УДК 057.087.1:621.391.26

ИССЛЕДОВАНИЕ ФАЗОВЫХ ХАРАКТЕРИСТИК ГОЛОСОВОГО СИГНАЛА ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ АУТЕНТИФИКАЦИИ

И.С. Павленко, Н.С. Пастушенко, О.Н. Файзулаева
Харьковский национальный университет радиоэлектроники

Abstract *– In the work, the modern systems of voice authentication have been analyzed and the directions for increasing their effectiveness have been defined. The use of the phase of the voice signal during digital processing can significantly improve the efficiency of modern authentication systems. The subject of the study is the process of digital signal processing in voice authentication systems. The scientific task of isolating and analyzing the phase information of the voice signal of the authentication system user is being solved. The purpose of the research is to study the phase characteristics of the voice signal and to identify their main features. The procedures for separating the phase signal against the noise of unknown intensity have been first developed and investigated, simulations have been performed and the operability of the indicated procedures has been shown in the presence of a priori information on the duration of the sawtooth signal. The accuracy of setting the signal duration is 10 counts. The received procedures make it possible to specify the beginning and the end of the phase signal, if necessary to compensate for anomalous measurements, and also to extract the frame of the voice signal from the noise sequence. The presented research results can be used in voice authentication systems as well as to improve the quality of speech recognition and speaker identification problems.*

Анотація *– Об'єктом дослідження є процес цифрової обробки сигналів у системах голосової аутентифікації. Вирішується наукове завдання виділення й аналізу фазової інформації голосового сигналу користувача системи аутентифікації. Ціль досліджень - аналіз фазових характеристик голосового сигналу й виявлення їхніх основних особливостей. Уперше розроблені й досліджені процедури виділення фазового сигналу на фоні шуму невідомої інтенсивності, проведено імітаційне моделювання й показане працездатність зазначених процедур при наявності апріорної інформації про тривалість пилкоподібного сигналу.*

Аннотация *– Объектом исследования является процесс цифровой обработки сигналов в системах голосовой аутентификации. Решается научная задача выделения и анализа фазовой информации голосового сигнала пользователя системы аутентификации. Цель исследований – анализ фазовых характеристик голосового сигнала и выявление их основных особенностей. Впервые разработаны и исследованы процедуры выделения фазового сигнала на фоне шума неизвестной интенсивности, проведено имитационное моделирование и показана работоспособность указанных процедур при наличии априорной информации о длительности пилообразного сигнала.*

Введение

В последнее время особо остро стоит проблема сохранности финансовых и информационных ресурсов, доступ к которым осуществляется с помощью сетевых технологий. Начальным барьером при сохранении различных ресурсов является система аутентификации, совершенствованию которой посвящено множество исследований в последние двадцать лет. Одно из основных направлений совершенствования систем аутентификации было связано с использованием статических биометрических признаков (внешний вид лица, папиллярный узор пальцев и радужная оболочка глаз). К сожалению, как показала практика, эти надежды не оправдались. Причины этого – низкие качественные характеристики и простота подделки анализируемого шаблона пользователя. В связи с этим, в последнее время основные исследования сосредоточены на использовании в системах аутентификации динамических

(поведенческих) биометрических признаков и, в первую очередь, голоса пользователя.

По отношению к иным биометрическим признакам, голосовые системы аутентификации имеют ряд существенных преимуществ, среди которых выделим. Простота и низкая стоимость оборудования, которое в настоящее время, как правило, установлено на всех используемых устройствах. Сложность подделки шаблона, который может оперативно изменяться и наращиваться в процессе принятия решения. Возможность применения всех достижений современной цифровой обработки сигналов, в том числе, и процедур последовательного анализа и др.

В современных голосовых системах аутентификации используются амплитудные и частотные характеристики голосового сигнала, а фазовая информация традиционно игнорируется [1]. Причиной этому может являться то, что современная модель акустической теории речеобразования представляется в виде взаимодействия спектра источника звука и спектральной характеристики фильтра, в роли которого выступает речевой тракт [1]. Очевидно, поэтому основные отличия пользователя (диктора) ищут в области амплитудно-частотного спектра голосового сигнала.

