

УДК 057.087.1:621.391.26

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ПРОГРАММНО-АППАРАТНЫХ СРЕДСТВ ВВОДА И ВЫДЕЛЕНИЯ РЕЧЕВОГО СИГНАЛА ПОЛЬЗОВАТЕЛЯ СИСТЕМ ГОЛОСОВОЙ АУТЕНТИФИКАЦИИ

О.Н. Файзулаева^а, И.Ш. Невлюдов^а

^аХарьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, Olga_Kharkov_2014@mail.ru

Аннотация. Рассматривается научная задача повышения отношения сигнал/шум речевого сигнала пользователя компьютерных систем и сетей в процессе его голосовой аутентификации. Объект исследования – процесс ввода и выделения речевого сигнала пользователя системы аутентификации в компьютерных системах и сетях. Исследуются методы и средства ввода и выделения речевого сигнала на фоне внешних помеховых сигналов. Экспериментально исследуются пути повышения качества речевого сигнала пользователя в системах голосовой аутентификации. Рассматриваются программно-аппаратные средства экспериментальной установки для ввода и выделения речевого сигнала на фоне внешних помеховых воздействий. Поскольку современные компьютерные средства, в том числе и мобильные, оснащены двухканальной звуковой картой, при вводе речевого сигнала использовались два микрофона. Расстояние между приемниками акустических волн составляло 20 мм и обеспечивало формирование одного лепестка диаграммы направленности микрофонной решетки в требуемой области регистрации речевого сигнала (от 100 Гц до 8 кГц). Как свидетельствуют результаты экспериментальных исследований, использование направленных свойств предложенной микрофонной решетки и применение пространственно-временной обработки регистрируемых сигналов с адаптивными весовыми коэффициентами позволило существенно снизить влияние помеховых сигналов. Представлены результаты экспериментальных исследований предложенных программно-аппаратных средств ввода и выделения речевых сигналов на фоне помеховых воздействий. Предложенные решения позволяют повысить значение отношения сигнал/шум регистрируемых полезных сигналов до 20 дБ при воздействии внешних помеховых сигналов в диапазоне частот от 4 до 8 кГц. Полученные результаты могут быть полезными специалистам, выполняющим исследование в области распознавания речи, а также различения диктора.

Ключевые слова: аутентификация, антенная решетка, диаграмма направленности, речевой сигнал, квадратурная обработка, микрофон.

EXPERIMENTAL STUDY OF FIRMWARE FOR INPUT AND EXTRACTION OF USER'S VOICE SIGNAL IN VOICE AUTHENTICATION SYSTEMS

O.N. Faizulaieva^а, I. Sh. Nevlyudov^а

^аKharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, Olga_Kharkov_2014@mail.ru

Abstract. Scientific task for improving the signal-to-noise ratio for user's voice signal in computer systems and networks during the process of user's voice authentication is considered. The object of study is the process of input and extraction of the voice signal of authentication system user in computer systems and networks. Methods and means for input and extraction of the voice signal on the background of external interference signals are investigated. Ways for quality improving of the user's voice signal in systems of voice authentication are investigated experimentally. Firmware means for experimental unit of input and extraction of the user's voice signal against external interference influence are considered. As modern computer means, including mobile, have two-channel audio card, two microphones are used in the voice signal input. The distance between sonic-wave sensors is 20 mm and it provides forming one direction pattern lobe of microphone array in a desired area of voice signal registration (from 100 Hz to 8 kHz). According to the results of experimental studies, the usage of directional properties of the proposed microphone array and space-time processing of the recorded signals with implementation of constant and adaptive weighting factors has made it possible to reduce considerably the influence of interference signals. The results of firmware experimental studies for input and extraction of the user's voice signal against external interference influence are shown. The proposed solutions will give the possibility to improve the value of the signal/noise ratio of the useful signals recorded up to 20 dB under the influence of external interference signals in the frequency range from 4 to 8 kHz. The results may be useful to specialists working in the field of voice recognition and speaker discrimination.

Keywords: authentication, array, direction pattern, voice signal, quadrature processing, microphone.

