содержание ключевыми моментами защищаемой информации (РПС как модель косвенного подобия);

– восстановление РР и качества звучания, искаженной шумами и помехами речи в т.ч. с использованием для этих целей голосовой базы данных конкретного диктора (РПС как модель косвенного подобия);

– тестирование речевых СКУД, систем автоматического распознавания речи и речевого управления на возможность незаконного проникновения и управления (РПС как модель косвенного подобия);

– обнаружение клонированных голосов и подтверждение подлинности РС и др. (РПС как комплекс моделей косвенного и условного подобия).

Наиболее часто в публикациях по защите РИ от несанкционированного прослушивания встречаются упоминания об отдельном особом виде РПС – речеподобных помехах (РПП), применяемых для систем активного акустозаграждения и речевого маскирования помещений конфиденциальных переговоров. Рассмотрим их подробнее.

Действительно, речеподобные помехи как особая разновидность речеподобных сигналов последнее время широко применяются в качестве маскирующих сигналов для активных методов защиты речевой информации от утечки по техническим акустическим каналам. При этом здесь чаще используются комбинированные маскирующие сигналы, включающие и «белый» шум и речеподобные сигналы. Соотношение между уровнями мощности «белого» шума и РПС рекомендуется обычно в пределах 3–12 дБ [1, 9].

Важным требованием к маскирующим сигналам является то, чтобы их основные характеристики имели случайный характер, Эти требования обусловлены необходимостью исключения какой-либо возможности шумоочистки перехваченных нарушителем акустических сигналов от маскирующих шумов и помех типа РПС. Хотя и не для всех методов восстановления разборчивости речи в шумах это справедливо [4]

К РПП как особой разновидности РПС для систем защиты речевой информации помимо согласованности с акустическими свойствами голоса определенного диктора дополнительно предъявляются требования, обусловленные характерными лингвистическими и фонетическими свойствами определенного языка.

Согласно работам [1, 9] лингвистические требования включают статистические данные о длине предложений, длине синтагм, длине слов, длительности пауз, а также вероятности появления определенных фонем для определенного языка, а фонетические требования обусловлены необходимостью формирования речеподобных сигналов со спектральными характеристиками, либо не отличающимися, либо кардинально отличающимися (как антиформанты от формант) от спектральных характеристик голоса диктора, речь которого необходимо защитить.

Стремятся также, чтобы генерируемые РПП были похожи на реальные РС не только по временным и спектральным характеристикам, но и по восприятию на слух [1, 9]. В ряде других исследований указывается, что при формировании РПП должны использоваться одинаковые по мощности образцы речи не менее 3-х дикторов.

Классификация РПП тоже пока не устоялась. Генерируются они способами, которые представляются разработчикам наиболее подходящими для имитации

акустического проявления речи. При этом зачастую отсутствуют необходимые разъяснения – достижение какого, именно, сходства с РС является целью формирования РПП в каждом отдельном случае [1, 9].

Так в работе [1] приводятся несколько определений РПП и описаний способов их формирования. Это и «маскирующий сигнал», и «речевой хор», состоящий из суммы нескольких речевых сигналов, и «речеподобный шум, алгоритм синтеза которого теоретически обоснован, представлен аналитическим выражением и адаптирован под первые две-три форманты интегрального формантного спектра русского языка», или «шум с огибающей амплитудного спектра, подобной речевому сигналу» в речевом «диапазоне частот от *175* до *5600 Гц*».

В работе [1] также отмечается, что для маскирования РС эффективной речеподобной помехой «…специалистами в основном предлагается создание трех видов такой помехи: «речеподобная помеха-1» – формируется из фрагментов речи трех дикторов радиовещательных станций при примерно равных уровнях смешиваемых сигналов; «речеподобная помеха-2» – формируется из одного доминирующего речевого сигнала или музыкального фрагмента и смеси фрагментов радиопередач с шумом; «речеподобная помеха-3» – формируется из фрагментов скрываемого речевого сигнала при многократном их наложении с различными уровнями».