Однако давно известно, что фаза сигнала содержит больше информации, чем амплитуда и частота [2]. К сожалению, в настоящее время использованию фазовой информации для решения задач аутентификации пользователя уделяется недостаточно внимания. Среди известных исследований и публикаций можно выделить работы Воробьева В.И., который в течение десятилетия уделяет большое внимание использованию фазовой информации, в том числе и при решении задач в области акустики [3-5]. Вместе с тем, до настоящего времени не раскрыты основные особенности обработки фазовой информации голосового сигнала пользователя. Поэтому в работе решается научная задача разработки процедур выделения и анализа фазовой информации голосового сигнала.

Цель данной работы – исследование фазовых характеристик голосового сигнала и выявление их основных особенностей.

Методика и результаты исследований голосового сигнала

Большинство источников, в том числе и голосовые связки, производят не простые, а сложные (комплексные) колебания, то есть колебания, характеризующиеся наличием более чем одной частоты. Поэтому голосовые сигналы представляют собой комплексные колебания, т.е. сложнейшие сочетания множества простых или чистых тонов и/или шумов. Как известно, голосовой сигнал всегда включает колебание с частотой основного тона, который характеризует частоту повторения полных колебательных циклов в единицу времени.

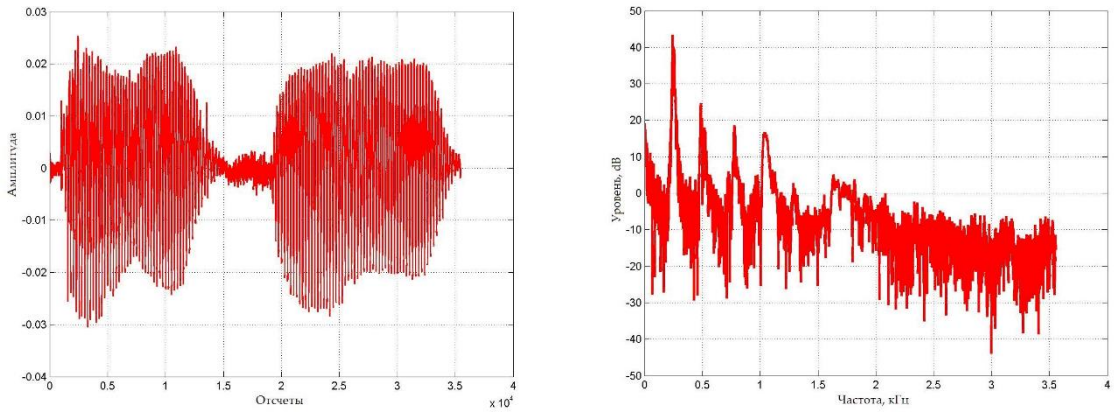
Частота основного тона зависит от напряженности голосовых связок, которые регулируются сокращением соответствующих мышц, или подъемом перстневидного хряща, а также перепада давлений по обе стороны глоттиса. Заметим, что голосовые связки, перстневидный хрящ и глоттис являются элементами гортани и являются

уникальными для каждого пользователя. Поэтому частота основного тона является определяющей в процессе голосовой аутентификации пользователя [1, 6].

Частота основного тона для всех голосов лежит в пределах 70 – 450 Гц. При произнесении речи она непрерывно изменяется в соответствии с ударением, подчеркиванием звуков и слов, а также с проявлением эмоций (вопрос, восклицание, удивление и т. д.). Изменение частоты основного тона называется интонацией. У каждого человека свой диапазон изменения основного тона (обычно он бывает немногим более октавы) и своя интонация [1, 6, 7].

Продемонстрируем роль частоты основного тона на примере цифры один, которую произносила женщина. Частота дискретизации составляла 64 кГц. Поэтому для более качественного анализа амплитудного спектра будем его ограничивать и далее обозначать, как «короткий» спектр. Указанный голосовой сигнал во временной области представлен на рис. 1 а, а его «короткий» амплитудный спектр на рис. 1 б.