Введение

В последнее время исследователи большое внимание уделяют совершенствованию систем аутентификации вычислительных систем и сетей на основе использования биометрических признаков пользователя. В ряде случаев отдается предпочтение системам голосовой аутентификации. Они имеют ряд преимуществ – простота устройств ввода, удобство ввода парольного сообщения, которое при необходимости может наращиваться, устойчивая работа в каналах связи и др. Вместе с тем современные системы голосовой аутентификации имеют показатель равной ошибки (EER – Equal Error Rates) в диапазоне 3–5%, что подчеркивает актуальность исследований в области их совершенствования [1].

В [2] предложено для совершенствования программно-аппаратных средств ввода и выделения речевого сигнала в системах голосовой аутентификации использовать два канала регистрации акустических волн, микрофоны в которых расположены на расстоянии $d = 20$ мм, с последующей квадратурной обработкой материалов регистрации.

Целью настоящей работы являются экспериментальные исследования предложенных программно-аппаратных решений и оценка их адекватности, достоверности и эффективности. Здесь же отметим, что в большинстве работ для голосовой аутентификации используются преимущественно спектральные характеристики речевого сигнала пользователя. Например, в [3] использовалась огибающая спектра голосового источника, в [4] предложен метод кепстрального преобразования спектра речевых сигналов, в [5] применялась модель, в которой спектрально-временные характеристики речевого сигнала анализируются гребенкой фильтров. Исходя из этого, ниже будут подвергаться сравнительному анализу спектральные характеристики речевого сигнала, зарегистрированного с помощью одного канала (микрофона) и предложенной системой ввода и выделения речевого сигнала на базе двух каналов (использование для регистрации двух микрофонов) [2]. При этом основное внимание будем уделять анализу диапазона спектра до 8 кГц, что обусловлено наличием отличительных признаков речевого сигнала пользователя в диапазоне от 100 Гц до 8 кГц [6, 7]. Для этого рассчитанный спектр, диапазон изменения которого определяется половиной частоты временной дискретизации, будем ограничивать частотой 8 кГц («короткий» спектр).

Программно-аппаратная схема экспериментальных исследований

Структурная схема программно-аппаратных средств экспериментальной установки представлена на рис. 1. Основу экспериментальной установки составляет ноутбук, к которому подключается выносное устройство с двумя микрофонами и схемой усиления регистрируемых колебаний. Питание выносного устройства осуществляется от USB-порта ноутбука, а регистрируемые стереосигналы поступают на вход его звуковой карты. В ноутбуке установлены и используются два класса программных средств – COOL EDIT и MATLAB. Рассмотрим более подробно назначение и функции экспериментальной установки.

В качестве приемника и усилителя акустических сигналов для каждого канала регистрации в выносном устройстве использовалась одна из практических схем микрофонного усилителя.

