

Сергей В. Дворянкин¹, Антон О. Антипенко²
Финансовый университет при Правительстве Российской Федерации
(Финансовый университет),
Ленинградский пр-кт, 49, Москва, 125167, Россия
¹e-mail: SVdvoryankin@fa.ru, <https://orcid.org/0000-0001-6908-0676>
²e-mail: An-go-55@yandex.ru, <https://orcid.org/0000-0001-8692-6637>

ПРИМЕНЕНИЕ ФАЗОВЫХ ХАРАКТЕРИСТИК ГОЛОСОВЫХ ВОКАЛИЗМОВ
В РЕШЕНИИ ЗАДАЧ ЗАЩИТЫ РЕЧЕВОЙ ИНФОРМАЦИИ

DOI: <http://dx.doi.org/10.26583/bit.2021.2.02>

Аннотация. В настоящее время из-за своего удобства речевые технологии приобретают всё большую популярность, в том числе применительно к системам обеспечения безопасности, однако из-за возможных проблем с безопасностью остаются сферы, в которых данная технология находит крайне ограниченное применение. **Цель** работы состоит в уточнении области применения фазовых характеристик голосовых вокализмов при решении различных задач в области защиты и обработки речевой информации. **Методы.** В работе использовались методы компьютерного и математического моделирования, цифровой обработки сигналов и изображений. **Результаты.** Предложено использовать фазовые характеристики речи в различных приложениях, в том числе для решения различных проблем безопасности, например, надёжной идентификации. **Выводы.** Представлены основные характеристики речевых сигналов. Рассмотрена Гильбертовская модель речевого сигнала, а также модель синусоидального описания речевого сигнала МакАуэля и Куатъери. Приведены области практического применения фазовых характеристик голосовых вокализмов (включая внедрение в фазограмму биометрической информации о говорящем).
Ключевые слова: голосовые вокализмы, защита речевой информации, разборчивость, речевой сигнал, фазовые характеристики, синусоидальная модель.

Для цитирования: ДВОРЯНКИН, Сергей В.; АНТИПЕНКО, Антон О. ПРИМЕНЕНИЕ ФАЗОВЫХ ХАРАКТЕРИСТИК ГОЛОСОВЫХ ВОКАЛИЗМОВ В РЕШЕНИИ ЗАДАЧ ЗАЩИТЫ РЕЧЕВОЙ ИНФОРМАЦИИ. *Безопасность информационных технологий*, [S.l.], v. 28, n. 2, p. 21–33, 2021. ISSN 2074-7136. Доступно на: <<https://bit.mephi.ru/index.php/bit/article/view/1337>>. Дата доступа: 09 apr. 2021. DOI: <http://dx.doi.org/10.26583/bit.2021.2.02>.

Sergey V. Dvoryankin¹, Anton O. Antipenko²
Financial University under the Government of the Russian Federation
(Financial University),
Leningradsky prospect, 49, Moscow, 125167, Russia
¹e-mail: SVdvoryankin@fa.ru, <https://orcid.org/0000-0001-6908-0676>
²e-mail: An-go-55@yandex.ru, <https://orcid.org/0000-0001-8692-6637>

**Applying the phase characteristics of voice vocalisms in solving problem of protection
of speech information**

DOI: <http://dx.doi.org/10.26583/bit.2021.2.02>

Annotation. At present, due to its convenience, speech technologies are gaining increasing popularity, including in relation to security systems. However, due to possible security problems, the areas in which this technology is extremely limited still remain. **The aim** of this work is to clarify the field of application of the phase characteristics of voice vocalisms in solving various problems of protection and processing of speech information. **Methods.** The methods of computer and mathematical modeling, digital processing of signals and images are used. **Results.** It is proposed to use the phase characteristics of speech in various applications, including security problems such as a reliable identification. **Conclusion.** The major characteristics of speech signals are presented. The Hilbert model of the speech signal is considered, as well as the model of the sinusoidal description of the speech signal by MacAuel and

Quatieri. The areas of practical application of the phase characteristics of voice vocalisms (including the introduction of biometric information about the speaker into the phaseogram) are described.

Keywords: voice vocalisms, speech information protection, intelligibility, speech signal, phase characteristics, sinusoidal model.

For citation: DVORYANKIN, Sergey V.; ANTIPENKO, Anton O. Applying the phase characteristics of voice vocalisms in solving problem of protection of speech information. IT Security (Russia), [S.l.], v. 28, n. 2, p. 21–33, 2021. ISSN 2074-7136. Available at: <<https://bit.mephi.ru/index.php/bit/article/view/1337>>. Date accessed: 09 apr. 2021. DOI: <http://dx.doi.org/10.26583/bit.2021.2.02>.

Введение

В настоящий момент количество сервисов, в которых могут найти применение речевые технологии, постоянно возрастает – это связано, главным образом, с лёгкостью их освоения и применения пользователями, а также сопутствующими удобствами [1]. Хорошим примером может служить внедрение функции голосовой аутентификации в приложениях дистанционного банковского обслуживания (ДБО), однако в следствие угрозы перехвата и утечки вводимой голосом ключевой информации, а также недостаточной надёжности данная технология не находит широкого применения.

