

УДК 057.087:621.391

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИНФОРМАТИВНОСТИ АМПЛИТУДНОГО СПЕКТРА ГОЛОСОВОГО СИГНАЛА ДЛЯ АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ



[Н.С. ПАСТУШЕНКО](#), [Б.Дж. МАЛОНГА](#), [О.Н. ФАЙЗУЛАЕВА](#)

Харьковский национальный
университет радиоэлектроники

Abstract – In recent years, the problem of reliable access to information and financial resources of various organizations has exacerbated. The use of the biometric characteristics of the user (fingerprints, facial and iris image) has not given the desired results. In this regard, there are current important studies devoted to the application of dynamic biometric features in access systems, and above all, the user voice signal, which allow quick increase of the analyzed sequence to provide the required reliability. The subject of research is methods of user voice authentication. The voice authentication systems use voice signal frequency band in the range of 0.1-0.3 kHz; 4-5 kHz; 6.5-7.8 kHz. A small part of the signs is concentrated in the region of 1 kHz. Now features of the user are analyzed in the spectral region. However, up to the present moment the informative value of the portions of the spectrum of the voice signal used for user authentication has not been estimated. The problem of experimental estimation for the informativeness of different portions of the voice signal amplitude spectrum is proposed and solved in the article. The voice signal of the numbers from 0 to 9 was investigated as the analyzed digital sequence. The estimated parameter is the cross-correlation coefficient of individual portions in the spectrum of the identical numbers sequence. The studies have shown that the most informative is the portion of the spectrum from 100 to 600 Hz. The area above 4 kHz is less informative. The spectrum has a large enough cross-correlation coefficient in the range of 1 to 2 kHz. It is found that the noise signal has little impact on the informativeness of the spectrum at low frequencies. In the future it is advisable to estimate the impact of frequency spectrum portion from 1 to 2 kHz on the quality of the user authentication.

Анотація – У статті поставлено та вирішено завдання експериментальної оцінки інформативності різних ділянок амплітудного спектра голосового сигналу. Оцінюваним параметром був коефіцієнт взаємної кореляції окремих ділянок спектра послідовності однакових цифр. Проведені дослідження показали, що найбільш інформативним є ділянка спектра від 100 до 600 Гц. Область вище 4 кГц є менш інформативною. Установлено, що шумовий сигнал незначно впливає на інформативність спектра в області низьких частот. Надалі доцільно оцінити вплив ділянки спектра частот від 1 до 2 кГц на якість аутентифікації користувача.

Аннотация – В статье поставлена и решена задача экспериментальной оценки информативности различных участков амплитудного спектра голосового сигнала. Оцениваемым параметром являлся коэффициент взаимной корреляции отдельных участков спектра последовательности одинаковых цифр. Проведенные исследования показали, что наиболее информативным является участок спектра от 100 до 600 Гц. Область выше 4 кГц является менее информативной. Установлено, что шумовой сигнал незначительно влияет на информативность спектра в области низких частот. В дальнейшем целесообразно оценить влияние участка спектра частот от 1 до 2 кГц на качество аутентификации пользователя.

Введение

В настоящее время обостряется проблема обеспечения безопасности финансов, информации, услуг и ресурсов, доступ к которым осуществляется с помощью современных телекоммуникационных и компьютерных систем разного назначения. Об этом свидетельствуют многочисленные периодические сообщения в прессе. Здесь же следует заметить, что западные финансовые учреждения пытаются не афишировать случаи хищения средств до 100 тысяч USD.

В связи с этим после известных драматических событий 11.09.2001 года странами G8 было принято решение, ориентированное на снижение рисков и повышение эффективности систем доступа различного назначения. Для этого было предложено использовать в таких системах биометрические характеристики пользователя. К сожалению, в качестве основных характеристик рекомендовано использовать физиологические (статические) признаки пользователя, а именно, папиллярный узор пальцев, изображение лица и радужную оболочку глаза. Обусловлено это тем, что дактилоскопия, как и фото, очень широко и эффективно применяется в криминалистике для идентификации преступников. Более того, накоплены и интенсивно пополняются большие базы отпечатков, особенно в странах Запада. Здесь же заметим, что указанные статические биометрические признаки обладают ограниченной информативностью.

