

УДК 004.934.5

РЕЧЕПОДОБНЫЕ СИГНАЛЫ: РАЗНОВИДНОСТИ, ОСНОВНЫЕ ПАРАМЕТРЫ, СПОСОБЫ ФОРМИРОВАНИЯ, ОБЛАСТИ ПРИМЕНЕНИЯ

В.И. ВОРОБЬЁВ¹, А.Г. ДАВЫДОВ², Г.В. ДАВЫДОВ¹

¹Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

²Объединенный институт проблем информатики НАН Беларуси
Сурганова, 6, Минск, 220072, Беларусь

Поступила в редакцию 23 января 2008

Анализируется вопрос о современном применении терминов "речеподобные" сигналы и помехи. Для термина "речеподобная" помеха предложено определение. Представлен алгоритм формирования речеподобных помех, соответствующих предложенному определению.

Ключевые слова: речевые и речеподобные сигналы (помехи), временное и частотное представления речевых сигналов, защита информации от утечки по акустическим каналам.

Введение

Термины "речеподобные сигналы (РПС) и речеподобные помехи (РПП)" в публикациях по речевым технологиям используются довольно часто [1–10]. В англоязычной литературе [11–14] распространено устойчивое выражение "speech-like signal" — буквально *речеподобный сигнал*. При ближайшем рассмотрении обнаруживается, что под одним и тем же названием понимаются самые разные акустические и электрические процессы. Это приводит к информационной дезориентации потребителей речевых технологий. Назрела настоятельная необходимость изменения этой неприемлемой ситуации.

Приведенные ниже материалы призваны способствовать упорядочению представлений относительно акустических и электрических процессов, которые *допустимо и целесообразно* трактовать как речеподобные сигналы.

Основные особенности речевых сигналов

Для корректного определения термина речеподобного сигнала как сигнала, подобного речи в четко оговоренном смысле, целесообразно предварительно рассмотреть основные характерные особенности речевых сигналов (РС). Обнаруживая наличие таких особенностей в анализируемых сигналах, вопрос об их принадлежности к РПС можно решать без привлечения экспертных оценок.

Используя гармоническую модель РС [15–17], любой звук $s(t)$ можно представить в виде:

$$s(t) = \sum_{p=1}^N A_p(t) \sin \left[2\pi p \int_0^t F_0(\tau) d\tau + \Phi_p(t) \right] + r(t), \quad t \in [0, T], \quad \tau \in [0, t], \quad (1)$$

где F_0 — мгновенная частота основного тона (ЧОТ) звука s ; A_p и Φ_p — сравнительно медленно меняющиеся амплитуда и фаза его p -й гармонической составляющей; N — число привлекаемых для анализа энергетически значимых гармонических составляющих звука s ; r — шумовая составляющая звука s ; T — время анализа звука s .

Приведенная формула пригодна для описания любых звуков речи, а на вокализованных участках РС помимо колебаний с ЧОТ позволяет наглядно отображать энергетически выраженные обертоны с частотами, кратными ЧОТ.

Среднеквадратическое значение звукового давления, создаваемого голосом человека, считается приблизительно равным 70 дБ, а уровень звукового давления посторонних шумов в лабораторных условиях составляет около 50–55 дБ. При этом характерное для практики речевого обмена отношение сигнал/шум примерно равно 15–20 дБ.

Основными признаками речевых сигналов во временной области являются [4]: всплесковый характер их временных изменений с длительностью всплесков 15–120 мс; группирование всплесков по 2–8 в пачки с промежутками между отдельными всплесками 10–50 мс при интервалах между пачками 0,01–1 с.

Спектральный состав речевых сигналов характеризуется следующими статистическими характеристиками ЧОТ (F_0) и формантных частот (F_1 – F_4):

Спектральный состав речевых сигналов

F_0	F_1	F_2	F_3	F_4
60–240 Гц;	150–850 Гц;	500–2500 Гц;	1500–3500 Гц;	2500–4500 Гц.

Приведенные диапазоны частот (F_0 – F_4) относятся к сонорным неносовым звукам речи при произнесении их средним мужским голосом [18].