Анализ представленного спектра позволяет выделить частоту основного тона на частоте примерно 250 Гц. Кроме этого, на этом рисунке можно выделить и три кратные частоты (максимумы на частотах 500, 1000 и 1500 Гц), которые обозначаются как обертоны (форманты). Амплитуда указанных максимумов спадает со скоростью от 6 до 12 дБ на октаву, что также является признаком пользователя. Линия, соединяющая вершины указанных гармоник амплитудно-частотного спектра, называется спектральной огибающей, коэффициенты которой используются для формирования шаблона пользователя [6].



а) б)

Рис. 1. Голосовой сигнал цифры «1» (а) и его «короткий» амплитудный спектр (б)

Таким образом, частота основного тона, интонация, устный почерк и тембр голоса служат для аутентификации пользователя, и степень достоверности такая же высокая, как по отпечаткам пальцев [8]. Здесь же следует отметить, что все это касается действительной области регистрируемых сигналов, мнимая часть, как и фазовая информация, игнорируются.

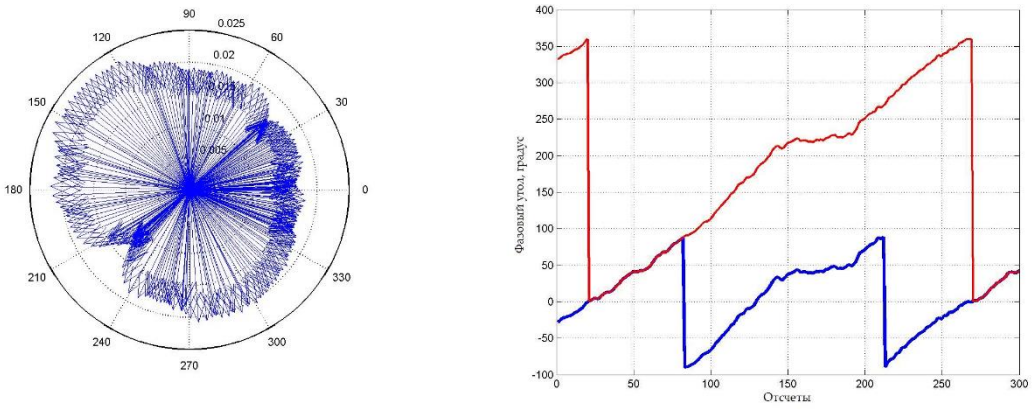
Рассматриваемый голосовой сигнал можно представить и в ином виде, где будут учитываться все его информативные параметры (амплитуда, частота и фаза). Для

этого необходимо выполнить преобразование Гильберта к зарегистрированной цифровой последовательности голосового сигнала, а затем рассчитать массив фазовой информации [9].

После цифровой обработки рассматриваемый сигнал можем представить на комплексной плоскости в виде ряда радиус-векторов в их обычном виде, то есть в виде стрелок, исходящих из начала координат. Каждый вектор будет иметь длину, которая характеризует амплитуду колебания, и угол, определяющий его текущую фазу. Эти параметры определяются действительной и мнимой частью аналитического (комплексного) сигнала. Более того, радиус-вектор для следующего отсчета будет иметь другую фазу и амплитуду, т.е. последовательность радиус-векторов будет осуществлять вращательное движение вокруг начала координат. При этом скорость вращения будет определять частоту анализируемого сигнала.

Фрагмент анализируемого сигнала на комплексной плоскости представлен на рис. 2 а, а его фазовая информация на рис. 2 б. Для рис. 2 б необходимо сделать следующее пояснение. Как известно, функция арктангенс на основе мнимой и действительной составляющей аналитического сигнала выдает значение фазового угла в пределах от -90° до 90° (синяя зависимость на рис. 2 б). Поэтому фазовую информацию необходимо преобразовать к диапазону углов от 0° до 360° (красная зависимость на рис. 2 б), что будет соответствовать фазовым углам, изображенным на рис. 2 а.

Здесь же обратим внимание и на недостатки преобразования Гильберта, на которые указано в [2], а именно «многие практически важные сигналы не являются минимально-фазовыми, вследствие чего метод на основе преобразования Гильберта имеет ограниченное применение».