Вместе с тем до настоящего времени из-за нехватки знаний о природе образования, а также сложности вычислений исследователями не уделялось должного внимания фазовым характеристикам речевого сигнала, хотя их применение позволит не только повысить надёжность голосовой аутентификации, но и решить некоторые другие прикладные задачи, связанные с речевой обработкой, а именно улучшения качества передачи голосовых сигналов в каналах речевой связи и защиты речевой информации.

1. Речевой сигнал и основные характеристики речевых вокализов

Будем считать, что речь – это исторически сложившаяся форма общения людей посредством языковых конструкций, создаваемых на основе определённых правил [2]. Множество исследований в части разработки высокоскоростных и надёжных систем обработки речи показали, что главным фактором их успеха является применение подходящего способа представления соответствующих сигналов, а также математических моделей.

Базовым компонентом речи являются звуки, которые, благодаря распространению на некоторые, иногда весьма значительные, расстояния в различных средах, например, в воздухе или строительных конструкциях, переносят информацию, воспринимаемую далее органами слуха человека либо различными средствами приёма акустических сигналов, в том числе разведки [3]. В настоящее время выделяют и используют различные характеристики речевого сигнала, полезные для решения задач речевой обработки, из них чаще всего особо отмечают следующие:

- частотные характеристики;
- амплитудные характеристики;
- энергетические характеристики;
- временные характеристики;
- фазовые характеристики.

Амплитудные характеристики речевого сигнала находят своё отражение в громкости звука – субъективном восприятии его силы, которое во многом зависит от частоты звуковых колебаний, а также давления (интенсивности) звука, и в меньшей степени от тембра, локализации в пространстве, спектрального состава и длительности [4, 5].

В соответствии с Международной организацией по стандартизации сон является единицей абсолютной шкалы громкости. Согласно стандарту, громкостью равной 1 сон называют громкость чистого непрерывного синусоидального тона частотой в 1 кГц и создающего звуковое давление равное 2 мПа. Часто уровень звукового давления выражается не в паскалях, а в децибелах (дБ) – это отношение величины звукового давления P к некоторому пороговому значению, а именно:

$$P_0 = 2 * 10^{-5} \text{ Па,}$$
$$P_{\text{дБ}} = 20 \lg \frac{P_{\text{Па}}}{P_0} \text{ дБ.}$$

Кроме того, существует ещё и относительная величина громкости звука – его уровень, выражаемый в фонах, который равен звуковому давлению в 1 дБ, создаваемому чистым непрерывным синусоидальным тоном частотой 1 кГц равногромкому измеряемому звуку.

На рис. 1 представлены изофоны – кривые равной громкости. Их значения определяются международным стандартом ISO 226 (русский аналог – ГОСТ Р ИСО 226–2009 «Акустика. Стандартные кривые равной громкости») и отражают зависимость уровня звукового давления от частоты при определённом уровне громкости. Применяя изофоны можно легко сопоставить уровень создаваемого звукового давления и уровень громкости тона определённой частоты. Допустим дана синусоидальная волна с частотой в 1000 Гц, которая создаёт звуковое давление в 20 дБ. Используя приведённые выше изофоны легко понять, что данный звук имеет громкость в 20 фон. Порог слышимости стандартного человеческого уха составляет ноль фон – на изофонах на рис. 1 он изображён пунктирной линией [6].

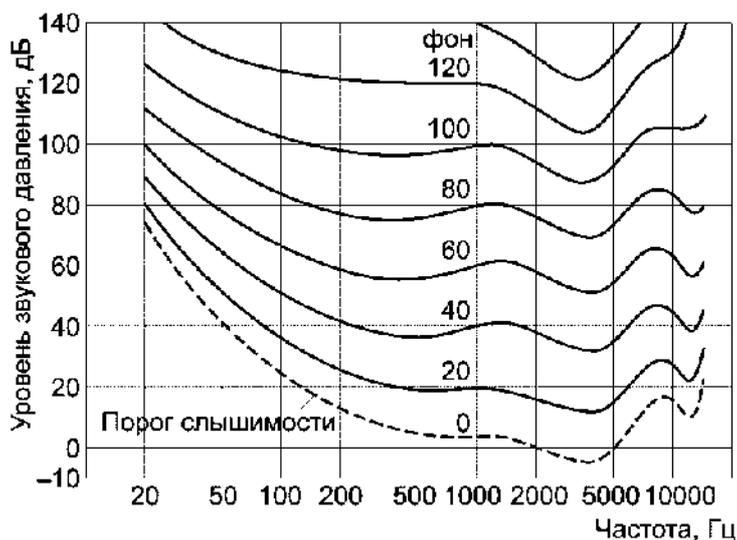


Рис. 1. Зависимость уровня громкости от звукового давления и частоты
Fig. 1. Dependence of the volume level on sound pressure and frequency

Темпом речи называется скорость произнесения элементов речи, таких как слова, слоги, звуки [7]. Эта характеристика речевого сигнала является одним из компонентов интонации и дополнительным элементом речи, при этом она сильно зависит от эмоционального состояния говорящего, его особенностей строения голосового аппарата, стиля произношения элементов речи, а также ситуации общения. Темп речи может

измеряться как средней длительностью элемента речи, так и числом произносимых элементов в момент времени [8]. Снижение темпа речи как правило достигается за счёт увеличения длительности гласных звуков, в то время как повышение осуществляется сокращением длительности и согласных, и гласных. Кроме того, темп речи используется и в качестве интонационного средства, позволяя выделить более важные моменты речи – как правило она становится более медленной, в то время как второстепенные моменты произносятся быстрее. Замедление темпа также характерно к концу высказывания, позволяя, тем самым, оформить его целостность. Вместе с тем темп можно рассматривать как характеристику степени слуховой отчётности и артикуляционной напряжённости, например, при быстром темпе речи слова обычно начинают выступать в неполных звуковых формах.