У женских голосов ЧОТ в среднем оказывается на октаву выше, чем для мужских, а частоты формант — выше на 17%. У детей в возрасте около десяти лет формантные частоты выше в среднем на 25%, чем у взрослого мужчины, а ЧОТ в среднем имеет значение около 300 Гц [18].

Усредненные спектры РС значительно изменяются при переходе к интервалам усреднения, сопоставимым по длительности с длительностью одного звука (около 100 мс).

Распределение звуков речи выбранного языка часто представляют в координатах первых двух формант, наделенных основной смыслоразличительной функцией (ролью) в речи (рис. 1). Отличия в спектральном составе РС различных языков и типов речи подробно освещены в [20–23].

Статистические характеристики встречаемости элементов речи в различных языках оказываются разными. Данные по статистическим характеристикам распределения фонем русского языка приведены в [20].

Кроме фонетических отличий РС различных языков имеют особенности в интонационных характеристиках. Такими характеристиками РС являются зависимости ЧОТ, кратковременного уровня интенсивности и скорости произношения от времени. Интонационные характеристики в значительной степени определяют стиль речи и дают возможность различать незнакомые языки.

Перечисленные основные характеристики РС 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 следует указывать по каким характеристикам формируемые РПС соответствуют РС.

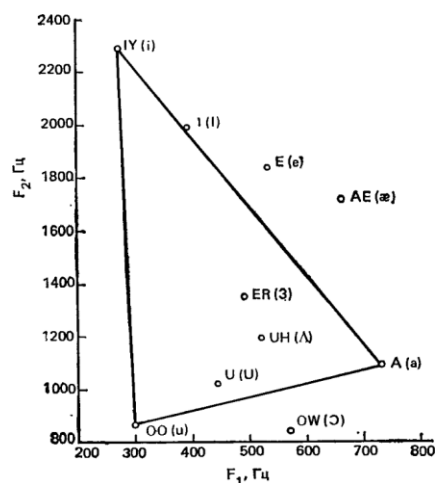


Рис. 1. Распределение средних формантных частот гласных звуков, произнесенных мужскими голосами [19]

Определение речеподобного сигнала

Согласно устоявшемуся определению [24], речь — исторически сложившаяся форма общения людей посредством языковых конструкций, создаваемых на основе определенных правил. В коммуникациях человека с помощью устной речи физическим посредником является РС. Основное его отличие от других видов сигналов состоит в том, что он несет информацию, которая может быть извлечена путем прослушивания речи. Весьма важно, что слуховая система человека отличается избирательностью к РС и помехоустойчивостью по отношению к мешающим акустическим воздействиям.

С учетом отмеченного и в согласии с этимологическими представлениями от *речеподобного сигнала* целесообразно потребовать акустического сходства с РС при условии отсутствия информации, которая может быть воспринята человеком.

Сформулированные требования к РПС, к сожалению, не дают непосредственной возможности установить метрику их подобия сигналам речевым. Отсутствие такой метрики является в настоящее время основным препятствием для становления единого и непротиворечивого понимания термина "речеподобный сигнал".

Трудности формирования этого термина связаны и еще с двумя обстоятельствами. Во-первых, на практике от РПС зачастую требуют сходства с РС лишь по отдельным, каждый раз вновь выбираемым характеристикам. Во-вторых, в литературе по речевым технологиям отсутствует должное методологическое разграничение подходов к выявлению присутствия РС в шуме и к шумоочистке этих сигналов. Такое положение не согласуется со сложившимися подходами к решению указанных задач в теории потенциальной помехоустойчивости, где проблемы обнаружения сигналов и минимизации ошибок их воспроизведения аргументировано рассматриваются автономно. Существенно различными при этом оказываются и результаты обработки сигналов, принимаемых на фоне помех. Так, при оптимальном обнаружении известного точно РС на фоне белого шума сигнал на выходе обнаружителя полностью утрачивает признаки речи. В связи с этим необходимо отметить, что в [10] для анализа эффективности маскирования РС различными видами помех аргументировано рассматривались биологические передатчик и приемник и корреляционный алгоритм. Там же резонно отмечена необходимость учитывать субъективные биологические свойства слуха: предмаскировку, постмаскировку, нелинейность, пороги слышимости и динамический диапазон.

