УДК 057.087.1:621.391.26

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ АМПЛИТУДНОГО И ФАЗОВОГО СПЕКТРОВ РЕЧЕВОГО СИГНАЛА ПОЛЬЗОВАТЕЛЯ СИСТЕМ ГОЛОСОВОЙ АУТЕНТИФИКАЦИИ

О.Н. ФАЙЗУЛАЕВА, Н.С. ПАСТУШЕНКО
Харьковский национальный университет радиоэлектроники

Abstract *– The scientific challenge of improving the quality of voice authentication of computer systems and networks using the phase component of the registered user speech signal is considered. The object of research is the process of digital signal processing in the voice authentication systems. The methods and procedures for digital processing of speech signals are studied in relation to the voice authentication systems. The voice signal of the user, which comprises a sequence of the same digits entered into the laptop via an amplifier and microphone, has been subject to processing. The procedures of digital processing included the Hilbert transform, the calculation of the speech signal phase and the construction of the amplitude and phase spectra. The cross-correlation coefficient has been used as the quantitative characteristics to assess the amplitude and phase spectra informativeness of the user voice signal. As the results of experimental studies have shown, the most informative area for both the amplitude and phase spectra is low frequencies (600 Hz). However, the amplitude spectrum informativeness is almost twice higher than the informativeness of the phase spectrum. The latter can be conditioned by the fact that for the calculation of the quadrature component of the speech signal Hilbert transform has been used, which does not always give satisfying results in processing of polyharmonic and nonstationary data. The obtained results can be useful for specialists, performing research in speech and speaker recognition.*

Анотація *– Розглядається наукове завдання підвищення якості голосової аутентифікації комп'ютерних систем і мереж за рахунок використання фазової складової мовного сигналу користувача, що реєструється. Об'єктом дослідження є процес цифрової обробки сигналів у системах голосової аутентифікації. Досліджуються методи й процедури цифрової обробки мовних сигналів, що застосовуються до систем голосової аутентифікації. Обробці піддавався мовний сигнал користувача. Процедури обробки включали перетворення Гільберта, розрахунок фази мовного сигналу та побудова амплітудного й фазового спектрів. Як свідчать результати експериментальних досліджень, найбільш інформативною є область низьких частот (до 600 Гц), як для амплітудного, так і для фазового спектрів. Разом з тим, інформативність амплітудного спектру майже у два рази перевищує інформативність фазового спектру.*

Аннотация *– Рассматривается научная задача повышения качества голосовой аутентификации компьютерных систем и сетей за счет использования фазовой составляющей регистрируемого речевого сигнала пользователя. Объектом исследования является процесс цифровой обработки сигналов в системах голосовой аутентификации. Исследуются методы и процедуры цифровой обработки речевых сигналов применительно к системам голосовой аутентификации. Обработке подвергался речевой сигнал пользователя. Процедуры обработки включали преобразование Гильберта, расчет фазы речевого сигнала и построение амплитудного и фазового спектров. Как свидетельствуют результаты экспериментальных исследований, наиболее информативной является область низких частот (до 600 Гц), как для амплитудного, так и для фазового спектра. Вместе с тем, информативность амплитудного спектра почти в два раза превышает информативность фазового спектра.*

Введение

Работа коммерческих и некоммерческих организаций, финансовых институтов и предприятий связана с широким использованием различных ресурсов и услуг, доступ к которым осуществляется с помощью современных телекоммуникационных систем. В связи с тем, что доступ к информационным и финансовым ресурсам осуществляется по открытым каналам связи, особое внимание необходимо уделять методам и средствам защиты. Основная мировая тенденция – ориентация на постро-

ние телекоммуникационных сетей на базе архитектуры TCP/IP, в рамках которой создаются защищенные каналы передачи данных, реализуемые на базе методов и средств шифрования и аутентификации.

При этом методы аутентификации являются первым барьером, который предназначен для борьбы со злоумышленниками и определяет права и возможности авторизованного пользователя. В последнее время для повышения надежности аутентификации используются биометрические признаки (образы) пользователя, и в первую очередь, внешний вид лица, папиллярный узор пальцев и радужная оболочка глаз. Принятое решение странами G8 по использованию в качестве основных указанных выше признаков, которые относятся к статическим биометрическим образам пользователя, очевидно, оказалось ошибочным. Поскольку эти признаки позволяют качественно решать задачи идентификации пользователя, образ которого хранится в базе. В тоже время, сохранение в тайне статических признаков человека и исключение их подделки может быть реализовано только через обеспечение анонимности пользователя. Более того, указанные биометрические признаки при обработке имеют ограниченный неизменяемый объем исходных данных.

