

УДК 057.087.1:621.391.26

ИССЛЕДОВАНИЕ ИНФОРМАТИВНОСТИ ФАЗОВЫХ ДАННЫХ ГОЛОСОВОГО СИГНАЛА ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ АУТЕНТИФИКАЦИИ



[Н.С. ПАСТУШЕНКО](#), [В.Г. ПЕДРО](#), [О.Н. ФАЙЗУЛАЕВА](#)

Харьковский национальный
университет радиоэлектроники

Abstract – Ways to improve the efficiency of modern voice authentication systems in various access systems are investigated in the article. One of the main ways to increase the effectiveness of authentication systems under consideration is associated with the use of voice signal phase. The object of the research is the process of digital signal processing in voice authentication systems. The scientific problem of forming and using phase data of a voice signal of an authentication system is solved. The purpose of the research is to evaluate the informative value of phase data of a voice signal and determine their main characteristics. The formant information on the amplitude spectrum of the experimental voice signal was processed and its qualitative and quantitative characteristics were obtained. Then, the phase data of the analyzed signal were calculated, the features of their use were revealed, and the phase spectrum was obtained on the basis of the calculation results. Processing of the phase spectrum has shown that it is easier to select at least one and a half times bigger formant of a voice signal that can be used to authenticate a user. The approximation of the maxima of formants is performed using linear and quadratic polynomials. The qualitative and quantitative evaluation of the amplitude and phase spectra formant information has confirmed the hypothesis that the phase data are more informative. The presented research results should be used in voice authentication systems, in improving speech recognition systems, as well as in solving speaker identification problems.

Анотація – У роботі досліджуються напрямки підвищення ефективності сучасних систем голосової аутентифікації в різних системах доступу. Один з основних напрямків підвищення ефективності розглянутих систем аутентифікації пов'язаний з використанням фази голосового сигналу. Об'єктом дослідження є процес цифрової обробки сигналів у системах голосової аутентифікації. Вирішується наукове завдання формування й використання фазових даних голосового сигналу системи аутентифікації. Метою досліджень є оцінка інформативності фазових даних голосового сигналу й визначення їхніх основних характеристик. Якiсна й кількісна оцінка формантної інформації амплітудного й фазового спектрів підтвердила гіпотезу про більшу інформативність фазових даних. Представлені результати досліджень доцільно використовувати в системах голосової аутентифікації, удосконалювання систем розпізнавання мови, а також при рішенні завдань ідентифікації диктора.

Аннотация – В работе исследуются направления повышения эффективности современных систем голосовой аутентификации в различных системах доступа. Одно из основных направлений повышения эффективности рассматриваемых систем аутентификации связано с использованием фазы голосового сигнала. Объектом исследования является процесс цифровой обработки сигналов в системах голосовой аутентификации. Решается научная задача формирования и использования фазовых данных голосового сигнала системы аутентификации. Целью исследований является оценка информативности фазовых данных сигнала и определение их основных характеристик. Качественная и количественная оценка формантной информации амплитудного и фазового спектров подтвердила гипотезу о большей информативности фазовых данных. Представленные результаты исследований целесообразно использовать в системах голосовой аутентификации, совершенствования систем распознавания речи, а также при решении задач идентификации диктора.

Введение

Число киберпреступлений в области инфокоммуникаций значительно растет, и главное – они становятся все более изощренными. Так, согласно годовому отчету по кибербезопасности Cisco 2018 [1], значительно возросло распространение вредоносного программного обеспечения (в частности, программ-вымогателей – ransomware), объемы передаваемого зашифрованного веб-трафика злоумышленников, увеличилось количество угроз со стороны электронной почты. Кроме того, необходимо отметить преступное использование облачных сервисов и других легитимных сете-

вых ресурсов, а также использование уязвимостей технологии Интернет вещей (Internet of Things, IoT) для осуществления атак DDoS.