Ниже рассматриваются примеры весьма разнообразного смыслового наполнения термина РПС в литературе.

РПС для тестирования каналов и средств передачи речевых сигналов

Стандарт Международного союза по телекоммуникациям ITU-T P.501 [11] к РПС относит так называемые "составные сигналы", предназначенные для испытания (тестирования) каналов связи в дуплексном режиме. Приведенное в стандарте P.501 описание испытательных РПС позволяет убедиться, что основание называться "речеподобными" они получили исключительно из-за близости их усредненных спектральных и временных характеристик к соответствующим характеристикам реальных РС. Сходство испытательных сигналов с речевыми, которое можно было бы обнаружить на слух, в стандарте P.501 не декларируется.

РПП для защиты речи от несанкционированного прослушивания

В средствах активной защиты РС от несанкционированного прослушивания стремятся использовать помеховые колебания с повышенным маскирующим действием и стохастическими проявлениями, затрудняющими шумоочистку полезных сигналов. Таковыми являются акустические колебания, похожие на реальные РС не только по временным и спектральным характеристикам, но и по восприятию на слух. При использовании РПС в качестве помех их следует и называть *речеподобными помехами* (РПП).

Для активной защиты речи от несанкционированного прослушивания применяются помехи самого различного рода [1–10]. Среди них РПП в последнее время получили значительное распространение. Классификация этих помех 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.

В работах [1–2, 8–10, 25, 26] содержится несколько определений РПП и описаний способов их формирования.

В [8] — это "речеподобный шум".

В [9] маскирующий сигнал "создают из исходного маскируемого сигнала путем модуляции шумовым сигналом моментов пересечения нуля маскируемым сигналом". Отмечается, что: "Модуляцию моментов пересечения нуля речевого сигнала можно производить путем *фазовой модуляции речевого сигнала шумовым* (курсив наш)". Нельзя не заметить, что в выделенной курсивом части цитаты из [9] говорится о неизвестном виде "фазовой" модуляции широкополосного (по признаку (ширина спектра)/(средняя частота спектра)) речевого сигнала.

В [10] упоминаются "речевой хор", состоящий из суммы нескольких речевых сигналов, и "речеподобный шум, алгоритм синтеза которого теоретически обоснован, представлен аналитическим выражением, адаптирован под первые три форманты интегрального формантного спектра русского языка".

В [25] маскирующее РС воздействие именуется, как «шумовая "речеподобная" помеха (шум с огибающей амплитудного спектра, подобной речевому сигналу)». Отмечается, что: "В соответствии с требованиями Государственной технической комиссии при Президенте Российской Федерации генератор помех должен формировать шумовые колебания в диапазоне частот от 175 до 5600 Гц".

По сведениям [26] для маскирования РС "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 и чем они подобны РС остается не ясным.

По понятным причинам от разработчиков нельзя ожидать полной прозрачности освещения принципов формирования предлагаемых ими маскирующих помех. Вместе с тем, по видимому, обоснованным является интерес к вопросу — насколько при восприятии на слух и по другим признакам вновь разрабатываемые РПП подобны речевым сигналам? Ведь именно от этого подобия, в конечном счете, зависят и маскирующая способность таких воздействий, и возможности очистки от них защищаемых сигналов.

Примеры видимого терминологического и методологического неблагополучия в обсуждаемой области речевых технологий, определенно, могут быть умножены. Как уже отмечалось, назрела настоятельная необходимость изменения этого положения.

В работах [4–5] предложен способ синтеза РПС, формируемого по случайному закону и по своим основным временным, спектральным характеристикам и восприятию на слух максимально подобного РС, но не содержащего смысловой информации.

В [5] для формирования РПС предложен алгоритм, основанный на управлении системой синтеза речи по тексту. Блок-схема этого алгоритма приведена на рис. 2. Как видно из рисунка, временные характеристики длительностей пауз между фоноабзацами, фразами и синтагмами составляют соответственно 0,3, 0,8 и 1,5 с, соответственно. Эти величины являются среднестатистическими для русского языка.