Поэтому в последнее время все больше внимания уделяется применению в системах доступа динамических биометрических признаков, таких как голос, клавиатурный почерк, личная подпись и др. Динамические биометрические признаки пользователя обладают неограниченным объемом анализируемых данных, например, за счет увеличения размеров исследуемого фрагмента речи. Более того, содержание анализируемого фрагмента и его размеры могут автоматически задаваться системой доступа, в зависимости от ее текущих качественных характеристик.

В связи с этим, в настоящее время все больше внимания уделяется исследованиям по использованию динамических биометрических признаков и в первую очередь, голосового сигнала пользователя. Системы голосовой аутентификации (СГА) обладают рядом дополнительных преимуществ, таких как: простота, удобство, экономичность, возможность удаленной аутентификации и др. СГА позволяют использовать все преимущества и достижения современной цифровой обработки сигналов. Здесь же следует заметить, что в современных голосовых системах аутентификации используются амплитудные характеристики, хотя давно известно, что фазовые характеристики являются более информативными [1]. В связи с бурным развитием цифровых сигнальных процессоров в последнее время проявился большой интерес к фазовым характеристикам и их более широкому использованию в цифровой обработке сигналов. К сожалению, в системах обработки речевых сигналов их фазовые характеристики традиционно игнорируются [2]. В статье [3] выполнен сравнительный анализ процедур оценки фазовых соотношений между колебаниями основного тона и обертонов речевых сигналов, которые авторы предлагают использовать для решения задач распознавания звуков речи и идентификации дикторов.

Цель настоящей статьи – сравнительная оценка информативности амплитудных и фазовых характеристик голосового сигнала пользователя применительно к СГА. Объектом исследования является процесс цифровой обработки сигналов в СГА.

В качестве количественной характеристики, вводимой для оценки информативности (количества сведений и знаний) амплитудных и фазовых характеристик голосового сигнала пользователя, будем использовать коэффициент взаимной корреляции. Данный коэффициент удовлетворяет всем требованиям, предъявляемых к количественному показателю: имеет название, математическое представление, физический смысл, размерность и возможные пределы изменений.

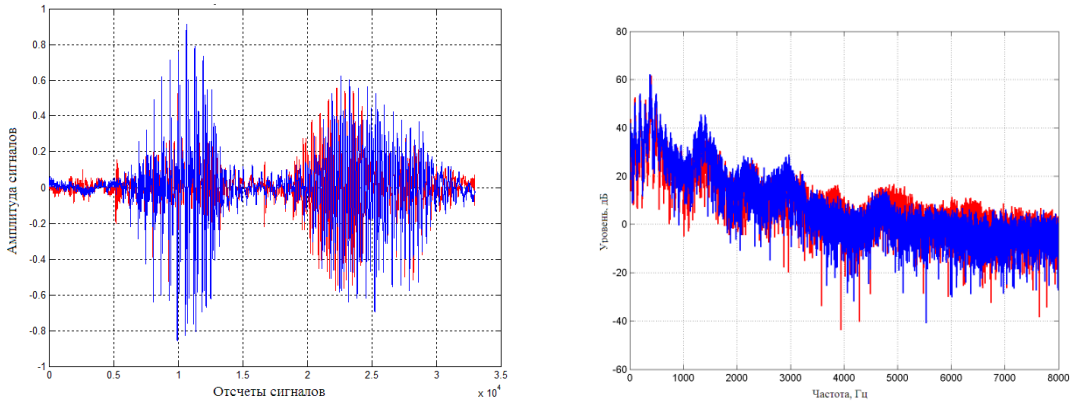
Известно, что в настоящее время в СГА используются преимущественно спектральные характеристики речевого сигнала пользователя. Например, в [4] использовалась огибающая амплитудного спектра (АС) голосового источника, в [5] предложен метод кепстрального преобразования АС речевых сигналов, в [6] применялась модель, в которой спектрально-временные характеристики речевого сигнала анализируются гребенкой фильтров, в [7, 8] исследовалось влияние двух микрофонов на качество процедур аутентификации.