За последнее десятилетие биометрические технологии стали активно применяться во многих областях, связанных с обеспечением безопасности доступа к информации и материальным объектам, а также в задачах уникальной идентификации личности. Во многом этому способствовало распространение микропроцессорных технологий. Вместе с тем, предложенные биометрические признаки не позволили существенно повысить надежность систем доступа. Обусловлено это тем, что как в криминалистике, так и любой биометрической системе основными характеристиками являются два числа – FAR (False Acceptance Rate, ложный доступ в систему) и FRR (False Rejection Rate, ложный отказ в доступе) [1]. Применительно к системе аутентификации первое число характеризует вероятность ложного совпадения биометрических характеристик двух людей. Второе – вероятность отказа доступа человеку, имеющему допуск. Система тем лучше, чем меньше значение FRR при одинаковых значениях FAR. В отличие от криминалистики система аутентификации должна быть устойчива к подделке. Последнее не присуще криминалистике. Устойчивость к подделке – это эмпирическая характеристика, обобщающая то, насколько легко обмануть биометрическую систему.

Применительно к рассматриваемым статическим биометрическим признакам можно констатировать, что они не удовлетворяют требованиям по устойчивости к подделке. Например, давно известны опыты японского криптографа Цутому Мацумото (Tsutomu Matsumoto), которые позволили вскрыть от 80 до 100% тестируемых дактилоскопических систем доступа. Из-за низкой устойчивости переходят от простейших к трехмерным изображениям пользователя. Появились сообщения о подделке радужной оболочки глаза, которые формируют на основе нескольких фото с помощью современной фотоаппаратуры.

В связи с этим в последнее время все больше внимания уделяется поведенческим (динамическим) признакам пользователя, а именно, подпись (форма букв, манера письма, нажим), голос, клавиатурный почерк и др. Основное преимущество указанных признаков – оперативное наращивание анализируемой последовательности по требованию системы. В общем случае можно утверждать, что указанные при-

знаки имеют неограниченную информативность. Это существенно влияет на снижение величин FRR и FAR, а также повышает устойчивость к подделке.

Указанное преимущество особо ярко проявляется для систем голосовой аутентификации (СГА) [2, 3]. Наряду с отмеченным выше СГА обладают рядом дополнительных преимуществ, таких как: простота, компактность, дешевизна, возможность удаленной аутентификации с использованием телефонных каналов связи и др. [4]. Вопросам развития и внедрения этих систем сегодня посвящены многочисленные исследования и разработки, отдельные вопросы которых рассмотрены в работах Г. Фанта, Р.М. Болла, Г.С. Рамишвилли, Ф. Россе, В.Н. Сорокина, Г. Холлиена и др. В работах [5, 6] установлено, что отличительные признаки пользователя в основном находятся за пределами стандартного телефонного канала (СТК) в диапазонах 0,1–0,3 кГц; 4–5 кГц; 6,5–7,8 кГц. Лишь незначительная часть отличительных признаков сосредоточена в районе 1 кГц. Поэтому речевой сигнал пользователя в СГА должен регистрироваться и анализироваться в широкой полосе частот (от 0,1 до 8 кГц).

Здесь же следует отметить, что в современных СГА для идентификации используются преимущественно спектральные характеристики речевого сигнала пользователя. Например, в [7] использовалась огибающая спектра голосового источника, в [8] предложен метод кепстрального преобразования спектра речевых сигналов, в [9] применялась модель, в которой спектрально-временные характеристики речевого сигнала анализируются гребенкой фильтров.