а) б)

Рис. 2. Радиус-векторы (а) и фазовая информация (б) фрагмента анализируемого сигнала

Кроме этого, известно, что голосовой сигнал является нестационарным, характеристики которого часто меняются во времени. Это также приводит к некачественным результатам, получаемым с помощью преобразования Гильберта при малом

отношении сигнал/шум (вначале и конце голосового сигнала) или когда действительная составляющая анализируемого сигнала приближается к нулю.

Проиллюстрируем сказанное с помощью экспериментальных данных. На рис. 3 представлен фрагмент анализируемого сигнала. В верхней части этого рисунка представлены зависимости (вещественная и мнимая составляющая) голосового сигнала и шума. В нижней части представлена зависимость фазового угла. Анализ представленных зависимостей подтверждает низкое качество определения фазового угла в начале сигнала. Кроме того, имеют место ошибки в определении фазового угла в диапазоне от 2250 до 2500 отсчетов. В тоже время следует отметить периодичность изменения фазового угла в диапазоне от 0^o до 360^o. Заметим, что при регистрации цифры происходят небольшие колебания вещественной составляющей. Последнее приводит к соответствующим изменениям в фазовом угле (см. рис. 3).

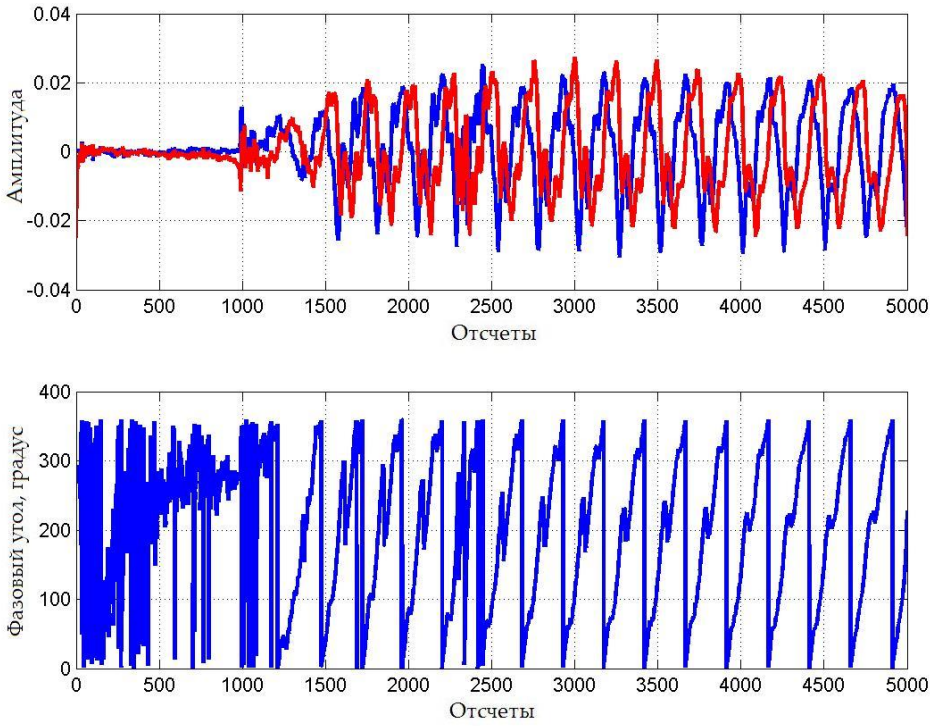


Рис. 3. Фрагмент регистрации и обработки голосового сигнала цифры один (низкое качество)

На рис. 4 представлен фрагмент той же цифры, которая была зарегистрирована у того же пользователя системы аутентификации. Качественная регистрация голосового сигнала приводит к отсутствию колебаний фазового угла. Следует отметить, что фазовые данные имеют форму пилообразного сигнала неизвестной длительности. Этот факт (априорную информации о форме ожидаемого сигнала) целесообразно использовать в процедурах выделения и анализа фазовых данных.