Соответствие спектральных характеристик РПП характеристикам РС обеспечивается применением выбранной базы аллофонов, которая формируется из реальных РС диктора. Фонетические особенности РС русского языка учитываются использованием условных вероятностей слогов. Конкретизация этого учета приведена ниже.

Первоначально, применительно к блок-схеме алгоритма, представленной на рис. 2, исследовано несколько подходов к формированию "речеподобного текста", называемого ниже "псевдотекстом". Наиболее простой подход заключался в формировании псевдослов в соответствии с безусловными вероятностями появления букв русского алфавита [20]. Пример псевдотекста, сформированного таким способом, имеет вид *"омтгбусн гаднн ааб р чйдвдну-*

гейнбмтию а т иреърн пкс схлтиеикуятиюжд рл аюас лкпнеб прунциъ ьизъ тк нкнйбеобот еммявицтн ма гуиотдесад а твисьлсинжн ймоттегадзо еошедеола ивауа млиц лпянуауот рбиут обнем ре овегопккаоонодоз ье ондусооююерешро чмумлккшеусоыйп". Можно убедиться, что озвучивание такого текста мало напоминает естественную речь.

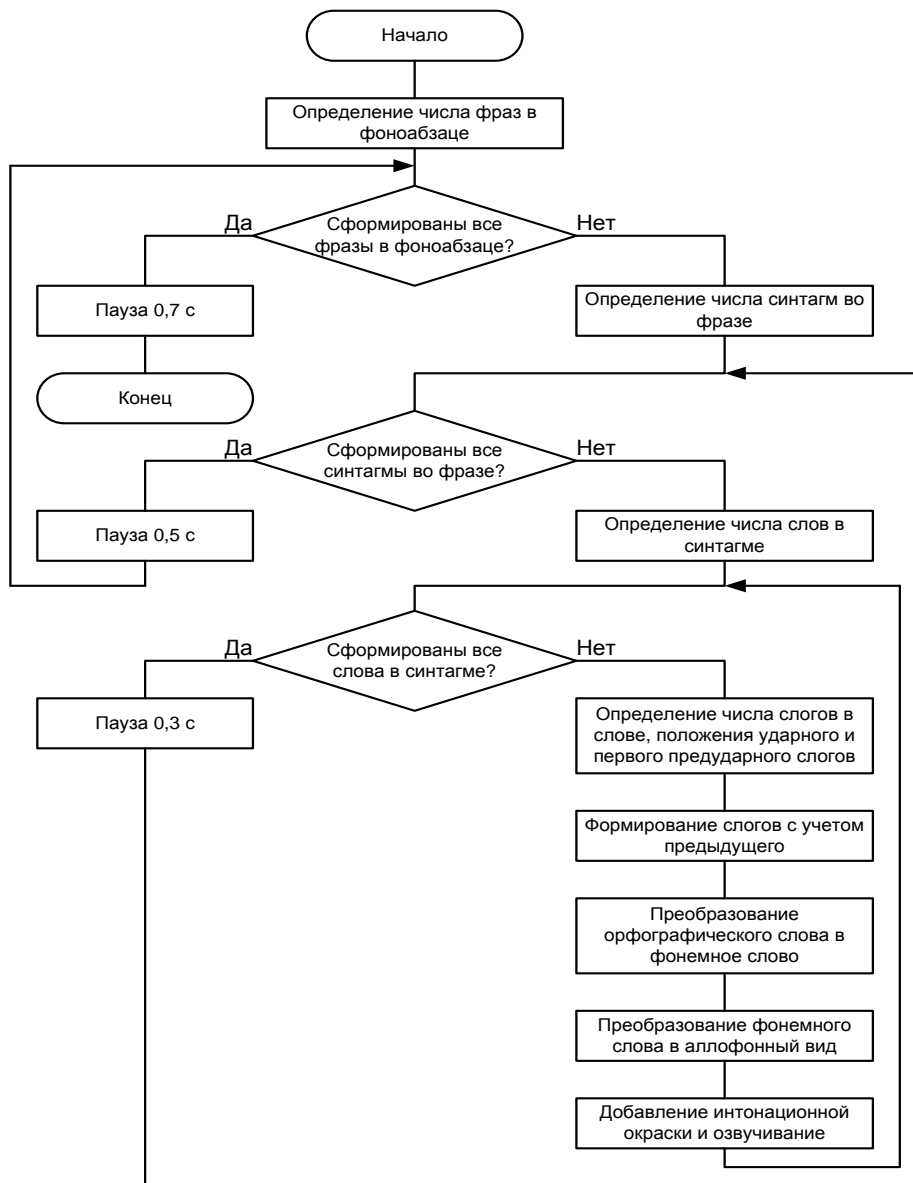


Рис. 2. Алгоритм формирования речеподобных сигналов

Для повышения сходства формируемых сигналов с речью был исследован способ формирования речеподобных сигналов, основанный на вероятностях появления звуков в различных частях слова [20]. Пример сформированного данным способом псевдотекста имеет вид "ним дликахаф шяпаныфехя ситья вапажихята ромахому ит сте еитопасьяти напроти нявето со фафо вуекийте понасеть невойеши рар пинееиком ласвайой самехай гамынимэ ту чиитисатьим тамыкакуф кепот е боноха к советаиш сынъ в". Синтезированный по данному тексту сигнал на слух оказывается более похожим на естественную речь, чем в предыдущем случае, но все еще не достаточно.

Псевдотекст, наиболее близкий к естественной речи, удалось получить при использовании вероятностей появления слогов русского языка. Для формирования таблицы безусловных вероятностей слогов большие объемы текстов (около $2 \cdot 10^9$ символов) были разбиты на слоги в соответствии с предложенными в [4, 5, 27] правилами. Несмотря на некоторую искусственность использованных правил, они позволили получить приемлемый результат: "тиебцеля по-