Для защиты финансовых ресурсов и конфиденциальной информации широко применяются пин-коды, пароли, идентификационные карточки, с помощью которых производится аутентификация пользователя. Однако, эти средства защиты не отличаются совершенством, поскольку их можно потерять или подделать. Поэтому в настоящее время широко используется биометрическая аутентификация пользователя, которая является решением вышеперечисленных проблем.

Первоначально основное внимание уделялось статическим биометрическим признакам (отпечатки пальцев, форма и размеры лица, узор радужной оболочки и сетчатки глаз), которые ранее плодотворно использовались в криминалистике. Из-за простой подделки и ограниченного объема анализируемых данных в последнее время в системах доступа предпочтение отдается динамическим (поведенческим) биометрическим признакам, и в первую очередь, голосовому сигналу.

Принцип работы систем голосовой аутентификации следующий. При регистрации в системе пользователь демонстрирует свои биометрические признаки (в данном случае голос). Эти признаки обрабатываются специальным алгоритмом и формируется шаблон пользователя, который хранится в соответствующей базе. Этот шаблон используется для проверки идентичности каждого, кто выдает себя за соответствующего пользователя. При прохождении процесса аутентификации пользователь произносит некий текст, который формируется системой аутентификации. В зависимости от совпадения или несовпадения совокупности предъявленных признаков с зарегистрированными в контрольном шаблоне предъявивший их признается законным (при совпадении) или незаконным (при несовпадении) пользователем.

Первое преимущество систем голосовой аутентификации перед другими биометрическими методами, основанными на анализе отпечатков пальцев или сетчатки глаза, состоит в возможности получения и проверки биометрических данных без применения специализированных и дорогостоящих технологий: достаточно иметь телефон или микрофон, подключенный к компьютеру. Таким образом, голосовая аутентификация не требует применения специализированного дорогостоящего оборудования. Все, что нужно – обычный микрофон. В отличие от других биометрических систем контроля доступа, голосовая аутентификация может производиться удаленно через телефонную сеть, Интернет, корпоративные сети.

И главное, в рассматриваемых системах аутентификации существует возможность оперативно изменять и наращивать контрольную фразу, а также использовать современные достижения цифровой обработки сигналов, которые широко и эффективно используются в радиолокации и радиосвязи. Указанное выше позволяет существенно улучшить качественные характеристики современных систем голосовой аутентификации, которые базируются на анализе амплитуды и частоты голосового сигнала, а фазу, как известно, традиционно игнорируют [2]. В то же время, давно известно, что фазовые данные сигнала являются более информативными [3]. Данная работа является продолжением исследований, которые представлены в [4], где авто-

ры обращали внимание на важность фазовой информации при обработке голосового сигнала.

Цель данной работы – исследование информативности фазовых данных голосового сигнала и их непротиворечивости амплитудно-частотной информации. Объект исследования – процесс цифровой обработки голосового сигнала пользователя системы аутентификации.

I. Методика и результаты исследований

В настоящее время при голосовой аутентификации рассчитывается амплитудный спектр, и по нему выявляются признаки пользователя в области низких частот (100 – 300 Гц), на частотах 4 – 5 и 6,5 – 7,8 кГц, а также в небольшой области – в районе 1 кГц [5-7].

При этом основное внимание уделяется анализу огибающей спектра и его среднему наклону, формантным частотам и их полосам, а также кепстру. Как известно, кепстр описывает форму огибающей спектра сигнала, в которой интегрируются характеристики источников возбуждения (голосового, турбулентного и импульсного) и формы речевого тракта. В экспериментах, в том числе, и по субъективному распознаванию голоса, было установлено, что огибающая спектра сильно влияет на узнаваемость голоса [5].