Частотный диапазон голоса человека измеряется в герцах, у мужчин он составляет от 80 до 150 Гц, у женщин от 120 до 400 Гц, у детей он сильно выше, отметим при этом, что диапазон разговорного голоса человека составляет только около 10% от общего диапазона голоса [9]. Заметим, что из-за особенностей человеческого слуха высокие голоса (с высокой частотой) воспринимаются более громко, нежели низкочастотные. Интересно, что окраска голоса, отражающаяся в его частоте, может служить индикатором не только эмоционального, но и психического состояние говорящего.

К энергетическим характеристикам речи относят интенсивность звука, а также плотность энергии. Количество энергии, проходящее в единицу времени через единицу площади, перпендикулярной к направлению распространения, называют силой звука или, по-другому, его интенсивностью, измеряется показатель в Вт/м². Интенсивность речевых сигналов зависит от амплитуды колебания голосовых связок человека и их напряжённости. Интенсивность звука снижается в случае уменьшения амплитуды колебаний. При этом различают несколько уровней интенсивности – от высокого до низкого. Интенсивность может быть как постоянна на отрезке времени, так и динамически меняться (плавно или резко). Под плотностью звуковой энергии подразумевают количество энергии акустических колебаний, находящейся в единице объёма (Дж/м³).

Фазовым характеристикам речи, которые более подробно описаны в следующем разделе настоящего исследования, до настоящего времени, за исключением нескольких работ [10, 11], практически не уделялось внимания учёных, в первую очередь из-за сложности их вычисления, а также недостаточности знаний об их образовании, при этом их анализ и применение может кардинально повысить надёжность голосовой аутентификации, в том числе в приложениях дистанционного банковского обслуживания. Наибольший интерес представляет оценка фазовых характеристик на вокализованных отрезках речевого сигнала.

2. Гильбертовская модель и фаза речевого сигнала

В целях выделения характеристик речевого сигнала и последующего приведения его к подходящей форме для анализа-синтеза разработано большое количество различных математических моделей, таким образом, что каждая из моделей удобна для решения своей узкой технической задачи [12, 13]. Для целей защиты речевой информации чаще всего применяются модели искусственного моделирования начальных фазовых характеристик, а также описания вокализованных участков речи. В настоящее время преобразование Фурье играет ключевую роль в области исследования акустического сигнала и является наиболее популярным методом его представления в частотной области [13]. При этом необходимо отметить существенный недостаток многих математических

моделей речевого сигнала – зачастую они не учитывают фазовые характеристики речевого сигнала.

Для исключения вышеназванного недостатка в статье предлагается использовать преобразования Гильберта и Фурье. Например, вокализованные участки речи можно представить следующим выражением [12]:

$$S(t) = \sum_{k=1}^{\infty} a_k(t) \cos k2\pi ft,$$

где f – частота основного тона, $a_k(t)$ – амплитуда k -ой гармоники основного тона, k – номер гармоники.

При этом для невокализованных звуков используют иную формулу:

$$S(t) = \int_0^{\infty} a_{\omega}(t) \cos \omega t dt,$$

где $a_{\omega}(t)$ – спектральная амплитудная плотность на частоте ω .

В общем случае принято использовать другое описание речевого сигнала [13]:

$$S(t) = a(t) \cos \varphi(t),$$

где $a(t)$ и $\varphi(t)$ – мгновенная амплитуда и мгновенная фаза соответственно, определяемые преобразованиями В.И. Коржикова, Гильберта или В.И. Тихонова [14].

С помощью преобразования Гильберта речь можно представить в аналитическом виде и таким образом найти значения параметров, отвечающих за разборчивость (опорных точек), однако при таком подходе фаза не является определённой по времени и частоте, что сильно усложняет вычисление начальной фазы. Таким образом традиционные модели описания речевых сигналов не подходят для решения задач вычисления и дальнейшего анализа фазовых характеристик, фактически для построения эффективного процесса верификации и распознавания голосов требуется нахождение и использование нового алгоритма.

Тем не менее речь на вокализованных участках можно представить как суперпозицию акустических узкополосных по Гильберту [12]:

$$s(t) = \sum s_i(t) = \sum A_i(t) \cos \varphi_i(t).$$

И затем для всех обертонов речи необходимо вычислить полную фазу $\varphi_i(t)$ по Гильберту согласно формуле:

$$\varphi_i(t) = \arctg \left(\frac{s_i(t)}{s_i^c(t)} \right) + 2k\pi,$$

где k – неопределённое целое число, а $s_i^c(t)$ – сопряжённая функция от $s_i(t)$ по Гильберту.

Представим полную фазу i -го обертона в виде:

$$\varphi_i(t) = \omega_i t + \theta_i(t) + \varphi_{0i},$$

где ω_i принимает кратные частоте основного тона значения, φ_{0i} – начальная фаза i -го обертона речи, $\theta_i(t)$ отвечает за нелинейную зависимость фазы от времени.