зовца амниваде фапенаяльно педостаеагомы всетоктие издере эне ко растогайе ма осоко незнидныонный стьеныйчно икдей мнокрорайв ворый чтопострой идныснейтса пгухйсите у тре аспозутьнтныя листрользува свяранививатся вбще золевнюа поктобрани нальшихнтао нетилняель о тимячкиткеюм атривова тинфибраво вымяфинибра длялямму".

Для дальнейшего повышения сходства формируемого по псевдотексту сигнала с речью были рассмотрены способы формирования псевдотекста по таблицам условных вероятностей встречаемости слогов и по нескольким таблицам безусловных вероятностей встречаемости начальных слогов, первых предударных слогов, ударных слогов и конечных слогов. Оба способа приводят к сходным результатам, лучшим, чем при формировании псевдотекста по безусловным вероятностям встречаемости слогов. Появление осмысленных слов в формируемом по случайному закону псевдотексте не является основанием для отказа от такого способа формирования РПП, так как известно, что эти слова не коррелируют с защищаемым РС.

Рассмотренный алгоритм смоделирован компьютерной программой синтеза РПП и реализован в устройстве защиты речевой информации "Прибой-Р".

Экспериментальная часть

Ниже приводится один фоноабзац псевдотекста РПП, сформированного в соответствии с рассмотренным выше алгоритмом (рис. 2).

"Себяще цезае́мо отнё. Науча́стро веденёюза обходётады. Дялю́щих. Перспе́циён определён запазываю́тсаяка. Соде́ржи едалию́ня рае́мен темнёдо.

Оначива́ети желтому́вел безан разноду наяганёчество. Видетро́ва обён развёиси гоо́браз. Госума иканалёис нес орноббще целенический явлён это́мско. Иеетсаячи утверочёньея естворётор ии́ци почтие. Этому́церлива приписат ужегодажен.

Образёем братазао́собсто. Послежитсаяче далосто средства́. Разёе отнюо́тдел игёни.

Бенга́ран общеёлапла котор. Возмо́ж это́мо яютсаячале ского́ распрёс. Каналие быст-рофа́ущес".