Как воспринимаются все упомянутые здесь РПП на слух (мешают ли проведению конфиденциальных переговоров, требуют ли форсирования голоса у выступающих и помогают проявлению других демаскирующих признаков) и чем они подобны РС недостаточно ясно. Также неясно как помехи этих видов «устойчивы» по отношению к современным технологиям шумоочистки. С этих позиций «речеподобная помеха-2» кажется наиболее уязвимой, как впрочем и другие РПП, уровень которых значительно превышает уровень защищаемого РС и поэтому признаку спектрограмма полезной смеси сигнала с шумом может быть разделена и защищаемый РС сможет быть восстановлен по реконструированной спектрограмме, особенно при использовании для этого предварительно созданной, собранной из открытых источников голосовой базы данных основного диктора.

По понятным причинам от разработчиков нельзя ожидать полной прозрачности освещения принципов формирования предлагаемых ими маскирующих помех, что не дает уверенности в эффективности их применения. Исследование влияния акустозаградающих речеподобных помех на защищенность конфиденциальных переговоров требует отдельной глубокой проработки. В любом случае в нашей классификации РПП можем отнести к моделям РПС косвенного подобия.

Требования, подобные тем, что предъявляются к РПП, можно применить и для РПС, создаваемых для тестирования систем защиты речевой информации от атак клонированного голоса легитимного пользователя.

В области телекоммуникаций и связи РПС используются в основном для тестирования качества каналов и средств передачи речевых сигналов, тестирования качества передачи речевой информации по линиям связи (РПС как комплекс моделей косвенного и условного подобия);

В соответствии с работами [1, 9] для тестирования линий связи основные требования к речеподобным сигналам заключаются в наличии в них трех основных компонент: звонких вокализованных сигналов с имитацией свойств голоса на участках речевых вокализмов (РВ); сигналов с постоянной спектральной плотностью мощности в частотной области; наличия определенных пауз для обеспечения свойств, присущих сигналам с амплитудной модуляцией. Этим требованиям вполне могут соответствовать

РПС косвенного и условного подобия, создаваемые путем синтеза по спектрограмме с заданной гармонической и/или формантной структурой.

Стандарт Международного союза по телекоммуникациям ITU-T P.501 [1] относит к РПС так называемые «составные сигналы», предназначенные для испытания (тестирования) каналов связи. Отмечается близость их усредненных спектральных и временных характеристик к соответствующим характеристикам реальных РС. Слуховое сходство звучания испытательных сигналов с речевыми в стандарте P.501 не декларируется. Подобного рода сигналы широко используются в объективных быстрых методах определения речевой разборчивости (РР) типа STI, RASTI в прямых и побочных каналах речевой связи. В них для оценки индекса модуляции для определения РР используются порядка десятка видов огибающих спектра РПС с низкими частотами [15].

Подобные выше рассмотренные тестовые сигналы можем отнести к комплексу моделей косвенного и условного подобия.

Рассмотрим другие области применения РПС в области связи и телекоммуникаций:

– компадирование и восстановление полосы частот РС без потери его информативности – разборчивости (или с частичной допустимой потерей) для систем помехоустойчивой радиосвязи, организации подземной акустической голосовой связи для шахтеров в условиях возможных и состоявшихся чрезвычайных ситуаций (РПС как комплекс моделей косвенного и условного подобия);

– повышение надежности радиосвязи в условиях помех (феддинг) и низкой пропускной способностью канала (РПС используются в качестве комплекса моделей косвенного и условного подобия как тестовые сигналы, так и как сигналы, полученные в результате применения процедур шумоочистки));

– квазиречевая связь («разговор») компьютеров, роботов друг с другом в диапазоне частот ультразвука, инфразвука, естественной речи (РПС как модель косвенного подобия);

– повышение качества звучания (разборчивости, натуральности и естественности) речеобразующих и модифицирующих голос аппаратов для инвалидов с повреждениями голосовых связок, трахеи (РПС как модель косвенного подобия);

– повышение качества работы слуховых аппаратов для инвалидов (РПС как модель косвенного подобия);

– коррекция голосового звучания в отличных от воздуха газовых средах (водолазы, космонавты) (РПС как модель косвенного подобия);

– подтверждение подлинности, достоверности речевых сообщений в системах речевой связи (РПС как комплекс моделей косвенного и условного подобия).

