



ФАЗОВЫЕ ХАРАКТЕРИСТИКИ ГОЛОСОВЫХ ВОКАЛИЗМОВ КАК СЛЕДУЮЩИЙ ВОЗМОЖНЫЙ ЭТАП РАЗВИТИЯ СИСТЕМ ЗАЩИТЫ РЕЧЕВОЙ СВЯЗИ

В статье отмечается, что экспоненциальное увеличение мощности вычислительных систем, в том числе домашнего использования, а также стремительное развитие алгоритмов машинного обучения, особенно в части анализа и синтеза речи, уже привели к появлению новых относительно доступных видов угроз обеспечению конфиденциальности и целостности системам речевой связи. Вместе с тем использование удобных речевых технологий в некоторых специфических задачах, например, биометрической аутентификации говорящего, сильно ограничено из-за возникающих проблем безопасности. В статье подробно рассмотрены характеристики речевых сигналов и приведены их основные модели (Гильбертовская модель, а также модель синусоидального описания речевого сигнала МакАуэля и Куатьери). Автор предлагает обратить внимание на фазовые характеристики голосовых вокализов как на один из возможных путей купирования новых угроз системам речевой связи. Внимание при написании статьи уделено их практическому использованию – описаны основные области применения, а также представлено программное обеспечение, позволяющее анализировать и использовать фазовые характеристики во многих задачах защиты речевой информации. В выводах автор подчёркивает необходимость продолжения исследования возможной области применения фазовых характеристик, в том числе для надёжной аутентификации диктора.

Ключевые слова: голосовые вокализмы, защита речевой информации, разборчивость, речевой сигнал, фазовые характеристики, синусоидальная модель.

PHASE CHARACTERISTICS OF VOICE VOCALISMS AS THE NEXT POSSIBLE STAGE IN THE DEVELOPMENT OF SPEECH COMMUNICATION PROTECTION SYSTEMS

The article notes that the exponential increase in the power of computing systems, including home use, as well as the rapid development of machine learning algorithms, especially in terms of speech analysis and synthesis, have already led to the emergence of new relatively affordable types of threats to confidentiality and integrity of speech communication systems. At the same time, the use of convenient speech technologies in some specific tasks, for example, biometric speaker authentication, is severely limited due to emerging security problems. The article discusses in detail the characteristic of speech signals and presents their main models (Hilbert's model and the model of the sinusoidal description of the speech signal by McAulay and Quatieri). The author proposes to pay attention to the phase characteristics of voice vocalisms as one of the possible ways to stop new threats to speech communication systems. Attention when writing this article is intended for their practical use - the possibility of using software that allows you to analyze and use phase characteristics in many problems of protecting speech information. In the conclusions, the author emphasizes the need to continue the study of a possible field of application of phase characteristics, including for reliable speaker authentication.

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

Введение

В последние несколько лет в мире наблюдается взрывной рост количества сервисов, использующих в своём составе речевые технологии – это могут быть как распространённые виртуальные ассистенты, способные качественно понимать человеческую речь и синтезировать голосовые ответы на простые вопросы, так и относительно новые системы, способные в интерактивном режиме общаться с человеком (заказать столик в ресторане, записать в парикмахерскую) [1]. Такие технологии легки в освоении и удобны для пользователя. В большинстве своём рост качества анализа и синтеза речи связан с развитием области нейронных сетей, а также экспоненциальным увеличением мощности обрабатывающих серверов.

Вместе с тем речевые технологии могут применяться злоумышленниками для нарушения конфиденциальности и целостности систем речевой связи [2]. При этом использование фазовых характеристик голосовых вокализов позволит не только купировать эти угрозы, но и добиться повышения качества синтеза речи, а также возможности использовать речевые технологии в новых областях, например, для голосовой аутентификации диктора.