При этом, как и в [9], будут подвергаться сравнительному анализу спектральные характеристики речевого сигнала, зарегистрированного с помощью одного канала (микрофона) [7]. При этом основное внимание будем уделять анализу диапазона АС до 8 кГц, что обусловлено наличием отличительных признаков пользователя в диапазоне от 100 Гц до 8 кГц [10, 11]. Для этого рассчитанный амплитудный (фазовый) спектр, диапазон изменения которого определяется половиной частоты временной дискретизации, будем ограничивать частотой 8 кГц («короткий» спектр). Наряду с АС, будем рассчитывать и фазовый спектр (ФС) речевых сигналов, а также оценим их информативность.

I. Методика и результаты исследований амплитудного и фазового спектров

Анализу подвергался речевой сигнал пользователя, который произносил цифры от 0 до 9. Ввод речевого сигнала осуществлялся с расстояния 0,7,…,1 м по нормали к оси микрофона в замкнутом помещении. Регистрация речевых сигналов осуществлялась с помощью ноутбука, к которому подключался микрофон с усилителем. В качестве помехового сигнала имел место акустический шум работы винчестера, а также внутренние шумы микрофона и усилителя. Частота дискретизации сигнала – 64 кГц. Отношение сигнал/шум экспериментальной последовательности составляло примерно 25 дБ. Естественно, предположить, что наиболее информативные участки «короткого» спектра в последовательности (одной и той же цифры) речевых сигналов должны совпадать и иметь, например, большой коэффициент взаимной корреляции. В этом будет заключаться исследуемая гипотеза.

Ниже представлены два речевых сигнала цифры «1» (рис. 1 а) и «короткий» спектр указанных сигналов (рис. 1 б). Здесь и далее изображение первого сигнала на графиках будет представлено красным цветом, а второго – синим.



а) б)

Рис. 1. Речевые сигналы цифры «1» (а) и их «короткий» амплитудный спектр (б)

В качестве оцениваемого параметра будем использовать коэффициент взаимной корреляции (КВК). Для расчета КВК использовалось известное соотношение для двух дискретных последовательностей [12]

$$k = \frac{\sum_{i=1}^{N}(K_i - m_e)\cdot(\hat{K}_i - m_r)}{\sqrt{\sum_{i=1}^{N}(K_i - m_e)^2 \cdot \sum_{i=1}^{N}(\hat{K}_i - m_r)^2}},$$

где K_i и $\hat{K}_i$ – анализируемые цифровые последовательности, $i = 1,...,N$ – номер отсчета анализируемой последовательности, N – количество анализируемых отсчетов, m_e, m_r – оценки математического ожидания анализируемых последовательностей. Ниже в качестве последовательностей рассматриваются либо речевые сигналы, либо их амплитудные (фазовые) спектры. Заметим, что КВК речевых сигналов равен 0,3. Низкий КВК анализируемых сигналов во временной области обусловлен тем, что не выполнены процедуры масштабирования и передискретизации. Указанные процедуры во временной области требуют значительных вычислительных затрат. КВК амплитудных спектров анализируемых сигналов равен 0,75, а «коротких» спектров – 0,83, т.е. в частотной области корреляция рассматриваемых сигналов выше.

Для анализа ФС исследуемых сигналов необходимо выполнить ряд дополнительных операций, таких как:

- расчет мнимой (квадратурной) составляющей аналитического сигнала;
- оценка фазы речевого сигнала в каждой точке регистрации;
- построение фазового спектра.

Для расчета мнимой составляющей использовалось преобразование Гильберта [13]. «Короткий» ФС анализируемых речевых сигналов представлен на рис. 2 а. КВК фазовых спектров анализируемых сигналов равен 0,5, а «коротких» спектров –

0,57. Указанные цифры свидетельствуют о меньшей информативности ФС речевого сигнала по отношению к амплитудному.