В области речевого управления РПС находят свое применение при решении следующих задач:

– создание искусственного голоса в том числе клона по образцу голоса конкретного диктора (РПС как модель косвенного подобия);

– сонификация (сигнальное озвучивание) результатов мониторинга работы кибер-физических и других систем (РПС как модель условного подобия);

– подтверждение подлинности, достоверности голосовых команд и речевых сообщений в системах речевого управления (РПС как комплекс моделей косвенного и условного подобия);

– синтез текста в речь (T2S) (РПС как модель косвенного подобия);

– автоматическое распознавание речи (РПС как модель косвенного подобия).

Здесь к создаваемым, синтезируемым РПС применяются такие же требования как для тестирования эффективности систем защиты речевой информации, находящиеся под атакой клонированного голоса.

В области общей методологии создания и применения РПС можно выделить следующие направления:

– моделирование систем и процессов слухового восприятия, речеобразования и речепреобразования для проработки вопросов безопасности (скремблирование, шифрование, тестирование и оценка эффективности средств защиты РИ), связи (сжатие-восстановление РС) и управления (обнаружение, аудио маркирование, маскирование, скрытная передача) с использованием РПС (РПС как комплекс моделей косвенного и условного подобия);

– исследование общих принципов и возможностей применения РПС в качестве тестовых сигналов в различных областях использования (РПС как комплекс моделей косвенного и условного подобия);

– исследование и создание методов синтеза (генерация) РПС с требуемыми свойствами для применения в сферах безопасности, управления, связи, рекламы, культуры (киноиндустрия), развлечений и др. (РПС как комплекс моделей косвенного и условного подобия).

Не исключены и другие области применения РПС как комплекса системных моделей в сфере безопасности, связи и управления в основе которых также лежит подход, связанный с подготовкой и формированием изображений узкополосных амплитудных спектрограмм с заданными свойствами с последующим синтезом по ним волновой формы нового РПС с требуемыми характеристиками.

Заключение

В статье представлены основные характеристики речевых и речеподобных сигналов. В зависимости от возможных приложений рассмотрены и изложены требования к речеподобным сигналам (РПС).

Определены системные модели РПС в рамках культуры речевых коммуникаций и существующей инфокоммуникационной среды

Рассмотрены трудности формирования термина речеподобный сигнал и сделана попытка его определения через свойства его системных моделей прямого, косвенного и условного подобия оригинальной речи.

Предложена классификация РПС по областям их применения и комплексу генеративных системных моделей для формирования речеподобных сигналов с заданными свойствами гармонической и формантной структуры их спектральной огибающей.

Рассмотрены основные методы синтеза волновой формы речеподобных сигналов по заданным изображениям их спектрограмм. Отмечена перспективность направления прямого синтеза посредством использования усовершенствованной узкополосной синусоидальной модели звуко- и речеобразования и технологии образного анализа-синтеза изображений спектрограмм [3, 16].

Приведены практические примеры перевода в РПС с последующей передачей по каналам речевой связи биометрических признаков пользователей удаленных ресурсов для их дополнительной аутентификации в системах контроля и управления доступом.

СПИСОК ЛИТЕРАТУРЫ:

1. Воробьёв В.И., Давыдов А.Г., Давыдов Г.В. Речеподобные сигналы: разновидности, основные параметры, способы формирования, области применения. / Доклады БГУИР. №3 (41). 2009. С. 9–6. URL: https://cyberleninka.ru/article/n/rechepodobnye-signaly-raznovidnosti-osnovnye-parametry-sposoby-formirovaniya-oblasti-primeneniya/viewer (дата обращения: 01.12.2019).
2. Устинов, Роман А. Особенности современных систем защиты речевой информации. / Безопасность информационных технологий, [S.l.], v. 24, n. 4. P. 71–79, nov. 2017. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/279 (дата обращения: 24.01.2019). DOI: http://dx.doi.org/10.26583/bit.2017.4.08.
3. Дворянкин, Сергей В.; Дворянкин, Никита С.; Устинов, Роман А. Развитие технологий образного анализа-синтеза акустической (речевой) информации в системах управления, безопасности и связи. Безопасность информационных технологий, [S.l.], v. 26, n. 1. P. 64–76, mar. 2019. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/1186> (дата обращения: 01.12.2019). DOI: http://dx.doi.org/10.26583/bit.2019.1.07.
4. Дворянкин С.В., Макаров Ю.К., Хорев А.А. Обоснование критериев эффективности защиты речевой информации от утечки по техническим каналам / Защита информации. Инсайд. 2007. № 2 (14). С. 18–25. URL: http://www.inside-zi.ru/pages/2_2007/18.html (дата обращения: 01.12.2019).
5. Прикладной системный анализ: учебное пособие / Ф.П. Тарасенко. 2-е изд., перераб. и доп. - Москва. КНОРУС, 2019. – 322 с.
6. Дворянкин, Никита Сергеевич. Анализ методов скрытного маркирования голосовых команд дистанционного речевого управления для подтверждения их подлинности. Безопасность информационных технологий, [S.l.], v. 24, n. 1. P. 18–27, apr. 2017. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/51> (дата обращения: 01.12.2019).
7. Дворянкин С.В., Нагорных И.М. К вопросу о технологии преобразования звук – изображение – звук. / Спецтехника и связь. 2013. № 1. С. 28–32. URL: https://elibrary.ru/item.asp?id=18876233. (дата обращения: 01.12.2019).
8. Spanias A.S.,"Speech Coding: A Tutorial Review, Proc. Ieee, 82 (1994), 1541–1582.
9. Давыдов Г.В., Попов В.А., Потапович А.В., Сейткулов Е.Н., Савченко И.В. Синтез речеподобных сигналов на белорусском языке. / Доклады БГУИР. №4 (90). 2015. Р. 9–16. URL: https://elibrary.ru/item.asp?id=29737598 (дата обращения: 01.12.2019).
10. Дворянкин С.В. Речевая подпись. М.: РИО МТУСИ. 2003. – 184 с.
11. Natural Tts Synthesis By Conditioning Wavenet On Mel Spectrogram Predictions Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, RJ Skerry-Ryan, Rif A. Saurous, Yannis Agiomyrgiannakis, and Yonghui Wu. URL: https://arxiv.org/pdf/1712.05884.pdf (дата обращения:12.02.2018).
12. Griffin D.W. and Lim J.S., “Signal estimation from modified short-time Fourier transform,” IEEE Transactions on Acoustics, Speech and Signal Processing. 1984. P. 236–243.
13. McAulay R.J., Quatieri T.F, Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE Trans. on Acoust., Speech and Signal Processing. 1988. Vol. ASSP-34. P. 744–754.
14. Malik, Hafiz. “Securing Voice-driven Interfaces against Fake (Cloned) Audio Attacks.” CoRR abs/1902.06782 (2019): n. Pag
15. Рашевский Я.И., Каргашин В.Л. Обзор зарубежных методов определения разборчивости речи. [Текст]: научное издание / Ярослав Игоревич Рашевский, Виктор Леонидович Каргашин // Спец. техн. – 2003. N 1. С. 44–52: 1 ил. – 4 URL: http://irbiscorp.spsl.nsc.ru/webirbis-cgi-cnb-new/cgiirbis_64.exe?Z21ID=&I21DBN=AB&P21DBN=AB&S21STN=1&S21REF=&S21FMT=fullwebr&C21COM=S&S21CNR=20&S21P01=0&S21P02=1&S21P03=A=&S21STR=Рашевский,%20Ярослав%20Игоревич (дата обращения: 01.12.2019).
16. Алюшин В.М., Дворянкин С.В. Технологии образного анализа в задачах цифровой обработки речевой информации. // Научная визуализация. 2013. Т. 5. № 3. С. 75–88. URL: http://sv-journal.org./2013-3/06.php?lang=en (дата обращения: 01.12.2019).

REFERENCES:

[1] Vorobyov V.I., Davydov A.G., Davydov G.V. Speech-like signals: varieties, main parameters, methods of formation, fields of application. BGUIR reports. №3 (41). 2009. P. 9–16. URL: https://cyberleninka.ru/article/n/rechepodobnye-signaly-raznovidnosti-osnovnye-parametry-sposoby-formirovaniya-oblasti-primeneniya/viewer (accessed: 01.12.2019) (in Russian).