Поэтому в дальнейшем рассчитаем амплитудно-частотный спектр по экспериментальному голосовому сигналу пользователя системы аутентификации и в области низких частот выделим формантные частоты, а также построим их огибающую. Следующим шагом будет восстановление квадратурной составляющей голосового сигнала и расчет его фазовых данных. Затем необходимо откорректировать фазовые данные и построить фазовый спектр. Последний шаг – выполним указанные выше операции с фазо-частотным спектром анализируемого сигнала.

Анализу будем подвергать экспериментальный голосовой сигнал пользователя системы аутентификации, который произносил цифру «один». Частота дискретизации – 64 кГц и отношение сигнал/шум – более 20 дБ.

Анализируемый сигнал и его амплитудно-частотный спектр представлены на рис. 1. При этом спектр сигнала ограничим сверху границей канала тональной частоты, чтобы более качественно выполнить анализ последовательности основных формант.

Результаты обработки формантной информации этого спектра представлены в табл. 1. Ярко выраженными в спектре являются четыре форманты, а пятая имеет очень низкий уровень спектральной мощности (несколько децибел). Частота основного тона составляет 243 Гц. Вторая форманта имеет частоту в два раза выше. К сожалению, последующие форманты не кратны частоте основного тона [8].

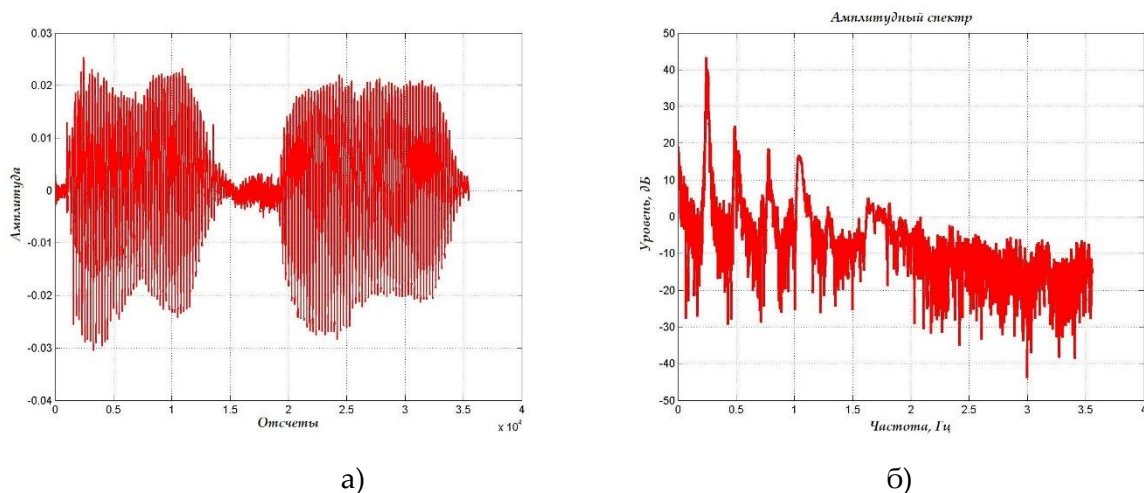


Рис. 1. Голосовой сигнал цифры «1» (а) и его короткий спектр (б)

Таблица 1. Характеристики формант амплитудного спектра

| | | | | |
|-------------|------|------|------|------|
| Уровень, дБ | 43,4 | 24,6 | 18,6 | 14,2 |
| Частота, Гц | 243 | 486 | 776 | 1025 |

Проведем расчет и анализ фазового спектра рассматриваемого сигнала. При этом надо сделать ряд замечаний. Фаза голосового сигнала не регистрируется. Поэтому ее, как правило, рассчитывают программно-алгоритмически. Для этого можем воспользоваться преобразованием Гильберта [9]

$$y(t) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau,$$

где $x(t)$ – зарегистрированный голосовой сигнал; $y(t)$ – квадратурная (мнимая) составляющая аналитического сигнала; t – независимая переменная, которая имеет физический смысл единицы времени; τ – переменная интегрирования.