Выполненный анализ работ в области голосовой аутентификации показал, что открытым остается вопрос оценки информативности амплитудного спектра речевого сигнала пользователя. Поэтому цель данной работы – анализ информативности различных частотных участков амплитудного спектра речевого сигнала и выявление диапазона частот, в котором сосредоточены основные признаки пользователя.

I. Методика и результаты исследований амплитудного спектра речевого сигнала

Анализу подвергался речевой сигнал пользователя цифр 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 составляла 64 кГц. Отношение сигнал/шум анализируемой последовательности составляло более 25 дБ.

При этом основное внимание будем уделять анализу диапазона спектра до 8 кГц, что обусловлено наличием отличительных признаков пользователя в его речевом сигнале (в диапазоне от 0,1 кГц до 8 кГц). Для этого рассчитанный спектр, диапазон изменения которого определяется половиной частоты временной дискретизации, будем ограничивать частотой 8 кГц (далее «короткий» спектр).

Очевидно 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» и «короткий» спектр одного из указанных сигналов (см. рис. 1).

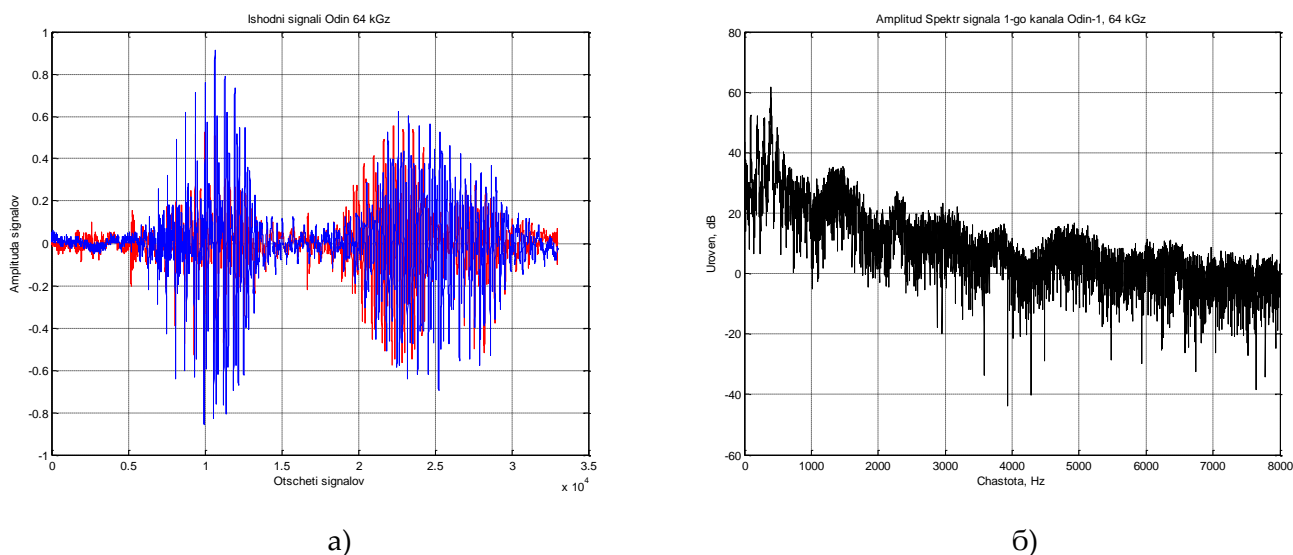


Рис. 1. Речевые сигналы цифры «1» и «короткий» амплитудный спектр
(а – исходные сигналы; б – спектр сигнала)

В качестве оцениваемого параметра будем использовать коэффициент взаимной корреляции. Для расчета коэффициента взаимной корреляции использовалось известное соотношение для двух дискретных последовательностей [10]:

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

где K_i и \hat{K}_i – анализируемые цифровые последовательности, $i = 1, \dots, N$ – 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 спектры.