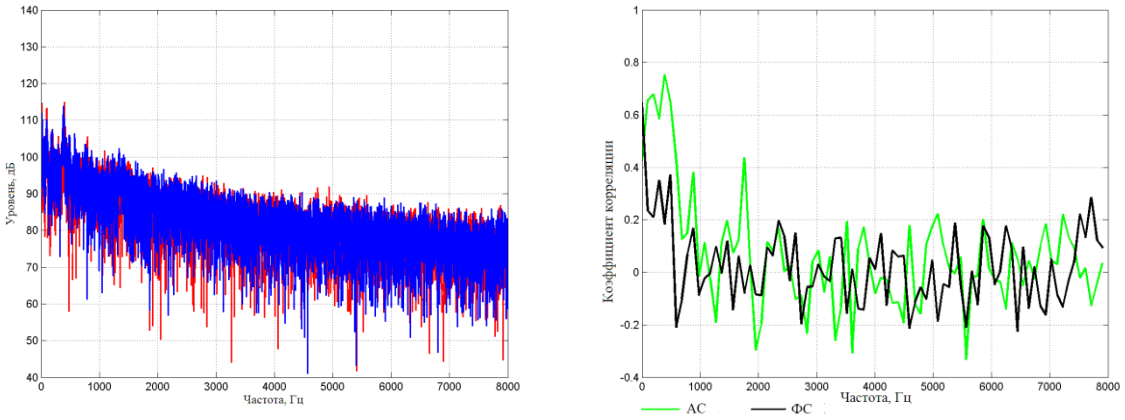
Проанализируем амплитудные и фазовые спектры исследуемых сигналов более детально. Для этого оценим КВК в «скользящем окне», которое включает часть элементов исходных спектров. После расчета и регистрации одного значения КВК «скользящее окно» сдвигается на один отсчет. Далее расчет величины k повторяется.

Здесь обратим внимание на порядок выбора размера «скользящего окна» (или выбора числа элементов последовательности), по которым будет осуществляться расчет текущей оценки коэффициента взаимной корреляции. Математические соображения, лежащие в основе выбора ширины «скользящего окна», должны отвечать двум противоречивым требованиям, а именно:

- размер «скользящего окна» должен быть достаточно широким для обеспечения хороших статистических свойств рассчитываемой оценки;

- размер «скользящего окна» должен быть как можно меньше для того, чтобы «прорисовывалась тонкая структура» линии регрессии, в частности вершины и щели мультиплетов, зависимости коэффициента взаимной корреляции.

Исследования показали, что для расчетов целесообразно выбрать ширину «скользящего окна» в 100 элементов. Это значение будет удовлетворять указанным выше требованиям. Графики полученных зависимостей КВК исследуемых спектров двух цифр «1», рассчитанные указанным выше способом, представлены на рис. 2 б.



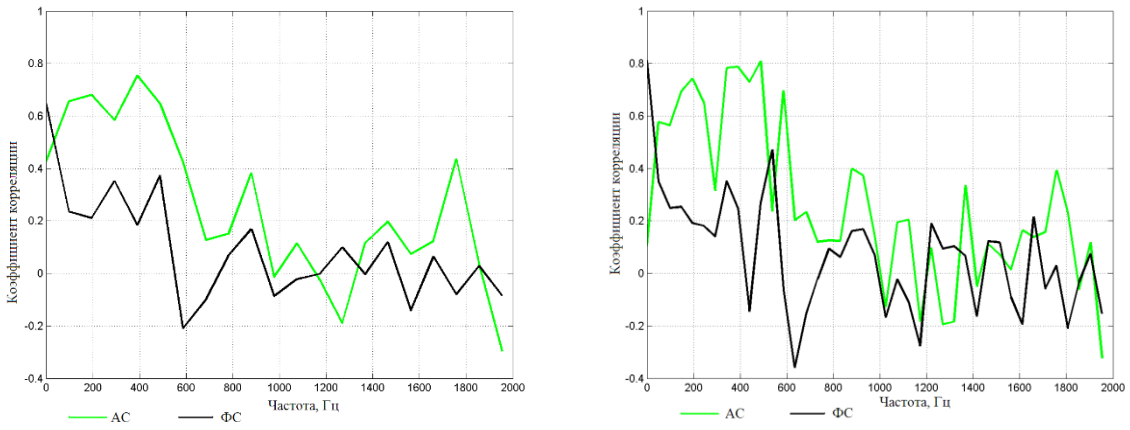
а) б)

Рис. 2. «Короткий» фазовый спектр (а) и коэффициент корреляции анализируемых спектров речевого сигнала цифры «1» (б)

Поскольку максимум КВК достигается в области низких частот (см. рис. 2 б), ниже проанализируем указанную часть спектров (см. рис. 3 а). Анализ рис. 3а свидетельствует о большей информативности АС по отношению к ФС. При этом максимум коэффициента корреляции АС анализируемых сигналов достигается в области 400 Гц. Здесь же находиться локальный минимум КВК для ФС.