Одной из самых важных задач при исследовании фазовых характеристик голосовых вокализов является определение зависимости изменения фазы от времени.

Для определения данной зависимости нужно взять производную второго порядка от косинуса фазы, тогда:

$$\frac{d(\cos \varphi_i(t))}{dt} = \sin \varphi_i(t) \frac{d\varphi_i(t)}{dt},$$
$$\frac{d^2(\cos \varphi_i(t))}{dt^2} = -\cos \varphi_i(t) \left(\frac{d\varphi_i(t)}{dt} \right)^2 + \sin \varphi_i(t) \frac{d^2 \varphi_i(t)}{dt^2}.$$

Результаты ранее проведённых экспериментов по нахождению второй производной от $\cos \varphi_i(t)$ явно указывают на то, что в окрестности точек пересечения нуля, где $\cos \varphi_i(t) \approx 1$, $\sin \varphi_i(t) \approx 1$ значение производной стабильно, практически не изменяется на коротком промежутке времени. Таким образом нелинейная зависимость фазы от времени на коротком промежутке имеет параболический вид:

$$\theta_i(t) = \frac{\delta_i}{2} t^2,$$

где $\delta_i = \frac{d^2 \varphi_i(t)}{dt^2}$.

Выявленная зависимость позволяет полную фазу каждого обертона сигнала приблизить к следующему выражению:

$$\varphi_i(t) = \omega_i t + \delta \frac{t^2}{2} + \varphi_{0i}.$$

Принимая во внимание вышеизложенное, общую формулу описания речевого сигнала на вокализованных участках можно представить в следующем виде:

$$s(t) = \sum_K A_k e^{\alpha_k t} \cos \left(\omega_k t + \delta_k \frac{t^2}{2} + \varphi_{0k} \right).$$

Эта модель хорошо показывает себя в решении задач идентификации говорящего (например, в речевых системах контроля и управления доступом) благодаря возможности оценки как потерь в речевом тракте (коэффициент α_k), так и частотно-фазовых параметров (коэффициенты $\omega_k, \delta_k, \varphi_{0k}$).

Очевидно, что главным прототипом для создания будущих эффективных информационных систем обработки и защиты речевой информации являются механизмы человеческого речевосприятия и речеобразования, вместе с тем некоторые из них так до конца и не изучены. Исследование данных механизмов позволит создавать новые функциональные модели анализа-синтеза речевой информации. Одним из примеров использования такого механизма является международный стандарт сжатия цифрового аудио и видеосигналов MPEG-4, в основу которого положен эффект скрытия энергетически слабых звуковых сигналов громкими [15]. Эти методы также могут быть приложены и к решению задач считывания, визуального представления и корректировки речевой информации.

Добавим, что к показателям, определяющим качество фиксируемой речи, относят её громкость, естественность, а также разборчивость. Показатели громкости речи уже были подробно разобраны в предыдущей части исследования, добавим лишь, что при определении качества этот показатель сравнивается с нормальным уровнем порога слышимости человеческого уха. Натуральность (естественность) речи является субъективным показателем качества, поскольку до настоящего времени надёжных математических методов оценки этого показателя не существует, однако существует

предположение, подкреплённое некоторым количеством научных работ, согласно которому этот показатель скрывается именно в фазовых характеристиках голосовых вокализов, при этом интересна их оценка на вокализованных участках речевого сигнала, как наиболее инвариантных к различным преобразованиям в системах обработки и передачи речи. Чаще всего качество речи измеряется её разборчивостью, которая является статистической характеристикой и определяется отношением числа правильно понятых речевых конструкций (слов, звуков, слогов, фраз) к их общему переданному по каналу количеству, при этом используется среднее значение, допускающее некоторые отклонения. Для различных работ, в том числе и по защите речевой информации, чаще всего применяют слоговую разборчивость, при этом существуют и другие типы, такие как фразовая и словесная. Все они измеряются в процентах и связаны определёнными соотношениями [16].

3. Модель синусоидального описания речевого сигнала МакАуэля и Куатъери

Наиболее исчерпывающее объяснение синусоидальных моделей описания речевых сигналов, которые также включают низкие частоты, были представлены в исследованиях МакАуэля и Куатъери. Синусоидальная модель, предложенная МакАуэлем и Куатъери, представляет речь как линейную комбинацию синусоид с изменяющимися во времени амплитудами, фазами и частотами [17]:

$$S_{SR}(n) = \sum_{K=1}^L A_k \cos(\Omega_k n + \varphi_k)$$

где SR обозначает синусоидальное представление.

Заметим, что число синусоид L изменяется во времени. Возможность уменьшить скорость передачи данных с использованием этой модели связана с тем, что голосовая речь как правило высокочастотна и, следовательно, она может быть представлена ограниченным набором синусоид. К тому же статическая структура (кратковременный спектр) шёпота (глухой речи) может быть представлена синусоидальной моделью с соответственно определёнными случайными фазами. Синусоидальная модель может быть связана с моделью системы-источника посредством замены упрощённой модели возбуждения голосового тракта на более общую модель, содержащую зависящую от L синусоидальных компонент, состоящих из амплитуд, частот и фаз. Выход из голосовой полосы фильтра (синтетическая речь) в стационарном состоянии может быть записан в соответствии с формулой. Основным предположением является то, что параметры синусоидальной модели медленно изменяются во времени по отношению к длительности импульса голосового тракта.