Заметим, что коэффициент корреляции анализируемых речевых сигналов равен 0,3. Низкий коэффициент корреляции анализируемых речевых сигналов обусловлен тем, что не выполнены процедуры масштабирования и передискретизации, которые во временной области выполнить достаточно сложно. Коэффициент корреляции амплитудных спектров анализируемых сигналов равен 0,75, т.е. в частотной

области корреляция рассматриваемых сигналов выше. Заметим, что коэффициент взаимной корреляции «коротких» спектров выше и равен 0,83.

Проведем более детальный анализ изменения коэффициента корреляции анализируемых спектров. Для этого будем оценивать коэффициент взаимной корреляции в «скользящем окне», которое включает часть элементов исходных спектров. После расчета одного значения коэффициента взаимной корреляции «скользящее окно» сдвигается на один отсчет. Далее расчет величины k повторяется.

Здесь обратим внимание на порядок выбора размера «скользящего окна» или выбор числа элементов последовательности, по которым будет осуществляться расчет текущей оценки коэффициента взаимной корреляции. Математические соображения, лежащие в основе выбора ширины «скользящего окна», отвечают 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.

Проведенные исследования показали, что для расчетов целесообразно выбрать ширину «скользящего окна» в 100 элементов. Последнее значение будет удовлетворять указанным выше требованиям.

Графики полученной зависимости коэффициента взаимной корреляции спектров двух цифр «1» от номера отсчета представлены на рис. 2.

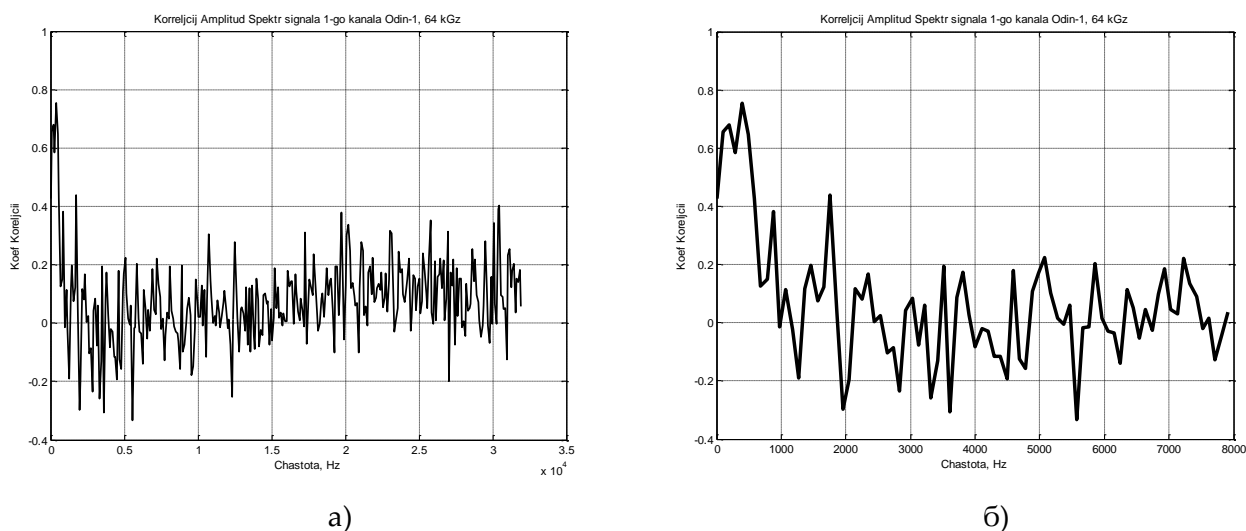


Рис. 2. Коэффициент корреляции амплитудных спектров речевого сигнала цифры «1»
(а – полного спектра; б – «короткого» спектра)

Анализ представленных спектров показывает, что наиболее информативные участки: диапазон до 500 Гц; небольшие и менее информативные участки у 1кГц и 2 кГц. Заметим, что частотные участки выше 4 кГц являются менее информативными.