Речь и основные показатели речевых сигналов

Одним из основополагающих аспектов развития человечества стало использование устной речи. Она позволяет передавать прошлый и настоящий общечеловеческий опыт, то есть, фактически, обмениваться знаниями и

достижениями между людьми. Пусть речь – форма общения, сложившаяся под воздействием совокупности исторических правил построения языковых конструкций [3]. Главным компонентом устной речи являются звуки. Благодаря их эффективному распространению в среде передачи, например в воздухе, мы органами слуха (либо иными приёмниками) можем улавливать некоторую информацию.

Из многих характеристик речевого сигнала чаще всего выделяют следующие основные:

- 1) Временные характеристики;
- 2) Частотные характеристики;
- 3) Амплитудные характеристики;
- 4) Энергетические характеристики.

Под временными характеристиками речи зачастую понимают её темп – скорость произношения различных элементов речи, таких как звуки, слоги, слова [3]. Темп речи считается одним из важных компонентов интонации (существенные части разговора стараются произносить медленнее, чем второстепенные моменты), он сильно зависит от эмоциональной окраски разговора, а также особенностей речевого аппарата оратора. Обычно его измеряют либо количеством произнесённых элементов речи в момент времени, либо средней продолжительностью произнесения

отдельного элемента речи. Следует отдельно отметить, что при быстрой речи часто снижается её разборчивость.

Частотная характеристика (диапазон) речевого сигнала среднестатистического человека лежит в области от 120 до 400 герц для женщин и от 80 до 150 герц для мужчин. Вместе с тем основную информацию несут лишь 10% от этих интервалов – данный диапазон называют диапазоном разговорного голоса [4]. Анализ частотного диапазона индивидуума позволяет косвенно оценить как его психическое, так и эмоциональное состояние.

Амплитудные характеристики речи отражаются на громкости звука, то есть, фактически, на субъективном восприятии его силы. Международной организацией по стандартизации (ISO) введена единица абсолютной шкалы громкости – сон, который равен громкости чистого непрерывного синусоидального тона частотой в 1 кГц и создающего звуковое давление равное 2 мПа. На практике чаще в качестве величины звукового давления используют не паскали, а децибелы (дБ):

$$P = 2 * 10^{-5} \text{ Па}$$

$$P_{\text{дБ}} = 20 \lg \frac{P_{\text{Па}}}{P_0} \text{ дБ}$$

Исходя из вышеприведённых формул под децибелами понимают отношение величины

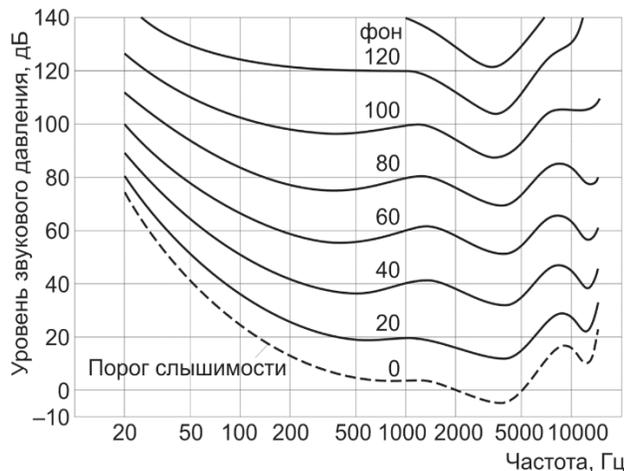


Рис. 1. Зависимость уровня громкости от звукового давления и частоты

звукового давления P к некоторому пороговому значению. Существует также и относительная величина громкости звука, выражаемая в фонах. Согласно российскому стандарту ГОСТ Р ИСО 226–2009 «Акустика. Стандартные кривые равной громкости» (аналог международного стандарта ISO 226) вводятся изофоны – кривые равной громкости, значения которых схематично представлены на рисунке 1.