Далее можем рассчитать фазу голосового сигнала с помощью следующего соотношения

$$\varphi(t) = \arctg \frac{y(t)}{x(t)}.$$

К сожалению, функция \arctg выдает значения угла в диапазоне от $-\pi/2$ до $\pi/2$. Поэтому для определения правильного значения фазового угла, который у голосового сигнала изменяется в пределах от 0 до 2π , необходимо угол $\varphi(t)$ соответствующим образом откорректировать. В противном случае фазовый спектр будет некорректным.

Теперь проанализируем фазовый спектр анализируемого сигнала. На рис. 2 представлены два спектра. Разница представленных фазовых спектров заключается в следующем. На рис. 2 а) представлен спектр, который рассчитан на основе угловых

данных, полученных по значениям на выходе функции \arctg . На данном рисунке нарушен порядок отображения формант, характер которых должен соответствовать изображению формант на рис. 1 б). На рис. 2 б) представлен фазовый спектр откорректированного фазового угла, который будем рассматривать ниже.

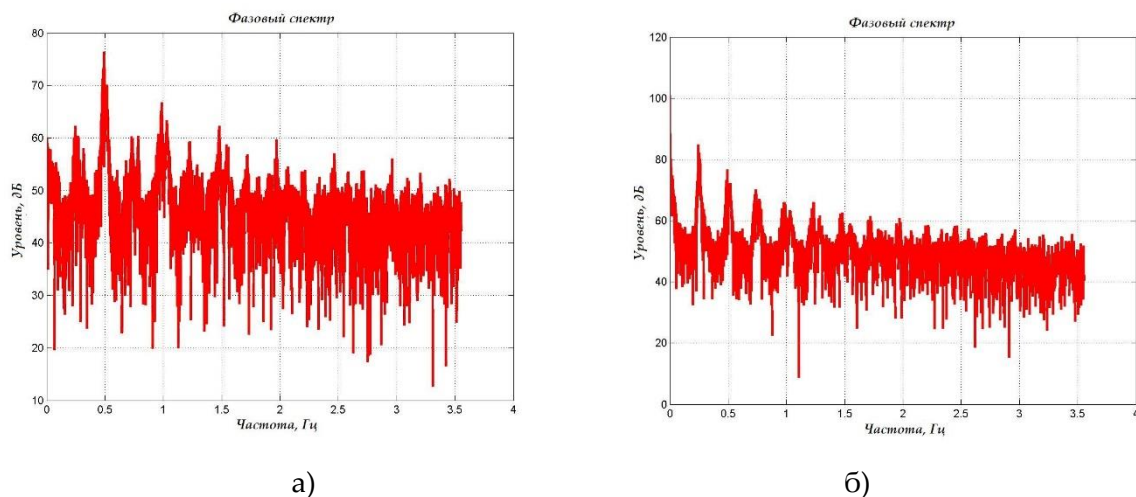


Рис. 2. Фазовый спектр голосового сигнала: а) на выходе функции \arctg ;
б) откорректированный угол

Результаты обработки формантной информации фазового спектра представлены в табл. 2. В данном спектре можно выделить шесть формант, а седьмая и восьмая имеют незначительное энергетическое отличие. Частота основного тона, как и в амплитудном спектре, составляет 243 Гц.

Уровень спектральной плотности выделенных максимумов в разы превышает уровень максимумов амплитудного спектра, что существенно упрощает процедуру их выделения. Количество выделенных формант у фазового спектра в полтора раза больше. Отмеченное выше свидетельствует о большей информативности фазового спектра голосового сигнала.

Таблица 2. Характеристики формант фазового спектра

| | | | | | | |
|-------------|------|------|------|-----|------|------|
| Уровень, дБ | 84,9 | 76,7 | 70,3 | 65 | 64 | 62 |
| Частота, Гц | 243 | 492 | 738 | 990 | 1217 | 1450 |

Теперь выполним аппроксимацию выделенных максимумов анализируемых спектров с помощью полиномов различной степени (построим огибающую спектра). В качестве критерия для выбора коэффициентов полинома будем использовать метод наименьших квадратов. Графики аппроксимации максимумов формант с помощью полиномов первой и второй степени представлены на рис. 3.