В связи с этим есть смысл более детально исследовать поведение оцениваемого показателя при разной ширине «скользящего окна» для диапазона частот до 2кГц (см. рис. 3).

Анализ представленных зависимостей свидетельствует, что меньшее окно дает возможность более точно определить максимумы коэффициента взаимной корреляции. Например, при ширине «скользящего окна» в 50 отсчетов вершины (максимумы) исследуемой зависимости являются более выраженными.

Зависимость, представленная на рис. 3б, дает возможность установить следующие максимумы: 200, 380, 500, 600, 900 и 1780 Гц. Часть из указанных максимумов совпадает с ранее известными результатами [5, 6], где наиболее информативными участками считают диапазон от 100 до 300 и 1000 Гц. Вместе с тем, выявлены еще несколько участков спектра, на которые следует обратить внимание: это частоты с максимумами у 500, 600 и 1780 Гц. При этом целесообразно убедиться, что данные максимумы не обусловлены содержанием анализируемой цифры. Для этого проанализируем аналогичные зависимости для цифры «2» (см. рис. 4).

Анализ зависимостей на рис. 4, в основном, подтверждает ранее сделанные выводы, т.е. целесообразно более внимательно относиться к области частот от 500 до 600 Гц, а также к диапазону от 1 до 2 кГц.

Теперь проанализируем помеховый и шумовой сигнал, который регистрировался между цифрами, вводимыми пользователем. На рис.5 представлен шумовой сигнал и его спектр.

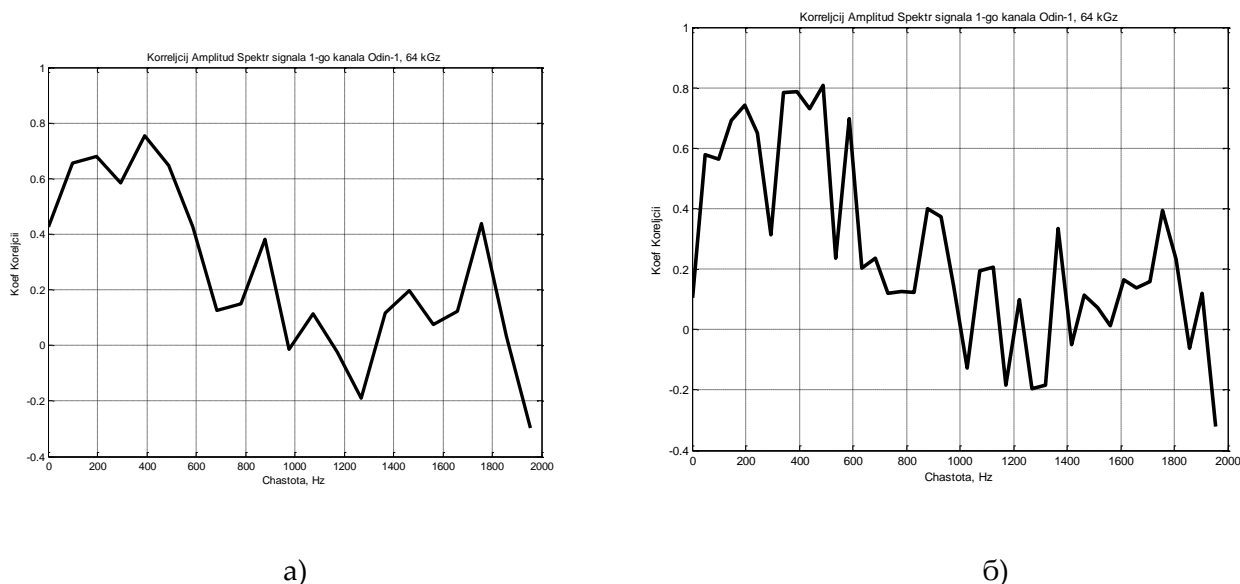
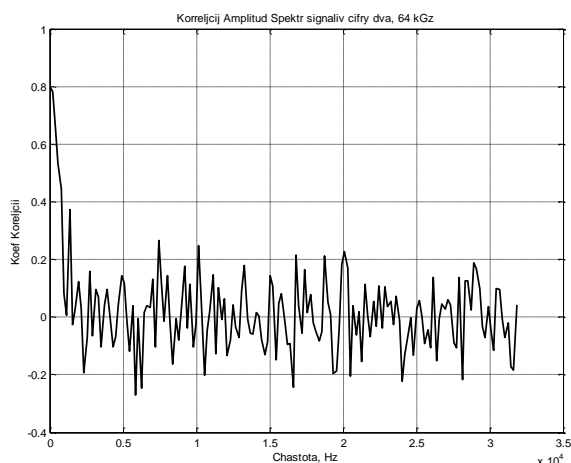
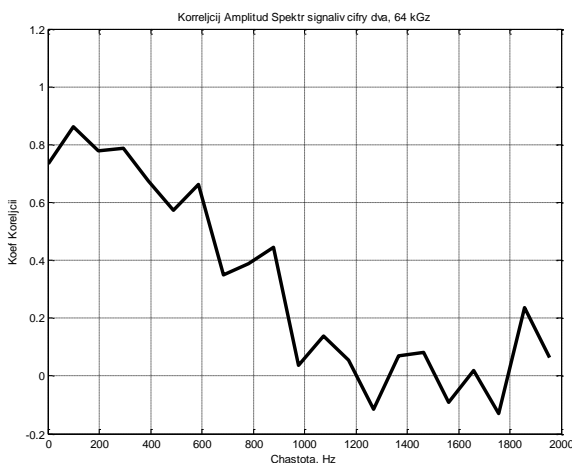