Представленный график отражает зависимость уровня звукового давления от частоты при определённом уровне громкости, позволяя, тем самым, легко сопоставить уровень создаваемого звукового давления и уровень громкости тона определённой частоты. Например, нам дана звуковая волна с частотой 500 Гц и уровнем звукового давления в 120 дБ, тогда по графику легко определить, что громкость звука равна 120 фон. Под

нулём фон понимают порог слышимости человеческого уха, на графике его изображают пунктирной линией [5].

В группу энергетических характеристик речевого сигнала обычно включают плотность энергии и интенсивность звука. Сумму энергии звуковых колебаний, находящейся в единице объёма, называют плотностью энергии и измеряют в Дж/м³. Количество энергии, проходящей за единицу времени через единицу площади, перпендикулярной к направлению распространения, называют интенсивностью (силой) звука – её измеряют 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 низкого до высокого. Интенсивность может быть как постоянна на отрезке времени, так и динамически меняться (плавно или резко).

Вместе с тем к основным характеристикам речи зачастую не относят фазовые характеристики, о которых и пойдёт речь далее – до недавнего времени исследователи практически не уделяли им внимания, за исключением нескольких работ [6]. Это связано со сложностью их вычисления и неясностью перспектив практического использования. Однако некоторые косвенные данные свидетельствуют, что включение именно этих характеристик в соответствующие математические модели речи позволит существенным образом повысить точность этих моделей, расширить их область применения, а также решить некоторые другие задачи защиты речевой информации.

Модели описания речевых сигналов

В целях выделения характеристик речевого сигнала и последующего приведения его к подходящей форме для анализа-синтеза разработано большое количество различных математических моделей, таким образом каждая из моделей удобна для решения своей узкой технической задачи.

Например, для целей искусственного синтеза голоса широкое распространение получили нейролингвистические модели, основанные на описании и моделировании участков речи с применением преобразования Фурье [7]. Однако необходимо отметить их существенный недостаток (как и подавляющего большинства других математических моделей речевого сигнала) – они либо вовсе

не учитывают фазовые характеристики, либо моделируют их искусственно.

Для исключения вышеназванного недостатка авторы предлагают использовать преобразования Гильберта. В результате вокализованные участки речи можно представить следующим выражением [8]:

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

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

Для невокализованных звуков необходимо использовать другую формулу:

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

в которой – спектральная амплитудная плотность на частоте .

Кроме того, часто используют иное описание речевого сигнала [8]:

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

в котором $a(t)$ и $\varphi(t)$ – мгновенная амплитуда и мгновенная фаза соответственно, определяемые преобразованиями В.И. Коржикова, Гильберта или В.И. Тихонова.

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

Наиболее исчерпывающее объяснение синусоидальных моделей описания речевых сигналов, которые также включают низкие частоты, были представлены в исследованиях МакАуэля и Куатьери. Синусоидальная модель, предложенная МакАуэлем и Куатьери, представляет речь как линейную комбинацию синусоид с изменяющимися во времени амплитудами, фазами и частотами [9]:

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

где \overline{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 голо-

совая речь как правило высокопериодична и, следовательно, она может быть представлена ограниченным набором синусоид. Основным предположением является то, что параметры синусоидальной модели медленно изменяются во времени по отношению к длительности импульса голосовая тракта.

МакАуэль и Куатьери показали, что высококачественное восстановление речевого сигнала может быть достигнуто путём использования синусоид с амплитудами, частотами и фазами, соответствующими пикам коротко-временного преобразования Фурье. Ширина окна Хемминга, равная 2.5 средним высотам, подходит и гарантирует, что синусоидальные волны хорошо определены. Более того голос, начиная со средней высоты, используется только для определения длины анализируемого окна.