Для детального исследования характера полученных зависимостей уменьшим размер «скользящего окна». На рис. 3б представлены зависимости, полученные для

размера «скользящего окна» равного 50 отсчетам. Анализ представленных на рис. 3б зависимостей свидетельствует о том, что уменьшение размера «скользящего окна» позволяет получить более ярко выраженные максимумы исследуемого показателя.



а) б)

Рис. 3. Коэффициент корреляции амплитудных и фазовых спектров речевого сигнала цифры «1» в области низких частот (а - окно 100 отсчетов; б - окно 50 отсчетов)

При этом для АС имеем максимум оценки КВК в районе 200 Гц и вначале канала «тональной» частоты. Для ФС максимумы исследуемой оценки находятся вначале канала «тональной» частоты и в районе 500 Гц.

Вместе с тем, информативность ФС голосовых сигналов существенного уступает АС, что противоречит результатам, полученным в [1] при обработке информации изображений. Причиной низкой информативности фазовых характеристик может быть их недостаточное качество. Для расчета фазы использовались данные преобразования Гильберта, которое может давать ошибочные результаты при обработке речевых сигналов, являющихся полигармоническими и не всегда стационарными. Данное предположение требует дополнительных исследований.

Выводы

Рассмотрена задача сравнительной оценки информативности амплитудного и фазового спектров речевого сигнала пользователя в системах голосовой аутентификации. Результаты получены в процессе цифровой обработки экспериментальных данных речевого сигнала пользователя, который вводился в компьютер с помощью микрофона и усилителя. Сравнение спектров выполнялось для одной и той же цифры, которая произносилась пользователем системы доступа несколько раз. В процессе анализа материалов регистрации речевых сигналов при решении задач аутентификации пользователя особое внимание целесообразно уделять области низких частот (до 600 Гц), где находятся максимумы коэффициента взаимной корреляции. Более информативным является амплитудный спектр речевого сигнала, который имеет примерно в два раза больший коэффициент корреляции.

Низкая информативность фазового спектра может быть обусловлена тем, что для расчета квадратурной составляющей речевого сигнала использовалось преобразование Гильберта. Вопросы оценки качества формирования квадратурной составляющей для речевого сигнала, который является полигармоническим и не всегда стационарным, требуют дальнейших исследований. Полученные результаты могут оказаться полезными и при решении иных задач, связанных с обработкой речевых сигналов, например, при распознавании речи, построении систем физического доступа.

Список литературы:

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

2. *Beigi H.* Fundamentals of Speaker Recognition – NY: Springer, 2011. – 1029 c.

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. *Sorokin V.N., Tsyplikhin A.I.* Speaker verification using the spectral and time parameters of voice signal // Journal of Communications Technology and Electronics. – 2010. – V. 55, N 12. – P. 1561–1574.

5. *Davis S., Mermelstein P.* Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Trans. Acoustics, Speech, Signal Process. – 1980. – V. 28. N 4. – P. 357–366.

6. *Patterson R.D., Holdsworth J.* A functional model of neural activity patterns and auditory images // Advances in Speech, Hearing and Language Processing. – 1996. – V. 3. – P. 547–563.

7. *Файзулаева О.Н., Невлюдов И.Ш.* Пути улучшения качества речевого сигнала пользователя систем голосовой аутентификации // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – Выпуск 2 (90). – С. 118–123.

8. *Файзулаева О.Н., Невлюдов И.Ш.* Экспериментальные исследования программно-аппаратных средств ввода и выделения речевого сигнала пользователя систем голосовой аутентификации // Научно-технический вестник информационных технологий, механики и оптики. – 2014. – Выпуск 5 (93). – С. 77–82.

9. *Пастушенко Н.С.* Экспериментальное исследование информативности амплитудного спектра голосового сигнала для аутентификации пользователя [Электронный ресурс] / Н.С. Пастушенко, Б.Д. Малонга, О.Н. Файзулаева // Проблеми телекомунікацій. – 2015. – № 2 (17). – С. 3–11. – Режим доступа к журн.: http://pt.journal.kh.ua/2015/2/1/152_pastushenko_research.pdf.

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

11. *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.

12. *Гмурман В.Е.* Теория вероятностей и математическая статистика. – М.: Высшая школа, 1999. – 479 с.

13. *Бендат Дж.* Прикладной анализ случайных данных: пер с англ. / Дж. Бендат, А. Пирсол: пер. с англ. под ред. И.Н. Коваленко. – М.: Мир, 1989. – 540 с.