Рис. 3. Коэффициент корреляции амплитудных спектров речевого сигнала цифры «1» для различной ширины «скользящего окна»
(а – 100 отсчетов; б – 50 отсчетов)



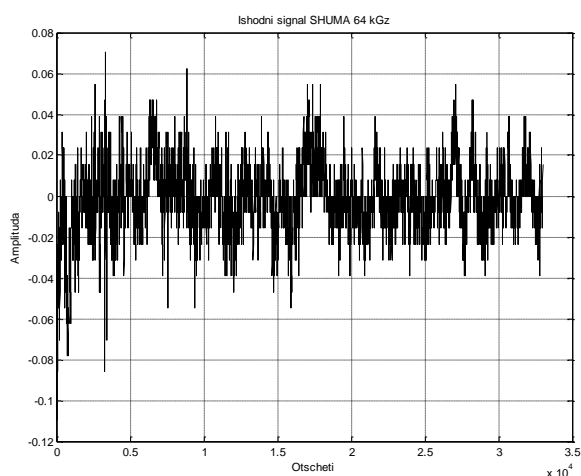
а)



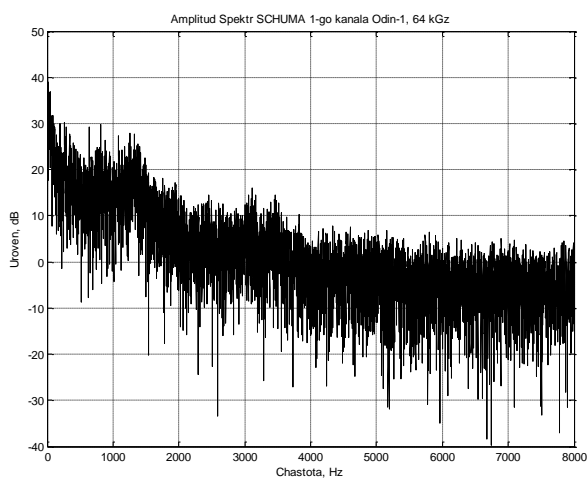
б)

Рис. 4. Коэффициент корреляции амплитудных спектров речевого сигнала цифры «2» для различной ширины «скользящего окна» (а – 100 отсчетов; б – 50 отсчетов)

Анализ представленного спектра свидетельствует, что основная интенсивность мешающих сигналов сосредоточена в области низких частот (там, где наибольший коэффициент корреляции).