Вышепредставленная синусоидальная модель описана в своей наиболее общей форме. Основные вклады в работе МакАуэля и Куатьери лежат в анализе минимального параметра синусоидальной модели, а также в разработке алгоритма отслеживания синусоидальных параметров от окна к окну. 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 разработаны новые алгоритмы интерполяции фазы и амплитуды для соответствия этим параметрам от одного окна к другому. Эксперименты с моделью показали, что при использовании адаптированного окна Хемминга шириной в 2.5 от средней высоты, и 1024 точки быстрого преобразования Фурье, которое обновляется каждые 10 мс, уже 80 синусоид могут быть использованы для синтеза голоса. Модель хорошо работает 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 [9]:

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

где $L(\Omega_0)$ – количество гармоник интересующей речевой полосы частот (обычно 4 кГц), Ω_0 – частота основного тона, а \overline{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.

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

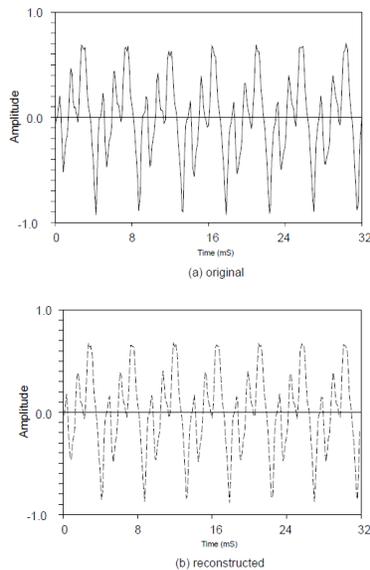


Рис. 2. Восстановление сигнала по набору гармоник

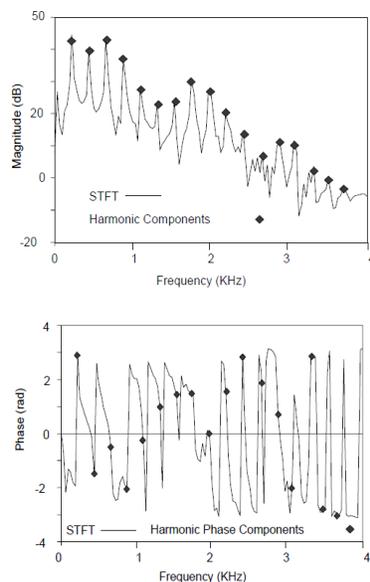


Рис. 3. Амплитуды и фазы участка речевого сигнала, выделенные с помощью дискретного преобразования Фурье

Алгоритм вычисления векторов приведённых начальных фаз на участках речевых вокализов

Одной из главных задач фоноскопических исследований речи является поиск таких параметров и признаков исследуемого фонообъекта, которые были бы инвариантны к реально встречающимся искажениям речевых сигналов. Наиболее актуально такая задача возникает при решении проблемы верификации личности говорящего [6].

Предложен следующий способ учёта вариативности акустического описания речевого сигнала. На вокализованных участках протяжённостью Δt в точках анализа с шагом r , речевой сигнал $S_r(t)$ может быть представлен суммой M его составляющих гармоник по формулам, представленным ранее. Для данного временного участка речи Δt и точки анализа r вводятся понятия вектора начальных фаз φ_r и вектора приведённых начальных фаз $\overline{\varphi}_r$:

$$\varphi_r = \begin{bmatrix} \varphi_{M-1} \\ \varphi_{M-2} \\ \vdots \\ \varphi_1 \\ \varphi_0 \end{bmatrix}$$

$$\overline{\varphi}_r = \begin{bmatrix} \varphi_{M-1} - \varphi_0 \\ \varphi_{M-2} - \varphi_0 \\ \vdots \\ \varphi_1 - \varphi_0 \\ 0 \end{bmatrix}$$

В качестве опорной начальной фазы в $\overline{\varphi}_r$ можно взять фазу любой гармонической составляющей речевого сигнала. Такое приведение фаз всех имеющихся на анализируемом участке гармоник к одной опорной необходимо для снятия неопределённости, связанной с выбором точки начала отсчёта при выполнении процедур анализа.