Временные реализации этого фоноабзаца, первой фразы и первой синтагмы представлены на рис. 3. Для формирования этих реализаций использовалась база аллофонов Б.М. Лобанова [23]. Частота основного тона в реализациях примерно равна 98 Гц.

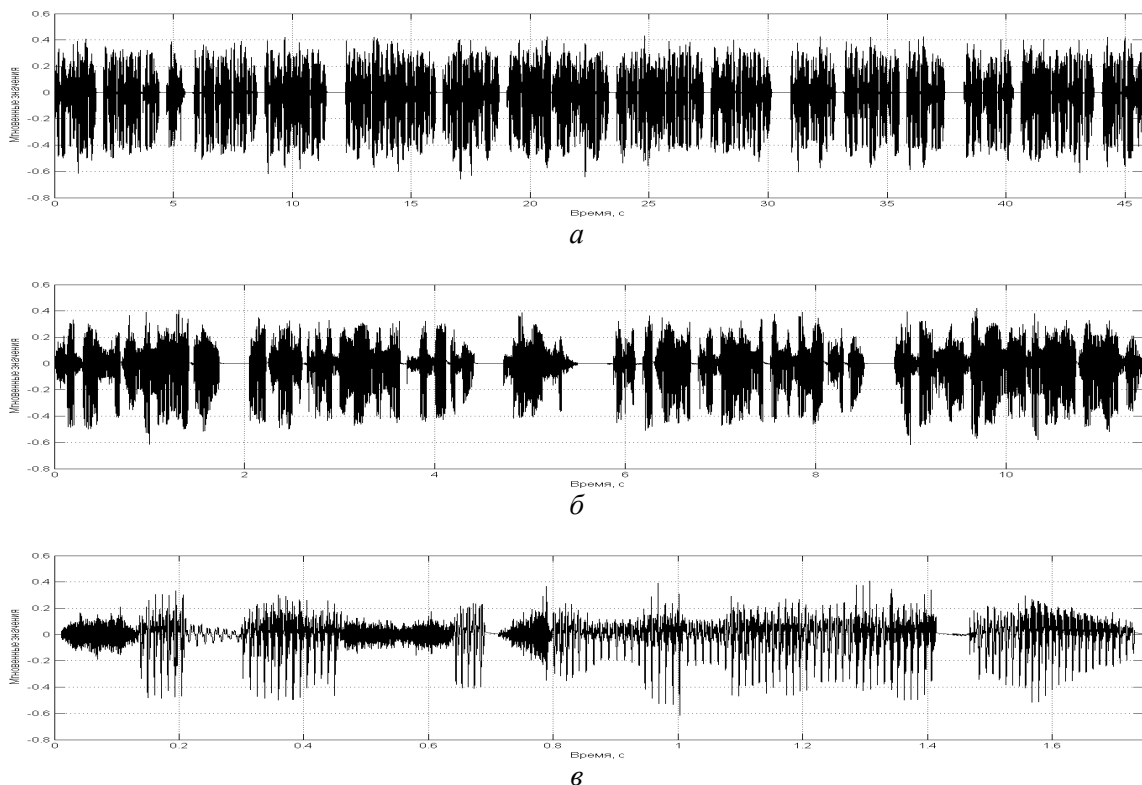


Рис. 3. Временные реализации фоноабзаца (а), фразы (б) и синтагмы (в) РПП

Оценки спектральной плотности мощности и автокорреляционной функции синтезированной РПП приведены на рис. 4.