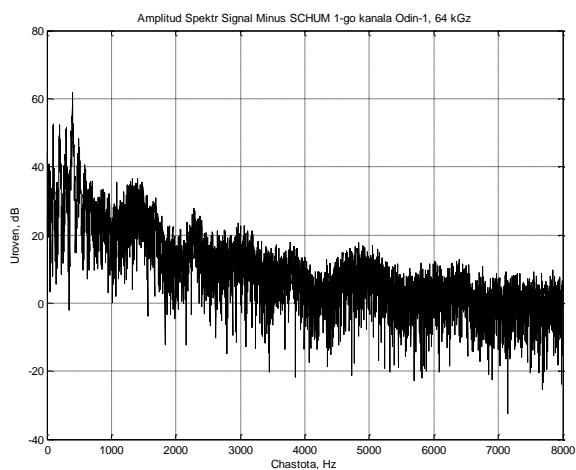
а)



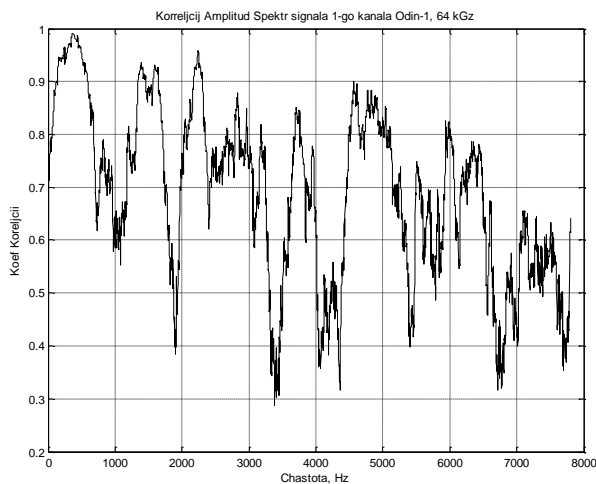
б)

Рис. 5. Шумовой сигнал и его спектр (а – сигнал; б – спектр)

Оценим влияние шумовой составляющей. Для этого из спектра смеси полезного сигнала и шума вычтем спектр шумовой составляющей, которая регистрировалась до (или после) полезного сигнала. После вычитания спектра шума получили результаты, представленные на рис. 6.



а)

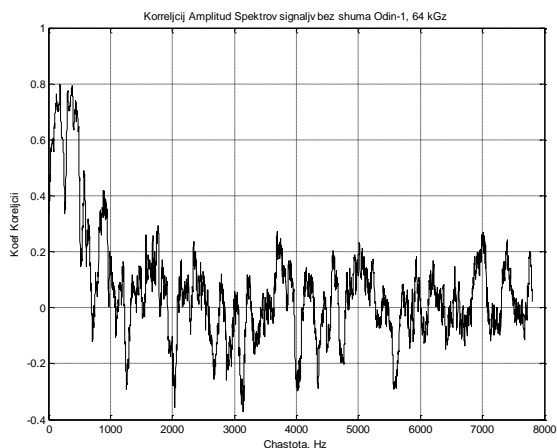


б)