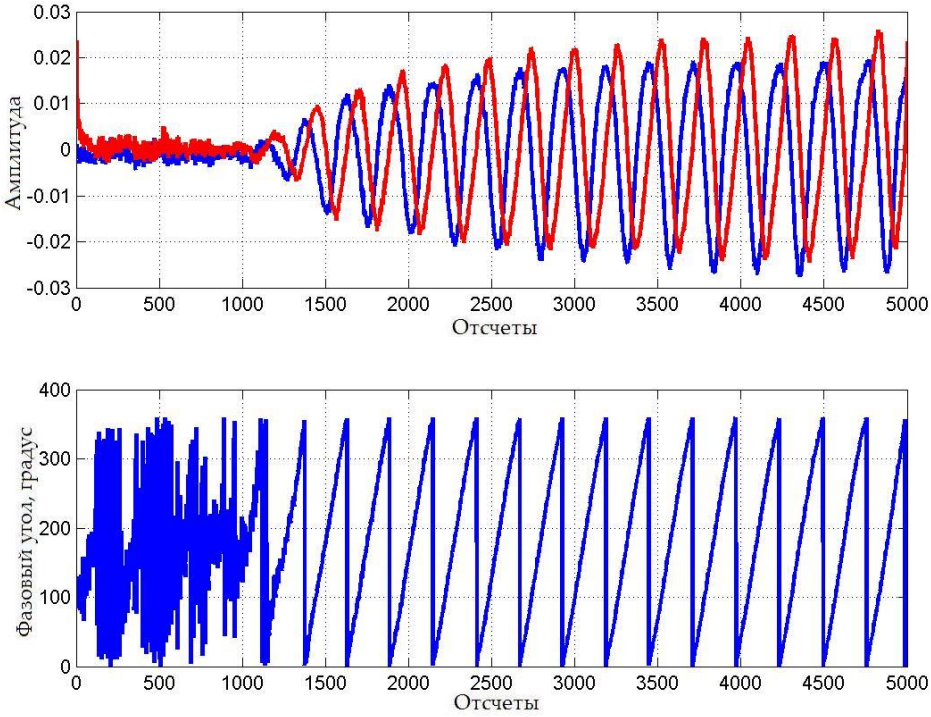
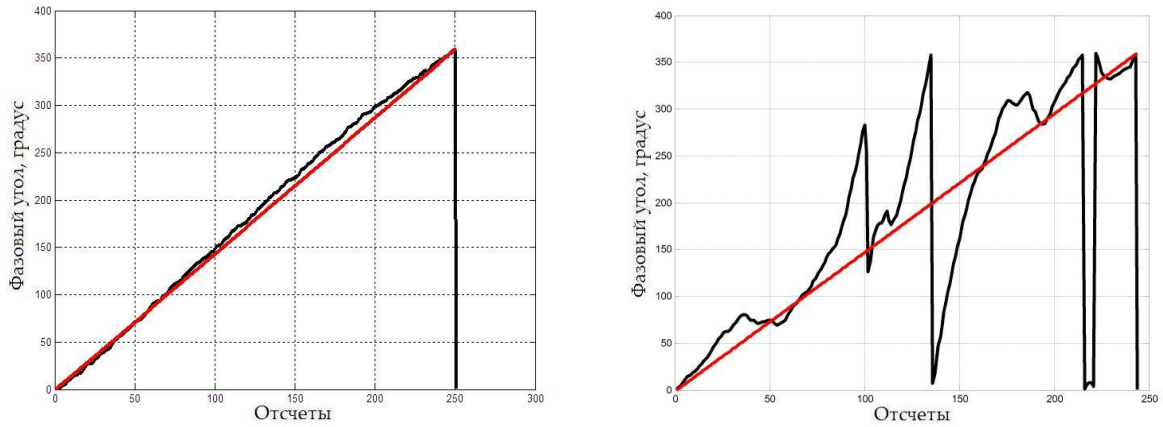


Рис. 4. Фрагмент регистрации и обработки голосового сигнала цифры один (высокое качество)

На рис. 5 представлены два периода изменения фазового угла, которые получены в результате расчета с использованием зарегистрированного голосового сигнала. Эти зависимости показаны черным цветом. Красным цветом на этих рисунках представлены зависимости ожидаемых значений фазового угла в соответствии с выдвинутой гипотезой о пилообразном изменении сигнала фазы.