[2] Ustinov, Roman A. Specific features of modern voice protection systems. IT Security (Russia), [S.l.], v. 24, n. 4. P. 71–79, nov. 2017. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/279>. (accessed: 24.01.2019). DOI: http://dx.doi.org/10.26583/bit.2017.4.08. (in Russian) (in Russian).
[3] Dvoryankin V.; Dvoryankin, Nikita S.; Ustinov, Roman A. Improvement of image analysis/synthesis technologies of acoustic (speech) information for the control, safety and communication systems. IT Security (Russia), [S.l.], v. 26, n. 1. P. 64–76, mar. 2019. ISSN 2074-7136. URL:<https://bit.mephi.ru/index.php/bit/article/view/1186> (accessed: 01.12.2019). DOI: http://dx.doi.org/10.26583/bit.2019.1.07 (in Russian).
[4] Dvoryankin S.V., Makarov Yu.K., Horev A.A. Justification of criteria of efficiency of protection of speech information from leakage on technical channels. Zashchita informatsii. Insider trading. 2007. No. 2 (14). P. 18–25. URL: http://www.inside-zi.ru/pages/2_2007/18.html (accessed: 01.12.2019) (in Russian).
[5] Applied system analysis: textbook / F. p. Tarasenko. 2nd ed., Rev. and dop. - Moscow. KNORUS, 2019. – 322 p. (in Russian).
[6] Dvoryankin, Nikita Sergeevich. Analysis methods of secretive labeling voice commands for remote voice control to confirm their authenticity. IT Security (Russia), [S.l.], v. 24, n. 1. P. 18–27, apr. 2017. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/51> (accessed: 01.12.2019) DOI:http://dx.doi.org/10.26583/bit.2017.1.03 (in Russian)
[7] Dvoryankin S.V., Nagornykh I.M. On the technology of sound – image – sound transformation. Spetstekhnika i sviaz. 2013. № 1. P. 28–32. URL: https://elibrary.ru/item.asp?id=18876233. (accessed: 01.12.2019) (in Russian).
[8] Spanias A.S., "Speech Coding: A Tutorial Review, Proc. Ieee, 82 (1994), 1541–1582.
[9] Davydau H.V., Papou V.A., Potapovich A.V., Seitkulov Y.N., Savchenko I.V. Synthesis of speech-like signals in the belarusian language. BGUIR reports. №4 (90). 2015. P. 9–16. URL: https://elibrary.ru/item.asp?id=29737598 (accessed: 01.12.2019) (in Russian).
[10] Dvoryankin S.V. Speech signature.M.: RIO MTUSI. 2003. – 184 p. (in Russian).
[11] Natural Tts Synthesis By Conditioning Wavenet On Mel Spectrogram Predictions Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, RJ Skerry-Ryan, Rif A. Saurous, Yannis Agiomyrgiannakis, and Yonghui Wu. URL: https://arxiv.org/pdf/1712.05884.pdf (accessed:12.02.2018).
[12] Griffin D.W. and Lim J.S., "Signal estimation from modified short-time Fourier transform," IEEE Transactions on Acoustics, Speech and Signal Processing, 1984. P. 236–243.
[13] McAulay R.J., Quatieri T.F, Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE Trans. on Acoust., Speech and Signal Processing. 1988. Vol. ASSP-34. P. 744–754.
[14] Malik, Hafiz. "Securing Voice-driven Interfaces against Fake (Cloned) Audio Attacks." CoRR abs/1902.06782 (2019): n. Pag
[15] Rashevsky Ya.I., Kargashin V.L. Review of foreign methods for determining speech intelligibility. [Text]: scientific edition Yaroslav Igorevich Rashevsky, Viktor Leonidovich Kargashin. Spetsialnaia tekhnika. – 2003. N 1. P. 44–52. URL: http://irbiscorp.spsl.nsc.ru/webirbis-cgi-cnb-new/cgiirbis_64.exe?Z21ID=&I21DBN=AB&P21DBN=AB&S21STN=1&S21REF=&S21FMT=fullwebr&C21COM=S&S21CNR=20&S21P01=0&S21P02=1&S21P03=A=&S21STR=Рашевский,%20Ярослав%20Игоревич (accessed:12.02.2018) (in Russian).
[16] Alyushin V., Dvoryankin S. The Technology of Figurative Analysis in the Problems of Speech Information Digital Processing. Scientific Visualization. 2013. T. 5. Vol. 3. P. 75–88. URL: http://sv-journal.org./2013-3/06.php?lang=en (accessed:12.02.2018).

Поступила в редакцию – 05 ноября 2019 г. Окончательный вариант – 05 декабря 2019 г.
Received – November 05, 2019. The final version – December 05, 2019.