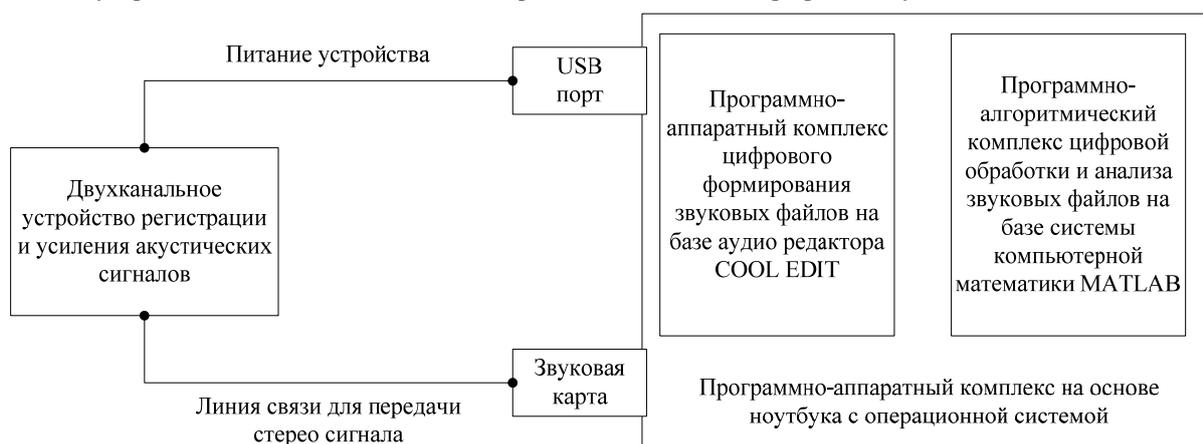


Рис. 1. Структурная схема экспериментальной установки

В ноутбуке в качестве программно-аппаратного обеспечения используется аудиоредактор COOL EDIT. Редактор является свободно распространяемым (ShareWare), его можно найти в Интернете на сайте www.syntrillium.com [8]. Он предельно прост в использовании, сохраняя при этом наиболее верные и качественные алгоритмы обработки звука. Редактор может работать с любым типом звуковых карт, при этом расширяя свои возможности в зависимости от класса качества звуковой карты. Аудиоредактор позволяет задать для звуковой карты ноутбука вид регистрируемого сигнала (моно или стерео), частоту временной дискретизации и количество бит на отсчет при регистрации, а также сформировать звуковой файл.

Сформированный файл далее поступает на цифровую обработку, процедуры которой реализуются в системе компьютерной математики MATLAB [9]. Преимущества системы MATLAB обусловлены тем, что можно использовать алгоритмический язык для реализации уникальных процедур цифровой обработки данных, а также применять готовые к использованию типовые функции для решения задач анализа цифровых данных, например, реализацию преобразования Гильберта, оценку спектра обрабатываемого сигнала и т.д.

Методика и результаты обработки материалов регистрации

Структурная схема системы выделения речевого сигнала, в которой реализована квадратурная обработка материалов регистрации, представлена на рис. 2 [2].

Ввод речевого сигнала пользователя осуществляется с расстояния 0,7–1 м по нормали к оси двухэлементной антенной решетки, т.е. в области максимума ее диаграммы направленности. Помеховый сигнал воздействует под углом к нормали оси антенной решетки.

Для формирования квадратурной составляющей используется преобразование Гильберта–Хуанга, которое не только обеспечивает качественную обработку сигналов с нестационарными и нелинейными искажениями, что присуще акустическим сигналам, но и дает возможность повысить отношение сигнал/шум (ОСШ) за счет исключения при необходимости зашумленных первых модовых функций [10–12].

Для формирования весовых коэффициентов ω_i используются адаптивные процедуры их расчета, которые имеют название алгоритма минимума среднеквадратического отклонения Уидроу–Хоффа [2]. При этом расчет адаптивных весовых коэффициентов осуществляется по шумовому сигналу, например, на этапе идентификации пользователя, а затем на этапе верификации последнему предлагается произнести заданную последовательность цифр.

Оценка качества ввода речевого сигнала производится на фоне помехового воздействия интенсивных шумов улицы, а также в замкнутом помещении. В последнем случае в качестве помеховых источников выступает шум работы коммутационной аппаратуры и серверов. В качестве полезного сигнала используется речевой сигнал цифр (от 0 до 9).

Далее представлены результаты экспериментальных исследований, полученные при частоте временной дискретизации 44 и 64 кГц.