а) б)

Рис. 5. Ожидаемая и расчетная зависимости фазового угла при отсутствии (а) и наличии (б) ошибок

Анализ представленных зависимостей подтверждает достоверность выдвинутой гипотезы об изменении фазового угла в виде пилообразного сигнала. Длительность периода изменения фазового угла зависит от характеристик голосового тракта пользователя и содержания сигнала, что может использоваться в системах аутентификации. Однако, в процессе анализа голосового сигнала имеют место отклонения от ожидаемой формы зависимости фазового угла (см. рис. 5б). При этом, как и ранее, красным цветом изображена ожидаемая зависимость фазового угла. Выявленные ошибки в определении фазы сигнала целесообразно откорректировать с учетом априорной информации, а затем на основе этих данных уточнить действительную и мнимую составляющие голосового сигнала.

Следует заметить, что математическая обработка фазового угла имеет ряд нерешенных задач, таких как, определение начала и конца зависимости фазового угла, устранение одиночных и групповых ошибок в определении значений фазового угла (см. рис. 5 б) и др. Определение начала и конца пилообразного сигнала фазового угла дает возможность получить его числовые оценки (математическое ожидание, дисперсию и др.), что может упростить процедуры аутентификации.

Для решения задачи определения начала и конца изменения фазового угла, который имеет форму пилообразного сигнала с медленным ростом, можно использовать известную процедуру – обнаружение детерминированного сигнала в шуме неизвестной интенсивности [10].

В рамках известной процедуры анализа это означает, что наблюдается выборка $X_n = \{x_0, ..., x_n\}$, которая может состоять либо из шумовых компонент, либо из компонент, получающихся при сложении сигнала с шумом. При этом n – объем анализируемой выборки, а модель исследуемого процесса имеет вид

$$X_n = \lambda S_n + \Xi_n$$

где $\lambda = 1$ с вероятностью p_1; $\lambda = 0$ с вероятностью p_2 $(p_1 + p_2 = 1)$; $s_i = s(t_i)$; $\xi_i = \xi(t_i)$. Элементы пилообразного сигнала фазового угла определяются следующим образом

$$s_i = \begin{cases} 360 i / k, & если \quad 0 \le i \le k; \\ \\ 0, & если \quad k < i < n. \end{cases}$$

Гауссов шум Ξ_n является некоррелированным с неизвестной дисперсией σ^2.

В процессе обработки выборки X_n необходимо принять решение относительно параметра λ. В соответствие, с критерием максимального правдоподобия данное решение принимается в процессе сравнения отношения правдоподобия, имеющего вид

$$\Lambda(X_n \Big| \hat{\sigma}_1^2, \hat{\sigma}_2^2) = \left(\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right)^{n/2} = \left[\frac{\sum_{i=0}^{n} x_i^2}{\sum_{i=0}^{n} (x_i - s_i)^2} \right]^{n/2}$$

с порогом С, зависящим от платы за правильные и ошибочные решения и априорных значений вероятностей p_1 и p_2. При этом

$$\lambda = \begin{cases} 0, & если \quad \Lambda < C; \\ 1, & если \quad \Lambda \geq C. \end{cases}$$

Здесь $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ – оценки дисперсии для гипотез, что X_n содержит мешающий шум Ξ_n или смесь шума Ξ_n и ожидаемого сигнала S_n соответственно.

Проблема применения рассмотренных процедур заключается в том, что не всегда известна точно величина k – длительность ожидаемого сигнала. В связи с этим, было проведено имитационное моделирование для двух случаев, когда величина k была известна точно или задана приблизительно.

В случае, когда величина k (длительность ожидаемого сигнала) была известна точно, рассмотренные процедуры позволяют с высокой достоверностью и точностью определять границы ожидаемого пилообразного сигнала в широком диапазоне отношений сигнал/шум.