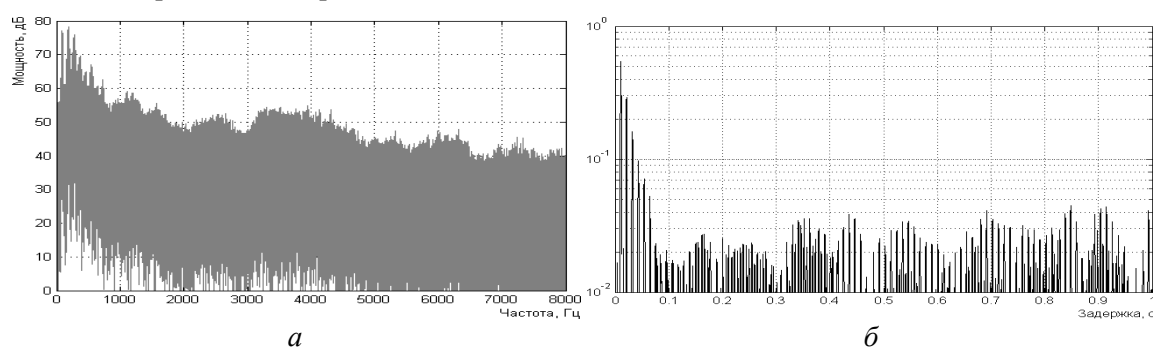


Рис. 4. Спектральная плотность мощности РПС для реализации общей длительностью 5 мин (а) и автокорреляционная функция для реализации длительностью 10 с (б)

Полученные к настоящему времени рабочие материалы позволяют заключить, что предложенный способ формирования РПП обеспечивает их приемлемое подобие реальным РС.

Устройства формирования РПП

К настоящему времени в Беларуси, Украине и Российской Федерации разработано и изготавливается значительное количество устройств активной защиты речевой информации от утечки по акустическому и вибрационному каналам. К таким устройствам относятся "Прибой", "Прибой-Р" (производятся в Беларуси), серия приборов "Базальт", "PSP-2А", "ПМ-2А" (производятся в Украине), "Барон", "Кабинет", "Шепот", "Заслон-2М", "Кедр", "Равнина-2К", "VNG-006", "VNG-012GL", "Скит-АР", "Скит-МВА", "Соната-АВ", "Шелест-4К", "Шорох-1", "Шорох-2", "ЛГШ-301", "ЛГШ-402", "ЛГШ-403", "SI-3001", "SI-3010", "SI-3030" (производятся в России).

Формируемые в соответствии с обсужденными здесь требованиями и предложенным определением РПП используются в устройстве "Прибой-Р" и последней модели устройства "Барон". Ранее (2005 г.) в устройстве "Барон" в качестве РПП применялась помеха в виде смеси трех сигналов радиовещательных станций.

Заключение

В современных публикациях содержатся самые различные употребления терминов "речеподобный сигнал — speech-like signal". Это приводит к информационной дезориентации потребителей речевых технологий. Для термина «"речеподобная" помеха» наиболее приемлемым определением в настоящее время является формулировка, приведенная в п. 3.8 "речеподобная" помеха предварительного стандарта СТБ П 34.101.28-2007. По мнению авторов, редакция этой формулировки может быть улучшена следующим образом: **"Синтезируемый по случайному закону акустический сигнал, который по своим основным характеристикам соответствует речевому сигналу, но не содержит смысловой информации"**.

SPEECH-LIKE SIGNALS: VARIETY, KEY PARAMETERS, WAYS OF GENERATION, APPLICATIONS WAY

V.I. VARABYEU, A.G. DAVYDAU, G.U. DAVYDAU

Abstract

The question on modern application of terms "speech-like" signals and noises is analyzed. Definition for the term "speech-like" noise is offered. The algorithm of speech-like noise generation corresponding to offered definition is presented.