МакАуэль и Куатъери показали, что высококачественное восстановление может быть достигнуто путём использования синусоид с амплитудами, частотами и фазами, соответствующими пикам коротковременного преобразования Фурье. Ширина окна Хемминга, равная 2,5 средним высотам, подходит и гарантирует, что синусоидальные волны хорошо определены. Более того, синусоидальная модель слабо зависит от высоты и голоса начиная со средней высоты используется только для определения длины анализируемого окна.

Синусоидальная модель была представлена выше в наиболее общей форме. Основные вклады в работе МакАуэля и Куатъери лежат в анализе минимального параметра синусоидальной модели, а также в разработке алгоритма отслеживания синусоидальных параметров от фрейма к фрейму. Прежде всего, поскольку число синусоид меняется с высотой, было установлено понятие «жизни» и «смерти»

синусоидальных компонентов для обеспечения соответствия динамических параметров. В добавок к этому были разработаны новые алгоритмы интерполяции фазы и амплитуды для соответствия этим параметрам от одного фрейма к другому. Эксперименты с синусоидальной моделью показали, что целых 80 синусоид могут быть использованы для синтеза. Эти эксперименты проводились при использовании адаптированного окна Хемминга шириной в 2,5 от средней высоты, и 1024 точки быстрого преобразования Фурье, которое обновляется каждые 10 мс. Синусоидальная анализо-синтезирующая модель работает достаточно хорошо с большим количеством сигналов (несколько динамик, музыка, биологические звуки), а также с речью в присутствии фонового шума.

Для низкочастотных приложений частоты синусоидальных волн могут быть ограничены, в итоге они получаются целыми, кратными основной частоте, то есть [17]:

$$S_{\widetilde{HR}}(n) = \sum_{k=1}^{L(\Omega_0)} \cos(\varphi_k + k\Omega_0 n) A_k,$$

где $L(\Omega_0)$ – количество гармоник интересующей речевой полосы частот (обычно 4 кГц), Ω_0 – частота основного тона, а \widetilde{HR} означает гармоническое представление сигнала. Гармоническое представление обеспечивает оптимальное множество частот только для идеально звучащих сегментов. Основное предположение в голосовой речи – шаг периода постоянен в течение всего периода анализа окна. Для не голосовой речи, множество синусоид, равноудалённых по частоте, как правило сохраняют статистику не голосовых сегментов. Предположение для не голосовой речи – частоты синусоид достаточно близки так что они следуют за изменениями измеряемой кратковременной спектральной плотностью.

Простой пример реконструкции сегмента голосовой речи при помощи линейной комбинации гармонических синусоид показан на рис. 2. Голосовой сегмент, сформированный с помощью 32 мс прямоугольных окон и амплитуд фаз синусоидальных волн, оценивался по пикам сегмента дискретного преобразования Фурье (рис. 3).

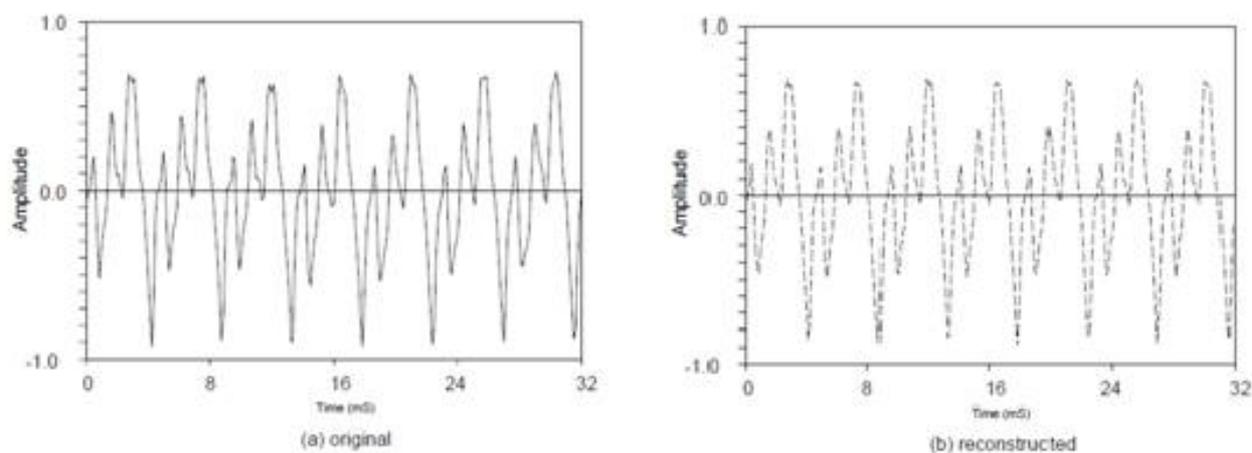


Рис. 2. Восстановление сигнала по набору гармоник
Fig. 2. Reconstruction of a signal from a set of harmonics