Выдвигаются следующие три гипотезы. Первая - вектор приведённых начальных фаз можно считать инвариантным на всём протяжении вокализованного участка и значения

приведённых фаз, кроме опорной, отличны от нуля. Вторая - вектор приведённых начальных фаз одинаков для одинаковых звуков и слов, произнесённых одним лицом. Третья - вектор приведённых начальных фаз одинаков для всех звукосочетаний данного лица.

Предлагается следующий метод нахождения вектора приведённых начальных фаз. Устранив из изначальной формулы $S_r(t)$ амплитуду гармоник, получаем следующее описание данного вокализованного участка речи протяжённостью Δt :

$$\overline{S}_r(t) = \sum_{k=0}^{M-1} \cos(\omega_k t + \varphi_k)$$

На рисунке 4 в полосе частот 4 кГц на временном интервале примерно равным 0.5 с при шаге анализа по времени 1.25 мс представлены сонограммы исходного сигнала $S_r(t)$ и сигнала $\overline{S}_r(t)$ с устранившей амплитудой гармоник.

Один из моментов времени t_0 на отрезке анализируемого вокализованного участка речи принимается за начало отсчёта. Далее строится система уравнений размерностью равной числу гармоник M , находящихся в интервале анализа:

$$\left\{ \begin{array}{l} \overline{S}_r(t_0) = \sum_{k=0}^{M-1} \cos(\omega_k t_0 + \varphi_k) \\ \overline{S}_r(t_0 + r) = \sum_{k=0}^{M-1} \cos(\omega_k(t_0 + r) + \varphi_k) \\ \vdots \\ \overline{S}_r(t_0 + (M-1) \cdot r) = \sum_{k=0}^{M-1} \cos(\omega_k(t_0 + (M-1) \cdot r) + \varphi_k) \\ \overline{S}_r(t_0 + M \cdot r) = \sum_{k=0}^{M-1} \cos(\omega_k(t_0 + M \cdot r) + \varphi_k) \end{array} \right.$$

Результатом решения этой системы являются вектора $\{\cos \varphi_r\}$ и $\{\cos \overline{\varphi}_r\}$. Процедура решения системы повторяется на всём протяжении вокализованного участка Δt для более точного определения вектора косинусов приведённых начальных фаз $\{\cos \overline{\varphi}_r\}$.

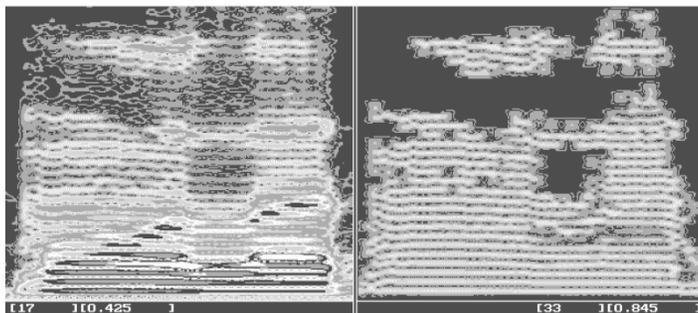


Рис. 4. Сонограмма исходного участка речи (левая половина) и сонограмма этого же участка с устранившим влиянием амплитудного спектра (правая половина)

Как пример выражение $\{\cos\overline{\varphi_r}\}$ может использоваться в качестве эталонного описания голоса в системах верификации - процесс будет состоять из сравнения заранее вычисленного эталонного вектора косинусов приведённых начальных фаз с вектором, вычисляемым в реальном времени из анализируемой речи.

Практическое применение фазовых характеристик голосовых вокализмов

Требования к комплексному подходу при проектировании и построении систем защиты информации, в том числе речевой, продиктован, прежде всего, широким спектром возможных угроз, а также постоянным усложнением как самих информационных систем, так и совершаемых злоумышленниками на них атак [6].