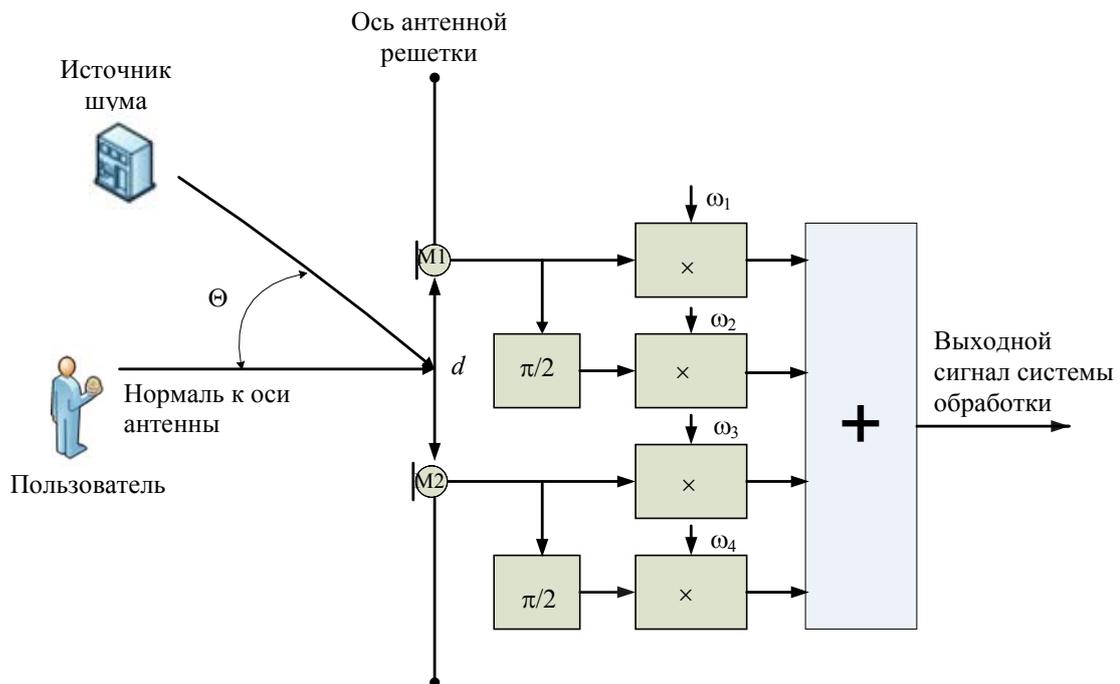


Рис. 2. Структурная схема системы ввода и выделения речевого сигнала: M1 и M2 – микрофоны; $(\pi/2)$ – программные формирователи квадратурной составляющей; ω_i – весовые коэффициенты; $i = 1, \dots, 4$; (\times) – умножители, реализующие весовую обработку; $(+)$ – сумматор

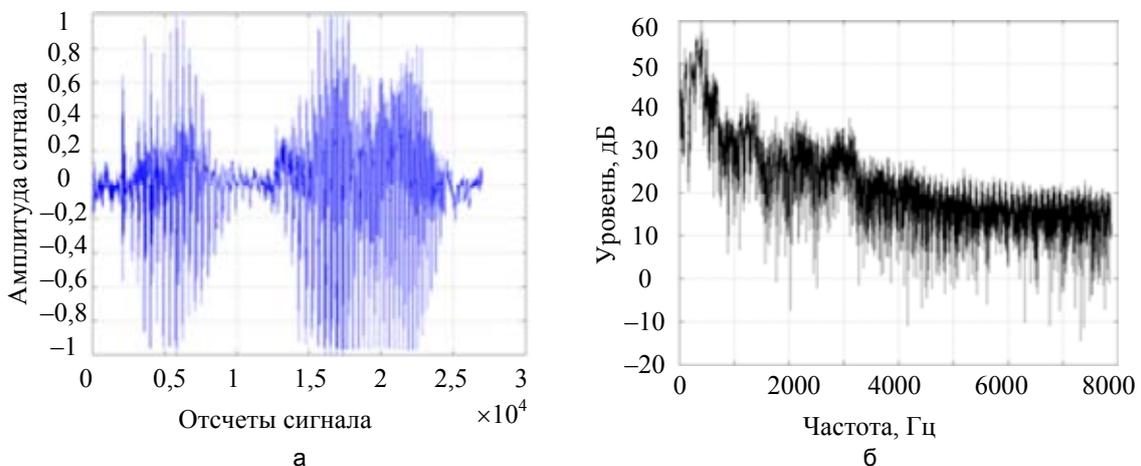


Рис. 3. Речевой сигнал цифры «1» пользователя системы аутентификации и его «короткий» спектр: речевой сигнал (а); «короткий» спектр (б). Частота дискретизации 44 кГц