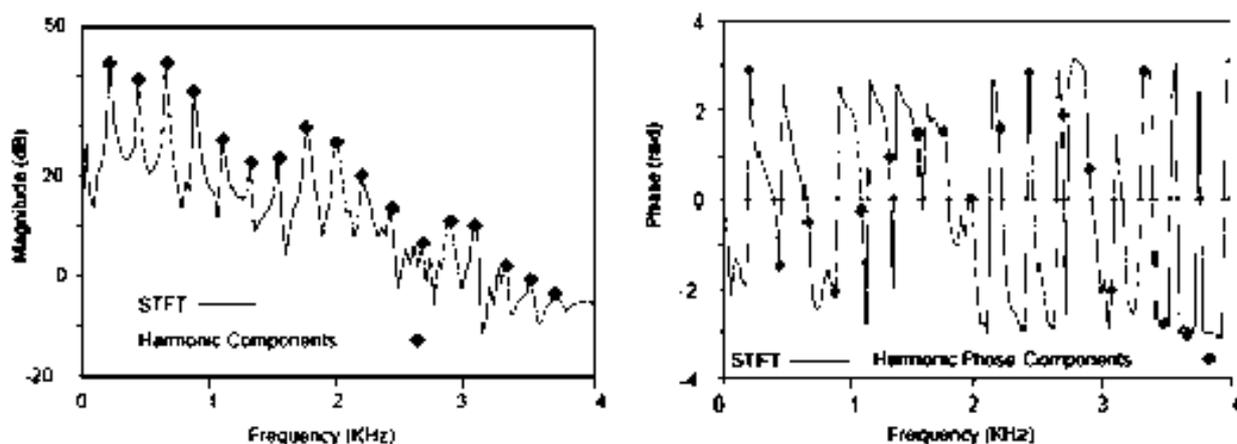


Рис. 3. Амплитуды и фазы участка речевого сигнала, выделенные с помощью дискретного преобразования Фурье

Fig. 3. Amplitudes and phases of a section of a speech signal isolated using a discrete Fourier transform

4. Области практического применения фазовых характеристик голосовых вокализов

Из-за сложности современных информационных систем, и, вследствие этого, широчайшего спектра возможных угроз в том числе речевой информации, информационная безопасность таких систем требует от проектировщика и разработчика применения комплексного подхода к вопросам защиты информации. Перед построением системы защиты необходимо разработать модель угроз и нарушителя, при этом в общем случае закладывается предположение, что злоумышленник обладает всеми возможными ресурсами, в том числе техническими и интеллектуальными, для осуществления несанкционированного доступа к конфиденциальной информации, циркулирующей в информационной системе и каналах связи [18].

На рис. 4 приведены некоторые методы защиты речевой информации от угроз целостности, конфиденциальной и доступности. Например, для купирования угрозы фейковых аудиозаписей (когда голосом известного человека озвучивают заранее подготовленный, обычно провокационный, текст), которые сейчас набирают популярность, можно использовать технологию речевой подписи. Для реализации данной технологии предлагается использовать фазовые характеристики голосовых вокализов. На рис. 5 приведён пример построенной фазограммы с наложением спектрограммы, фактически это является неким графическим изображением звукового сигнала и позволяет его рассмотреть в многомерном пространстве – со стороны классической частоты и времени, а с другой стороны фазы и мощности (в градациях серого).

Кроме того, удобным такой формат представляется, в том числе, потому, что в него можно встроить биометрическую информацию о говорящем, например, рукописную подпись (рис. 6) или изображение отпечатка пальца.



Рис. 4. Методы защиты речевой информации
Fig. 4. Methods for protecting speech information

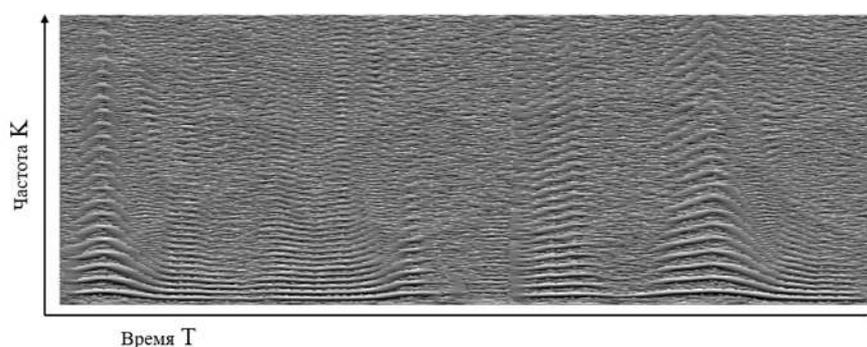


Рис. 5. Пример фазограммы с наложением спектрограммы
Fig. 5. An example of a phase chart with a spectrogram overlay

Применение фазовых характеристик речи позволит не только переосмыслить принципы использования технологии речевой подписи, но и решить некоторые другие прикладные задачи, связанные с речевой обработкой, в частности существенно повысить надёжность голосовой аутентификации и распознавания речи, а также улучшить защиту

речевого сигнала от несанкционированного доступа в выделенных защищаемых помещениях.