При этом на рисунках показано следующее: зелеными точками обозначены координаты максимумов формантных частот; сплошной красной линией – полином первой степени; штриховой синей кривой – полиномом второй степени. Сумма

квадратов невязок для амплитудного спектра составляет: 4,6 – линейный полином; 1,9 – квадратичный полином.

Для фазового спектра указанные величины имеют значения 0,9 и 0,06 соответственно. Следует отметить, что квадратичный полином с высоким качеством аппроксимирует максимумы формант фазового спектра.

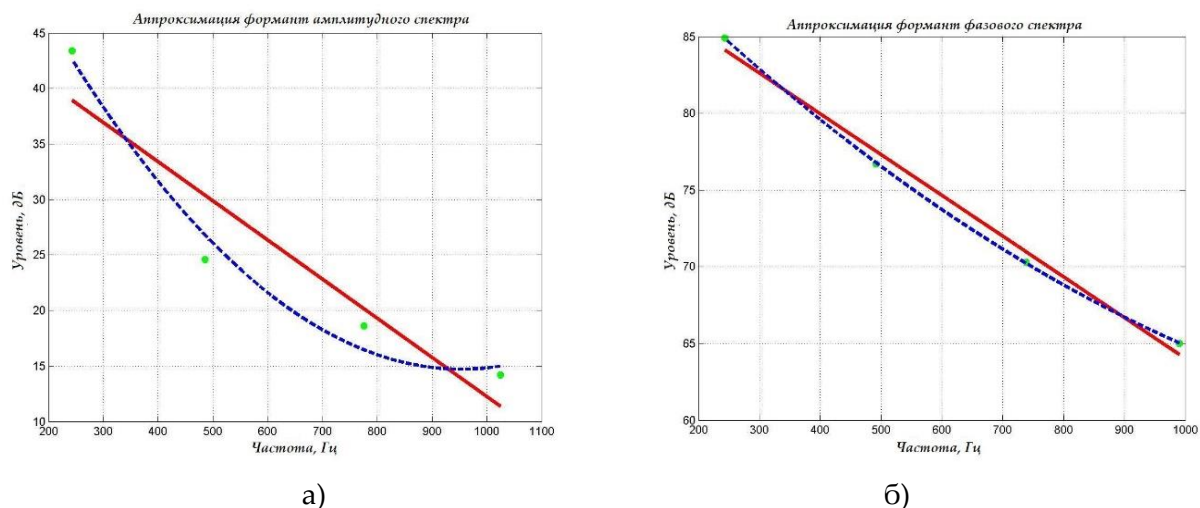


Рис. 3. Аппроксимация максимумов формант: а) амплитудного спектра; б) фазового спектра

Теперь оценим, как согласуются данные фазового и амплитудного спектров с помощью критерия хи-квадрат (χ^2). При этом данные амплитудного спектра будем считать теоретическими. С этой целью для указанных полиномов рассчитаем значения уровня спектральной мощности в точках от 200 Гц до 1000 Гц с шагом в 100 Гц. Затем для рассчитанных значений выполним нормировку с помощью соотношения

$$\beta(z) = \frac{z - z_{\min}}{z_{\max} - z_{\min}}.$$

В результате мы получим анализируемые данные, которые приведены к интервалу от 0 до 1, что позволит воспользоваться критерием χ^2 . Результаты расчетов представлены в табл. 3.

Анализ данных, представленных в табл. 3, свидетельствует о равенстве данных для линейных полиномов амплитудного и фазового спектров, т.е. огибающие спектров имеют одинаковые коэффициенты. Последнее свидетельствует о равенстве характеристик рассматриваемых спектров.