Литература

1. Иванов В.М., Хорев А.А. // Вопросы защиты информации. 1999. № 4.
2. Хорев А.А. Способы и средства защиты информации. М., 2000.
3. Железняк В.К., Макаров Ю.К., Хорев А.А. // Специальная техника. 2000. № 4. С. 39–45.
4. Воробьев В.И., Давыдов А.Г. // XI сессия Российского акустического общества: Сб. тр., Москва, 19–23 ноября 2001 г. / М., 2001. Т. 3. С. 83–86.
5. Воробьев В.И., Давыдов А.Г., Лобанов Б.М. // XIII сессия Российского акустического общества: Сб. тр. Н. Новгород, 25–29 августа 2003 г. / М., 2003. Т. 3. С. 110–114.
6. Устройство защиты речевой информации от утечки по вибрационным и акустическим каналам. Пат. Респ. Беларусь № 3053. МПК7 H 04K 3/00, G 10K 11/00 / В.И. Воробьев, А.Г. Давыдов, Г.В. Давыдов, А.И. Ивонин, Д.В. Лещенко, Б.М. Лобанов, Л.М. Лыньков, В.А. Попов, А.В. Потапович // Официальный бюл. / Нац. центр интеллектуальной собственности. 2006. № 5. С. 184.
7. СТБ П 34.101.28-2007 Информационные технологии. Средства защиты речевой информации виброакустические. Классификация и общие технические требования.
8. Куницын, И.В., Лобашев А.К. // Бюро научно-технической информации. [Электронный ресурс]. 2000. Режим доступа: <http://www.bnti.ru/showart.asp?aid=867&lvl=03.03.04>. Дата доступа: 21.12.2008.
9. RU 2308159 C2 10.10.2007 Бюл.28: Способ формирования речеподобного маскирующего сигнала // Авторы и патентообладатели Железняк В.К., Червинский В.М.
10. Прокофьев М., Журавлев В. // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. 2006. № 2 (13).
11. ITU-T Recommendation P.501 (08/96): Telecommunication standardization sector of International Telecommunication Union (ITU) / Series P: Telephone transmission quality: Objective measuring apparatus. Test signals for use in telephony.
12. Sircar P., Prasad K., Harshavardhan B. // Proc. of XIV European Signal Processing Conf., Florence, Italy, 4–8 September, 2006.
13. Malliopoulos C., Bacamidis S. // Institute of language and speech processing [Electronic resource]. — Mode of access: <http://www.ilsp.gr/papers1en.html>. Date of access: 21.12.2008.
14. Krot A.M., Prokudina N.N. // Digital Signal Processing: Proceedings of 13th International Conference, Conference Centre "P.M. Nomikos" Santorini, Hellas, Greece. 2–4 July 1997. Vol. 2. P. 587–590.
15. Павловец А.Н., Зубрыцки П., Петровский А.А. // Докл. БГУИР. 2007. № 4. С. 19–34.
16. Азаров И.С., Петровский А.А. // Речевые технологии. 2008. № 1. С. 67–77.
17. Воробьев В.И., Давыдов Г.В., Шамгин Ю.В. // Докл. БГУИР. 2006. № 2. С. 64–68.
18. Фант Г. Акустическая теория речеобразования. М., 1964.
19. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. / Пер. с англ. под ред. М.В. Назарова и Ю.Н. Прохорова. М., 1981.
20. Сапожков М.А. Речевой сигнал в кибернетике и связи. М., 1963.
21. Евдокимова В.В., Смирнов В.А., Фроликов Л.С. // Совет молодых ученых СПбГУ [Электронный ресурс]. 2008. Режим доступа: http://www.sovmu.spbu.ru/main/sno/uzmf2/uzmf2_20.pdf. Дата доступа: 21.12.2008.
22. Баляйтите Д. // Вильнюсский университет [Электронный ресурс]. 2008. Режим доступа: http://filologija.vukhf.lt/5-10/doc/2.91%20Balsaityte%20RED_VM.doc. Дата доступа: 21.12.2008.
23. Лобанов Б.М., Елисеєва О.Е. Речевой интерфейс интеллектуальных систем: Учеб. пособие / Под науч. ред. В.В. Голенкова. Минск, 2006.
24. Выготский Л.С. Мышление и речь. М., 1999.
25. Хорев А.А. Макаров Ю.К. // Специальная техника [Электронный ресурс]. 2008. Режим доступа: <http://st.ess.ru/publications/articles/horev/horev.htm>. Дата доступа: 21.12.2008.
26. Хорев А.А. // Безопасность информационных технологий [Электронный ресурс]. 2008. Режим доступа: http://www.security.ukrnet.net/d-book-2/ch_10.pdf. Дата доступа: 21.12.2008.
27. Воробьев В.И., Давыдов А.Г. // Акустический журнал. 2002. Т. 48, № 5. С. 701–704.