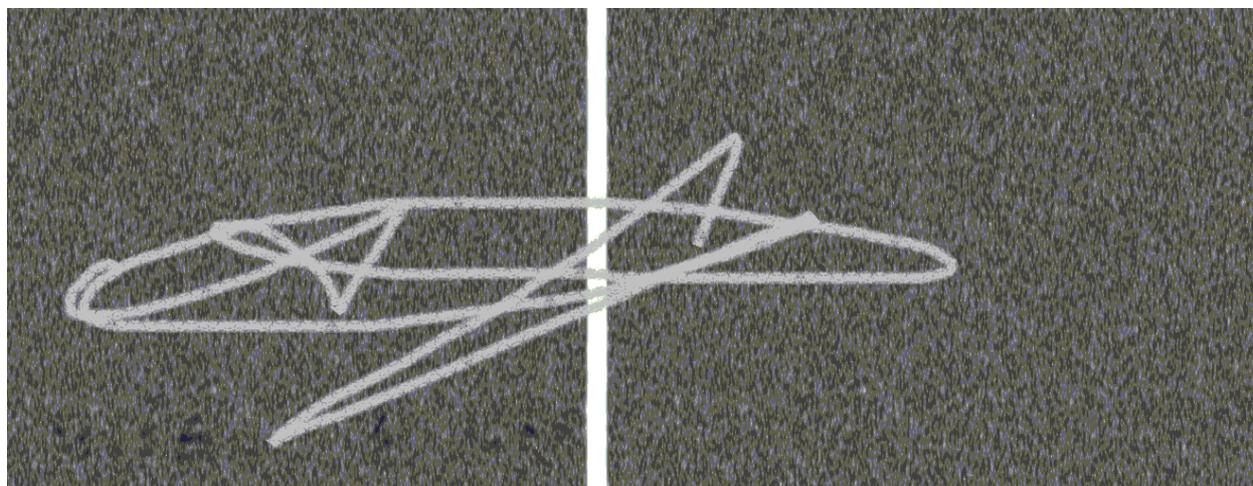


Рис. 6. Пример фазограммы с наложением рукописной подписи
Fig. 6. An example of a phasogram with overlaid handwritten signature

Заключение

В статье подробно представлены основные характеристики речевых сигналов (частотные, амплитудные, энергетические, временные и фазовые). Разобрана Гильбертовская модель речевого сигнала (в том числе для вокализованных и невокализованных звуковых сигналов) и принцип вычисления фазы по Гильберту. Выведена общая формула описания речевого сигнала на вокализованных участках. Представлена модель синусоидального описания речевого сигнала МакАуэля и Куатьери. Приведены примеры реконструкции сегмента голосовой речи при помощи линейной комбинации гармонических синусоид.

В заключение необходимо отметить, что фазовые характеристики голосовых вокализов, а равно результаты настоящего исследования, полученные согласно поставленной цели, могут найти широкое применение при построении удобных и безопасных систем голосовой аутентификации говорящего (например, внедрение в фазограмму биометрии), а также в автоматических системах анализа, искусственного синтеза и распознавания речи.

СПИСОК ЛИТЕРАТУРЫ:

1. Федоринин М. Они нас слышат: куда развиваются речевые технологии? URL: <https://www.forbes.ru/tekhnologii/331035-oni-nas-slyshat-kuda-razvivayutsya-rechevye-tekhnologii> (дата обращения: 25.03.2021).
2. Жинкин Н.И. Речь как проводник информации. М.: Наука, 1982. – 159 с.
3. Хорев А.А. Технические каналы утечки акустической (речевой) информации // Специальная техника. 2009. № 5. С. 12–26. URL: <https://elibrary.ru/item.asp?id=13011605> (дата обращения: 25.03.2021).
4. 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, [S.l.], Т. 26, № 1. С. 64–76, 2019. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/1186> (дата обращения: 25.03.2021). DOI: <http://dx.doi.org/10.26583/bit.2019.1.07>.
5. Козлачков С.Б., Дворянкин С.В., Бонч-Бруевич А.М. Принципы формирования тестовых речевых сигналов при оценках эффективности технологий шумочистки // Вопросы кибербезопасности. 2018.

- № 3 (27). С. 9–15. URL: <https://cyberleninka.ru/article/n/printsipy-formirovaniya-testovyh-rechevyh-signalov-pri-otsenkah-effektivnosti-tehnologiy-shumoochistki> (дата обращения: 25.03.2021).
6. Gelfand S.A. Hearing: An introduction to psychological and physiological acoustics. 2nd edition. New York and Basel: Marcel Dekker, Inc., 1990. – 488 p.
 7. Ахманова О.С. Словарь лингвистических терминов. М.: КомКнига, 2007. – 607 с.
 8. Кармин Г. Презентации в стиле TED: 9 приёмов лучших в мире выступлений. М.: Альпина Паблишер, 2015. – 253 с.
 9. Вартанян И.А. Звук — слух — мозг. Ленинград: Наука, 1981. – 176 с.
 10. Дворянкин, Сергей В. и др. Системное моделирование речеподобных сигналов и его применение в сфере безопасности, связи и управления // Безопасность информационных технологий. [S.l.], Т. 26, № 4. С. 101–119, 2019. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/1236> (дата обращения: 25.03.2021). DOI: <http://dx.doi.org/10.26583/bit.2019.4.08>.
 11. Митянок В.В., Коновалов Л.В. Применение фазового анализа звуков речи для распознавания человека по его голосу // Техническая акустика. 2013. № 4. С. 1–10. URL: <https://elibrary.ru/item.asp?id=21084028> (дата обращения: 25.03.2021).
 12. Оппенгейм А.В. Применение цифровой обработки сигналов. М.: Мир, 1980. – 552 с.
 13. Сергиенко А.Б. Цифровая обработка сигналов. СПб.: Питер, 2003. – 604 с.
 14. Коржик В.И. Расширенное преобразование Гильберта и его применения в теории сигналов // Проблемы передачи информации. 1969. Т. 5, № 4. С. 3–18. URL: http://www.mathnet.ru/php/archive.phtml?wshow=paper&jmid=ppi&paperid=1817&option_lang=rus (дата обращения: 25.03.2021 года).
 15. Touradj E., Fernando P. The MPEG-4 Book. Prentice Hall Professional, 2002. – 896 p.
 16. Покровский Н.Б. Расчёт и измерение разборчивости речи. М.: Связьиздат, 1962. – 390 с.
 17. R. McAulay and T. Quatieri, Speech analysis/Synthesis based on a sinusoidal representation, in IEEE Transactions on Acoustics, Speech, and Signal Processing. V. 34, no. 4. P. 744–754, August 1986. DOI: <http://dx.doi.org/10.1109/TASSP.1986.1164910>.
 18. Антипенко А.О. О методах обеспечения информационной безопасности в системах речевой связи // Modern Science. 2019. № 5–3. С. 187–195. URL: <https://elibrary.ru/item.asp?id=38250230> (дата обращения: 25.03.2021).