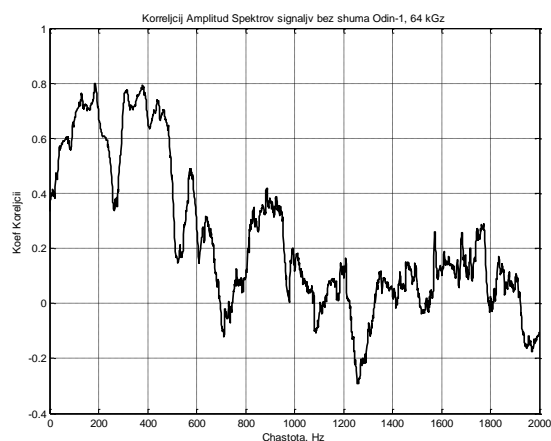
Рис. 6. Спектр первого сигнала без шума и его коэффициент корреляции (а – спектр сигнала; б – коэффициент корреляции спектров сигнала с шумом и без шума)

Провалы в графике коэффициента взаимной корреляции указывают на области, где шум воздействует на полезный сигнал. Аналогичные результаты получены и для второго сигнала цифры «1».

Теперь проанализируем коэффициент взаимной корреляции двух сигналов после вычитания шума (см. рис. 7).



а)



б)

Рис. 7. Коэффициент взаимной корреляции амплитудных спектров двух сигналов цифры «1» без шума

Сравнительный анализ зависимостей, представленных на рис. 3б и 7б, свидетельствует о незначительном влиянии шумовой составляющей на анализируемый показатель. Таким образом, при достаточно высоком отношении сигнал/шум анализируемых последовательностей информативные участки спектра сохраняют свои свойства.

Выводы

Рассмотрена задача оценки информативности участков амплитудного спектра голосового сигнала в интересах аутентификации пользователя. В качестве показателя использовался коэффициент взаимной корреляции голосовых сигналов последовательности цифрового ряда от «0» до «9». Проведенные исследования показали, что наиболее информативными являются участки амплитудного спектра в области низких частот. При этом, наряду с известным участком спектра в области до 300 Гц, который используется для аутентификации пользователя, необходимо иметь в виду следующее. В области частот 500, 600 и около 2000 Гц есть участки спектра, коэффициент взаимной корреляции которых более чем в два раза превышает оцениваемый параметр в диапазоне частот выше 4 кГц. Шумовой компонент регистрируемого сигнала незначительно влияет на информативность анализируемых участков амплитудного спектра. Дальнейшие исследования целесообразно будет ориентировать на оценку влияния выявленных областей спектра на качество аутентификации пользователя.

Список литературы:

1. Пастушенко О.Н., Невлюдов И.Ш. Анализ качественных показателей биометрических систем аутентификации пользователей // Проблемы телекоммуникаций. – 2012. – № 4 (9). – С. 96-103. – Режим доступа: http://pt.journal.kh.ua/2012/4/1/124_pastushenko_biometric.pdf
2. Файзулаева О.М. Методи підвищення якості виділення мовних сигналів для голосової аутентифікації користувачів: Автореф. дис. ... кандидата техн. наук / Харківський національний університет радіоелектроніки. – Х., 2015. – 20 с.
3. ГОСТ 52633-2006 «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации». – М: Стандартинформ, 2007. – 17 с.
4. Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. – 2012. – Т. 12, № 1. – С. 1-30.
5. Besacier L., Bonastre J.-F. Subband architecture for automatic speaker recognition // Signal Process. – 2000. – Vol. 80. – P. 1245-1259.
6. Lu X., Dang J. An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification // Speech Communication. – 2007. – Vol. 50, No. 4. – P. 312-322.
7. Sorokin V.N., Tsyplikhin A.I. Speaker verification using the spectral and time parameters of voice signal // Journal of Communications Technology and Electronics. – 2010. – Vol. 55, No. 12. – P. 1561-1574.
8. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Trans. Acoustics, Speech, Signal Process. – 1980. – Vol. 28, No. 4. – P. 357-366.
9. Patterson R.D., Holdsworth J. A functional model of neural activity patterns and auditory images // Advances in Speech, Hearing and Language Processing. – 1996. – Vol. 3. – P. 547-563.
10. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1999. – 479 с.