Угрозы речевой информации делятся на угрозы её доступности, целостности и конфи-

денциальности, некоторые методы защиты от них представлены на рисунке 5. В последние несколько лет в виду как активного развития алгоритмов машинного обучения, так и стремительного увеличения мощности вычислительных машин, особое внимание злоумышленников привлекает именно нарушение целостности речевой информации. Например, в 2020 году преступники с помощью технологии машинного обучения смоделировали голос одного из состоятельных клиентов банка ОАЭ, а затем осуществили звонок от его лица управляющему, который, узнав голос своего доверителя, без дополнительных проверок перевёл около 400 тысяч долларов США на сторонний банковский счёт [2]. Вместе с тем это далеко не единичная история – аналогичный случай произошёл в 2019 году в Великобритании, и тогда мошенникам удалось похитить около 240 тысяч долларов США [10].



Рис. 5. Различные методы защиты речевой информации от угроз целостности, конфиденциальности и доступности

При этом фазовые характеристики голосовых вокализмов могут быть применены для купирования многих угроз речевой информации, в том числе и приведённых выше, путём реализации их в соответствующих методах защиты. В частности, на основе эталонного вектора косинусов приведённых начальных фаз возможно построить технологию речевой подписи.

В целях проведения дальнейших исследований разрабатывается программный комплекс, позволяющий автоматизировано проводить анализ голосовых вокализмов, вычислять их фазовые характеристики (в том числе эталонный вектор начальных фаз), на основе полученных данных строить фазограммы с наложением спектрограмм, а также совершать различные звуковые преобразования (шумоо-

чистку). Программное обеспечение разрабатывается на языке программирования C++ с применением фреймворка Qt 5. Такой подход позволяет с одной стороны эффективно распоряжаться ресурсами автоматизированного рабочего места, а с другой стороны разраба-

тывать приложение, работающее на всех популярных платформах и поддерживающее построение интерфейса под различные разрешения и соотношения экранов. Снимок начального экрана программы в среде MicrosoftWindows представлен на рисунке 6.