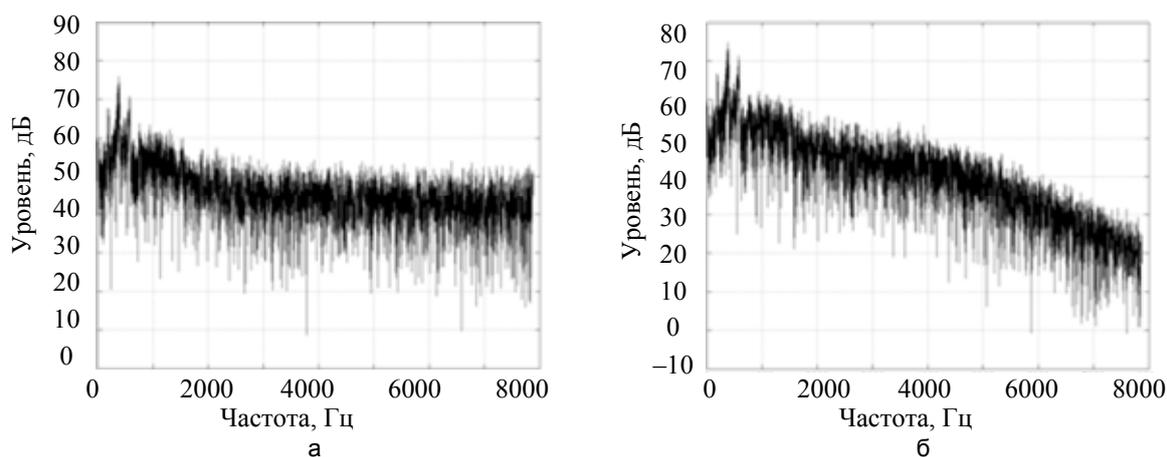


Рис. 4. Спектр смеси, зарегистрированной одним каналом (микрофоном), и спектр выходного сигнала предлагаемой системы цифровой обработки двух каналов: спектр смеси, зарегистрированной одним каналом (а); спектр выходного сигнала системы обработки (б). Частота дискретизации 44 кГц

Для примера на рис. 3, а, представлен речевой сигнал цифры «1» с частотой временной дискретизации 44 кГц, а на рис. 3, б, его «короткий» спектр.

На рис. 4, а, представлен «короткий» спектр смеси указанного речевого сигнала и шума улицы, который зарегистрирован одним из двух каналов (один микрофон). ОСШ рассматриваемой смеси составляет единицы децибел. На рис. 4, б, представлен «короткий» спектр результатов цифровой обработки двух каналов с учетом квадратурных составляющих и последующего их весового суммирования, т.е. выходной сигнал предлагаемой системы обработки (рис. 2).

Качественный анализ спектров рис. 3, б, и рис. 4, а, свидетельствует, что в области высоких частот (более 4 кГц) спектральная плотность мощности для одного канала регистрации увеличивается примерно на 20 дБ. В то же время результаты предлагаемой системы выделения речевого сигнала свидетельствуют о значительном подавлении высокочастотных составляющих, обусловленных влиянием шума.

Для количественного анализа результатов регистрации и обработки были использованы следующие показатели оценки качества ввода и выделения речевого сигнала на фоне шумовых последовательностей: коэффициент взаимной корреляции и средний «квадрат невязки».

Коэффициент взаимной корреляции оценивался между последовательностью «короткого» спектра речевого сигнала цифр (от 0 до 9) и последовательностью смеси указанного сигнала и помехового воздействия на выходе одного канала регистрации (или результатов предлагаемой схемы обработки цифровых данных двух каналов). Для расчета коэффициента взаимной корреляции использовалось известное соотношение для двух дискретных последовательностей

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

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.

Зависимость коэффициента взаимной корреляции от ОСШ представлены для различных частот временной дискретизации на рис. 5.

Из анализа представленных зависимостей следует, что коэффициент взаимной корреляции слабо изменяется при низком ОСШ. По этой причине был введен еще один показатель – средний «квадрат невязки» (η), для расчета которого использовалось соотношение

$$\eta = \frac{\sum_{i=1}^N (K_i - \hat{K}_i)^2}{N}.$$

При построении зависимостей среднего «квадрата невязок» от ОСШ выполнялась известная нормировка:

$$\hat{\eta} = \frac{\eta - \eta_{\min}}{\eta_{\max} - \eta_{\min}},$$

где η_{\min} , η_{\max} – минимальное и максимальное значение анализируемых показателей. Указанная нормировка позволит привести средний «квадрат невязок» к диапазону от 0 до 1 (нормированная «невязка»), которая, в свою очередь, упростит процесс сравнительного анализа полученных результатов.