REFERENCES:

- [1] Fedorinin M. They hear us: where are speech technologies developing? URL: <https://www.forbes.ru/tekhnologii/331035-oni-nas-slyshat-kuda-razvivayutsya-rechevye-tekhnologii> (accessed: 25.03.2021) (in Russian).
- [2] Zhinkin N.I. Speech as a conductor of information. М.: Nauka, 1982. – 159 p. (in Russian).
- [3] Horev A.A. Technical channels of acoustic (speech) information leakage. Special'naja tehnika, 2009, no. 5. P. 12–26. URL: <https://elibrary.ru/item.asp?id=13011605> (accessed: 25.03.2021) (in Russian).
- [4] Dvoryankin, Sergey 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, no. 1. P. 64–76, 2019. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/1186> (accessed 25.03.2021). DOI: <http://dx.doi.org/10.26583/bit.2019.1.07> (in Russian).
- [5] Kozlachkov S.B., Dvoryankin S.V., Bonch-Bruevich A.M. Principles of Formation of Test Speech Signals in Evaluating the Effectiveness of Noise Cleaning Technologies. Voprosy kiberbezopasnosti, 2018, no. 3 (27), P. 9–15. URL: <https://cyberleninka.ru/article/n/printsipy-formirovaniya-testovyh-rechevyh-signalov-pri-otsenkah-effektivnosti-tehnologiy-shumoochistki> (accessed: 25.03.2021) (in Russian).
- [6] Gelfand S.A. Hearing: An introduction to psychological and physiological acoustics. 2nd edition. New York and Basel: Marcel Dekker, Inc., 1990. – 488 p.
- [7] Ahmanova O.S. Dictionary of linguistic terms. М.: KomKniga, 2007. – 607 p. (in Russian).
- [8] Karmin G. TED Presentations: 9 Tricks of the World's Best Speaking. М.: Al'pina Publisher, 2015. – 253 p. (in Russian).
- [9] Vartanjan I.A. Sound - hearing - brain. Leningrad: Nauka, 1981. – 176 p. (in Russian).
- [10] 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, no. 4. P. 101–119, 2019. ISSN 2074-7136. URL: <https://bit.mephi.ru/index.php/bit/article/view/1236>. (accessed: 25.03.2021). DOI: <http://dx.doi.org/10.26583/bit.2019.4.08> (in Russian).

- [11] Mitjanok V.V., Kononov L.V. Application of phase analysis of speech sounds to recognize a person by his voice. *Tehnicheskaja akustika*, 2013, no. 4. P. 1–10. URL: <https://elibrary.ru/item.asp?id=21084028> (accessed: 25.03.2021) (in Russian).
- [12] Oppengejm A.V. *Digital Signal Processing Applications*. M.: Mir, 1980. – 552 p. (in Russian).
- [13] Sergienko A.B. *Digital signal processing*. SPb.: Piter, 2003. – 604 p. (in Russian).
- [14] Korzhik V.I. Extended Hilbert transform and its applications in signal theory. *Problems of Information Transmission*, 1969. V. 5, no. 4. P. 1–14. URL: http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=ppi&paperid=1817&option_lang=rus (accessed: 25.03.2021) (in Russian).
- [15] Touradj E., Fernando P. *The MPEG-4 Book*. Prentice Hall Professional, 2002. – 896 p.
- [16] Pokrovskij N.B. *Calculation and measurement of speech intelligibility*. M.: Svjaz'izdat, 1962. – 390 p. (in Russian).
- [17] R. McAulay and T. Quatieri, Speech analysis/Synthesis based on a sinusoidal representation, in *IEEE Transactions on Acoustics, Speech, and Signal Processing*. V. 34, no. 4. P. 744–754, August 1986. DOI: <http://dx.doi.org/10.1109/TASSP.1986.1164910>.
- [18] Antipenko A.O. On methods of ensuring information security in voice communication systems. *Modern Science*, 2019, no. 5–3. P. 187–195. URL: <https://elibrary.ru/item.asp?id=38250230> (accessed: 25.03.2021) (in Russian).

*Поступила в редакцию – 29 января 2021 г. Окончательный вариант – 01 апреля 2021 г.
Received – January 29, 2021. The final version – April 01, 2021.*