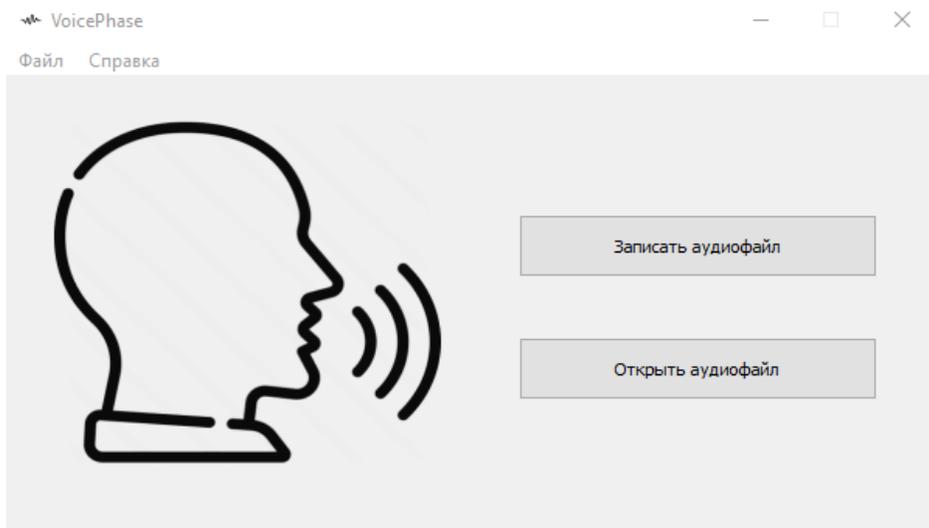


Рис. 6. Начальное окно программы в среде ОС MicrosoftWindows

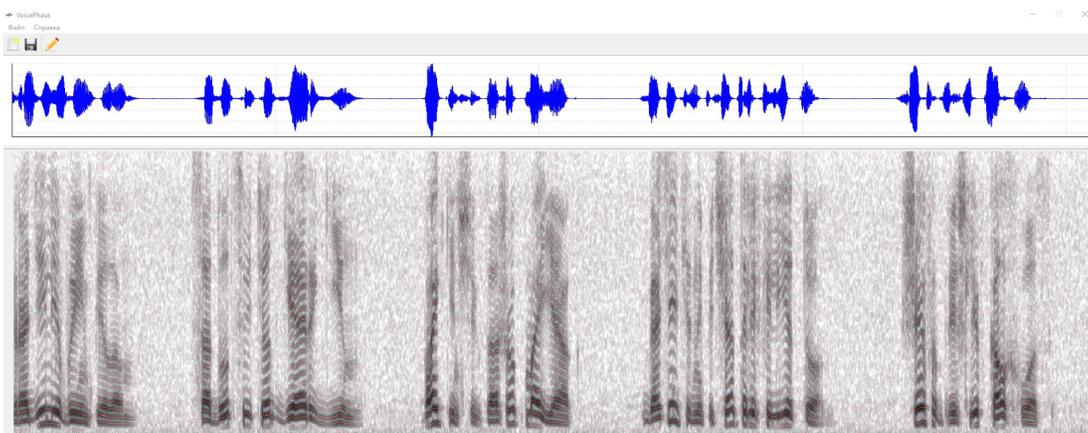


Рис. 7. Основное рабочее окно программы с анализом аудиофайла

На рисунке 7 приведён скриншот основного рабочего окна программы, в нижней его половине построена фазограмма аудиосигнала с наложением спектрограммы. Этот рисунок является многомерной визуализацией звука (в данном случае речи), поскольку кроме частоты и времени в нём в градациях серого выражены ещё мощность и фаза. Подобный формат представления кажется удобным, поскольку в него можно встроить биометрическую информацию о говорящем, например, рукописную подпись или изображение отпечатка пальца.

Авторы считают, что использование фазовых характеристик голосовых вокализмов

может дать новый толчок к развитию различных сервисов, использующих технологии синтеза и анализа речи – повысить надёжность и удобство использования речевой подписи, уточнить модели систем распознавания и синтеза голоса, усовершенствовать защиту выделенных для конфиденциальных переговоров помещений от прослушивания.

Заключение

В статье рассмотрены характеристики речевых сигналов и приведены их основные модели (Гильбертовская модель, а также модель синусоидального описания речевого сигнала МакАуэля и Куатъери). Описан алго-

ритм вычисления векторов приведённых начальных фаз на участках речевых вокализов, а также приведены примеры реконструкции сегмента голосовой речи при помощи линейной комбинации гармонических синусоид. Представлено специальное программное обеспечение, предназначенное для работы с фазовыми характеристиками голосовых вокализов.

Необходимо отметить, что фазовые характеристики голосовых вокализов могут найти широкое применение во многих прикладных областях – начиная с построения удобных и безопасных систем голосовой аутентификации говорящего, заканчивая системами автоматического анализа, а также синтеза голоса.