Когда величина k была задана с ошибкой в несколько отсчетов (до ± 10) рассматриваемые процедуры работают устойчиво, но при этом необходимо изменить величину порога (в сторону увеличения). При неточном задании порога количество регистрируемых сигналов может кратно увеличивается (в два, три раза), а регистрируемое начало сигнала отличается (на один, два отсчета соответственно).

Здесь же заметим, что точную длительность фазового сигнала можно уточнить с помощью итерационных процедур.

Рассмотренные процедуры выделения фазового угла можно использовать и для решения иных задач актуальных для обработки речевых сигналов. Например, обнаружения начала и конца фрейма, который содержит речевой сигнал; разделения анализируемого слова на слоги и т.д.

Выводы

Решалась научная задача разработки процедур выделения и анализа фазовой информации голосового сигнала пользователя. Рассмотрено текущее состояние современных систем голосовой аутентификации, которые базируются на анализе амплитудной и частотной информации в спектральной области. К сожалению, традиционно фазовые данные регистрируемых сигналов игнорируются, хотя, как известно, они несут больше информации. Дополнительным достоинством фазовой информации является и то, что форма сигнала априори известна, а именно, пилообразный сигнал неизвестной длительности.

Использование фазовой информации в системах аутентификации требует решения множества задач от расчета фазовых углов до выявления признаков пользователя. Впервые разработаны и исследованы процедуры выделения фазового сигнала на фоне шума неизвестной интенсивности, проведено имитационное моделирование и показана работоспособность указанных процедур при наличии априорной информации о длительности пилообразного сигнала. Точность задания длительности фазового сигнала составляет ± 10 отсчетов. Полученные процедуры дают возможность уточнить начало и конец фазового сигнала, при необходимости компенсировать аномальные измерения, а также выделить фрейм голосового сигнала из шумовой последовательности.

Дальнейшие исследования будут ориентированы на поиск отличительных признаков пользователя по фазовой информации его голосового сигнала.

Список литературы:

1. *Beigi H.* Fundamentals of Speaker Recognition. – NY: Springer, 2011. – 1029 с.

2. *Оппенхейм А.В., Лим Дж.С.* Важность фазы при обработке сигналов // ТИИЭР. – Т. 69 (1981). – № 5. – С. 39–54.

3. *Борисенко С.Ю., Воробьев В.И., Давыдов А.Г.* Сравнение некоторых способов анализа фазовых соотношений между квазигармоническими составляющими речевых сигналов // Сборник трудов 1-ой Всероссийской акустической конференции. – 2004. – С. 2-7.

4. *Воробьев В.И., Давыдов Г.В., Шамгин Ю.В.* Фазовые соотношения между основным тоном и обертонами гласных звуков // Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2006. – № 2(14).– С. 64-68.

5. *Воробьев В.И.* Межкомпонентная фазовая обработка речевых сигналов во временной и частотной областях // Акустика речи. Медицинская и биологическая акустика. Архитектурная и строительная акустика. Шумы и вибрации. Аэроакустика / Сборник трудов XIX сессии Российского акустического общества. Т. 3. – М.: ГЕОС, 2007. – С. 46-49.

6. *Сорокин В.Н., Вьюгин В.В., Тананыкин А.А.* Распознавание личности по голосу: аналитический обзор // Информационные процессы. – 2012. – Т. 12, № 1. – С. 1–30.

7. *Фант Г.* Акустическая теория речеобразования / Г.Фант. Пер.с англ. – М.: Наука, Главная редакция физико-математической литературы, 1964. – 284 с.

8. *Гудонавичус Р.В.* Распознавание речевых сигналов по их структурным свойствам / Р.В. Гудонавичус, П.П. Кемешис, А.Б. Читавичус. – Л.: Энергия, 1977. – 300 с.

9. *Вайнштейн Л.А.* Разделение частот в теории колебаний волн / Л.А. Вайнштейн, Д.Е. Вакман. – М.: Наука, Главная редакция физико-математической литературы, 1983. — 288 с.

10. *Репин В.Г.* Статистический синтез при априорной неопределенности и адаптация информационных систем / В.Г. Репин, Г.П. Тартаковский. – М.: Советское радио, 1977. – 432 с.