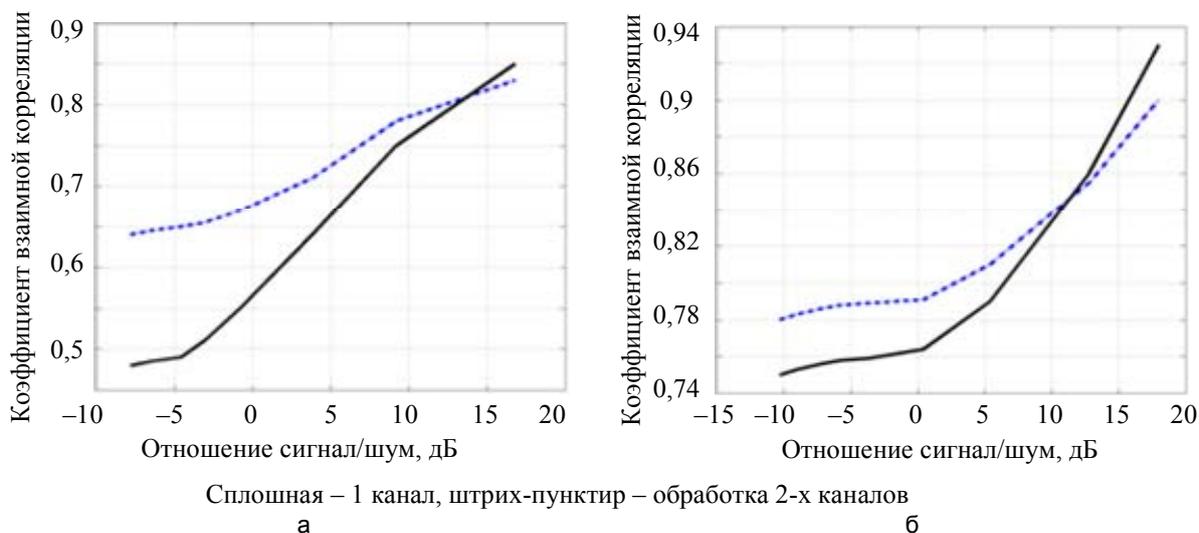


Рис. 5. Экспериментальные зависимости коэффициента взаимной корреляции от отношения сигнал/шум: частота временной дискретизации 44 кГц (а); частота временной дискретизации 64 кГц (б)

Зависимости нормированной «невязки» для частот временной дискретизации 44 и 64 кГц от ОСШ представлены на рис. 6, а, б, соответственно. Анализ представленных зависимостей позволяет сделать вывод, что предложенная система ввода и выделения речевого сигнала пользователя является более эффективной по отношению регистрации на один микрофон при обработке последовательностей с ОСШ менее 10 дБ. При ОСШ более 10 дБ возникают проблемы с формированием адаптивных весовых коэффициентов по слабому шумовому сигналу, а также сказывается влияние неидентичности характеристик каналов регистрации. В этом случае при низкой частоте временной дискретизации (менее 44 кГц) можно ограничиться применением одного канала регистрации.

Представленные результаты относятся к случаю выделения речевого сигнала цифры «1» на фоне интенсивных шумов улицы. Аналогичные зависимости получаются при обработке речевых сигналов других цифр на фоне интенсивных шумов коммутационной аппаратуры и сервера.