Литература

1. Chauhan S. Evolution of Speech Recognition Technology // Электронный ресурс <https://readwrite.com/2019/10/22/evolution-of-speech-technology> (дата обращения: 05.11.2021 года).
2. Quach K. Bank manager tricked into handing \$35m to scammers using fake 'deep voice' tech // Электронный ресурс https://www.theregister.com/2021/10/16/ai_in_brief/ (дата обращения: 05.11.2021 года).
3. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. Москва: РГГУ, 2001. 592 с.
4. Дидковский В.С. Акустическая экспертиза каналов речевой коммуникации. Монография. Киев: Имекс-ЛТД, 2008. 420 с.
5. Царегородцев А.В. Техническая защита информации. Москва: Финансовый университет при Правительстве РФ, 2014. 276 с.
6. Дворянkin С. В., Уленгов С. В., Устинов Р. А., Дворянkin Н. С., Антипенко А. О. Системное моделирование речеподобных сигналов и его применение в сфере безопасности, связи и управления // Безопасность информационных технологий. 2019. Т. 26, №4. С. 101-119.
7. Голубинский А.Н. Математические модели речевых сигналов для верификации и идентификации личности по голосу. Воронеж: Воронежский государственный университет, 2010. 363 с.
8. Коржик В. И. Расширенное преобразование Гильберта и его применения в теории сигналов // Проблемы передачи информации. 1969. Т. 5, №4. С. 3-18.
9. McAulay R. J., Quatieri T. F. Speech analysis/Synthesis based on a sinusoidal representation // Article in IEEE Transactions on Acoustics Speech and Signal Processing. 1986. ASSP-34(4). pp. 744-754.
10. Stupp C. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case // Электронный ресурс <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402> (дата обращения: 05.11.2021).

References

1. Chauhan S. Evolution of Speech Recognition Technology // Electronic resource <https://readwrite.com/2019/10/22/evolution-of-speech-technology> (accessed: 05.11.2021).
2. Quach K. Bank manager tricked into handing \$35m to scammers using fake 'deep voice' tech // Electronic resource https://www.theregister.com/2021/10/16/ai_in_brief/ (accessed: 05.11.2021).
3. Kodzasov S. V., Krivnova O. F. Obshchaya fonetika. Moskva: RGGU, 2001. 592 s.
4. Didkovskij V.S. Akusticheskaya ekspertiza kanalov rechevoj kommunikacii. Monografiya. Kiev: Imeks-LTD, 2008. 420 s.
5. Caregorodcev A. V. Tekhnicheskaya zashchita informacii. Moskva: Finansovyy universitet pri Pravitel'stve RF, 2014. 276 s.
6. Dvoryankin S. V., Ulengov S. V., Ustinov R. A., Dvoryankin N. S., Antipenko A. O. Sistemnoe modelirovanie rechapodobnyh signalov i ego primenenie v sfere bezopasnosti, svyazi i upravleniya // Bezopasnost' informacionnyh tekhnologij. 2019. T. 26, №4. S. 101-119.
7. Golubinskij A. N. Matematicheskie modeli rechevyh signalov dlya verifikacii i identifikacii lichnosti po golosu. Voronezh: Voronezhskij gosudarstvennyj universitet, 2010. 363 s.
8. Korzhik V. I. Rasshirennoe preobrazovanie Gil'berta i ego primeneniya v teorii signalov // Problemy peredachi informacii. 1969. T. 5, №4. C. 3-18.
9. McAulay R. J., Quatieri T. F. Speech analysis/Synthesis based on a sinusoidal representation. Article in IEEE Transactions on Acoustics Speech and Signal Processing, 1986, ASSP-34(4), pp. 744-754.
10. Stupp C. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case // Electronic resource

САВЕЛЬЕВ Иван Андреевич, кандидат технических наук, доцент Департамента информационной безопасности, Финансовый университет при Правительстве Российской Федерации. 125993, г. Москва, Ленинградский проспект, 49. E-mail: IASavelyev@fa.ru

АНТИПЕНКО Антон Олегович, аспирант Департамента информационной безопасности, Финансовый университет при Правительстве Российской Федерации. 125993, г. Москва, Ленинградский проспект, 49. E-mail: An-go-55@yandex.ru

SAVELYEV Ivan Andreevich, candidate of technical sciences, Associate Professor of the Department of Information Security, Financial University under the Government of the Russian Federation. 125993, Moscow, Leningradsky Prospekt, 49. E-mail: IASavelyev@fa.ru

ANTIPENKO Anton Olegovich, postgraduate student of the Department of Information Security, Financial University under the Government of the Russian Federation. 125993, Moscow, Leningradsky Prospekt, 49. E-mail: An-go-55@yandex.ru