Для квадратичного полинома воспользуемся критерием хи-квадрат. Результаты расчетов показывают, что экспериментальное значение функции χ^2 равно 1,71. Сравнение указанной величины с табличной при уровне значимости равным 0,01, говорит о том, что полученные результаты не дают повода сомневаться в соответствии квадратичной аппроксимации фазового и амплитудного спектров в области низких частот.

Таблица 3. Результаты аппроксимации формант

| Частота, Гц | Амплитудный спектр, дБ | | Фазовый спектр, дБ | |
|-------------|------------------------|----------------------|--------------------|----------------------|
| | Линейный полином | Квадратичный полином | Линейный полином | Квадратичный полином |
| 200 | 1 | 1 | 1 | 1 |
| 300 | 0,88 | 0,75 | 0,88 | 0,84 |
| 400 | 0,75 | 0,54 | 0,75 | 0,68 |
| 500 | 0,63 | 0,36 | 0,63 | 0,54 |
| 600 | 0,5 | 0,22 | 0,5 | 0,41 |
| 700 | 0,38 | 0,11 | 0,38 | 0,29 |
| 800 | 0,25 | 0,04 | 0,25 | 0,18 |
| 900 | 0,125 | 0,001 | 0,125 | 0,08 |
| 1000 | 0 | 0 | 0 | 0 |

Выводы

В проведенных исследованиях оценивалась информативность фазовых данных голосового сигнала пользователя системы аутентификации. Установлено, что при использовании фазовых данных обязательной процедурой должна быть коррекция угловых данных на выходе функции \arctg к диапазону от 0 до 2π . В противном случае получим искаженные результаты.

Уровень спектральной мощности формантных максимумов фазового спектра в несколько раз превышает максимумы амплитудного спектра, что существенно упрощает процедуры их выделения. Количество выделяемых формант в фазовом спектре в полтора раза больше по сравнению с амплитудным спектром.

Выполнена аппроксимация максимумов формант с помощью линейных и квадратичных полиномов. Аппроксимация линейными полиномами дает идентичные результаты, т.е. наклон кривых одинаков. Аппроксимация для квадратичных полиномов дает результаты, которые согласуются по критерию χ^2 . При этом аппроксимация квадратичным полиномом очень точно совпадает с экспериментальными данными фазового спектра голосового сигнала.

Таким образом, фазовый спектр в области низких частот (до 1100 Гц) имеет большую информативность по отношению к амплитудному. Поэтому использование фазовой информации голосового сигнала пользователя может существенно улучшить качественные характеристики систем аутентификации.

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

Список литературы:

1. Отчет Cisco по информационной безопасности за 2018 год. – 68 с – Режим доступа: https://www.cisco.com/c/dam/global/ru_ru/assets/offers/assets/cisco_2018_acr_ru.pdf.
2. *Beigi H.* Fundamentals of Speaker Recognition. – NY: Springer, 2011. – 1029 с.

3. *Оттенхейм А.В., Лим Дж.С.* Важность фазы при обработке сигналов // ТИИЭР. – 1981. – Т. 69, № 5. – С. 39–54.

4. *Павленко И.С.* Исследование фазовых характеристик голосового сигнала пользователя системы аутентификации [Электронный ресурс] / И.С. Павленко, Н.С. Пастушенко, О.Н. Файзулаева // Проблемы телекоммуникаций. – 2017. – № 2 (21). – С. 52 - 60. – Режим доступа к журн.: http://pt.journal.kh.ua/2017/2/1/172_pavlenko_signal.pdf.

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

6. *Besacier L., Bonastre J.-F.* Subband architecture for automatic speaker recognition // Signal Process. – 2000. – V. 80. – P. 1245–1259.

7. *Lu X., Dang J.* An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification // Speech Communication. – 2007. – V. 50, N 4. – P. 312–322.

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

9. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных: Пер. с англ. – М.: Мир, 1989. – 540 с.