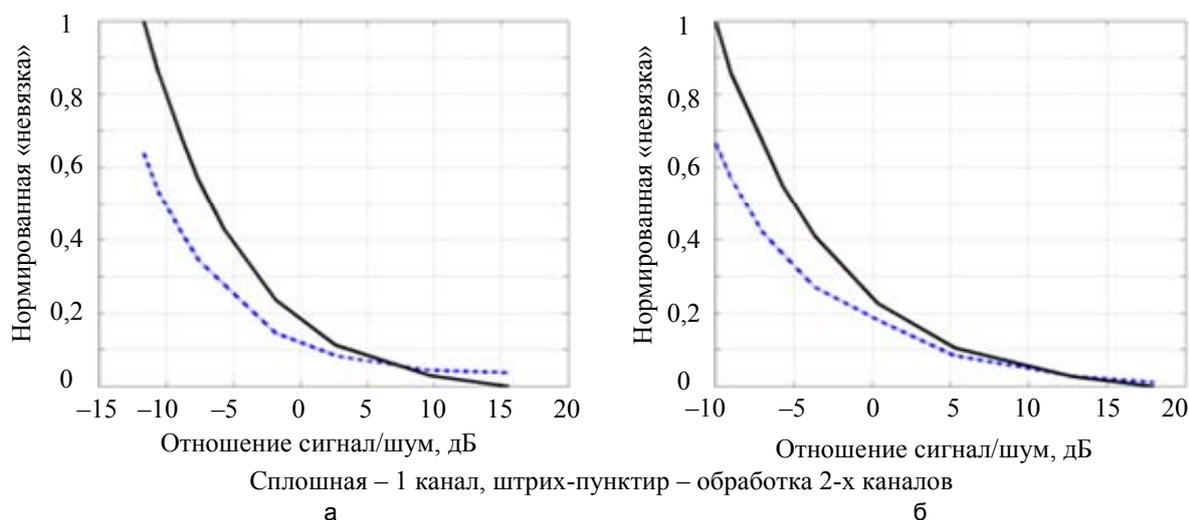


Рис. 6. Экспериментальные зависимости нормированной «невязки» от отношения сигнал/шум: частота временной дискретизации 44 кГц (а); частота временной дискретизации 64 кГц (б)

Заключение

Рассмотрены результаты экспериментальных исследований программно-аппаратных средств ввода и выделения речевого сигнала пользователя в системах голосовой аутентификации. Аппаратные средства ввода речевого сигнала включали два микрофона, расположенные на расстоянии 20 мм, и схемы предварительного усиления сигналов звуковой частоты каждого канала, которые подключались к звуко-

вой карте ноутбука. Программные средства содержали аудиоредактор COOL EDIT и систему компьютерной математики MATLAB. Аудиоредактор COOL EDIT позволял сформировать звуковой файл с заданными характеристиками, а в системе MATLAB была реализована цифровая обработка, включающая учет квадратурных составляющих регистрируемых последовательностей с использованием преобразования Гильберта–Хуанга и расчет адаптивных весовых коэффициентов на основе алгоритма минимума среднеквадратического отклонения Уидроу–Хоффа.

Предложенные технические решения позволяют повысить значение отношения сигнал/шум регистрируемых речевых сигналов до 20 дБ в диапазоне частот от 4 до 8 кГц. Особо эффективными предложенные технические решения являются при обработке цифровых последовательностей регистрируемых сигналов при отношении сигнал/шум менее 10 дБ.

Полученные результаты могут оказаться полезными и при решении иных задач, связанных с обработкой речевых сигналов в компьютерах, например, при различении диктора, построении систем физического доступа.

Дальнейшие исследования будут связаны с разработкой процедур компенсации помеховых сигналов в области низких частот.

Литература

1. Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. 2012. Т. 12. № 1. С. 1–30.
2. Файзулаева О.Н., Невлюдов И.Ш. Пути улучшения качества речевого сигнала пользователя систем голосовой аутентификации // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 118–123.
3. 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.
4. Davis S.B., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Transactions on Acoustics, Speech, Signal Process. 1980. V. ASSP-28. N 4. P. 357–366.
5. Patterson R.D., Holdsworth J. A functional model of neural activity patterns and auditory images: In Advances in Speech, Hearing and Language Processing. 1996. V. 3. P. 547–563.
6. Besacier L., Bonastre J.-F. Subband architecture for automatic speaker recognition // Signal Processing. 2000. V. 80. N 7. 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. Петелин Р.Ю., Петелин Ю.В. Cool Edit Pro 2. Секреты мастерства. СПб. БХВ-Петербург, Арлит, 2002. 432 с.
9. Дьяконов В.П. MATLAB 6/6.1/6.5 + Simulink 4/5 Основы применения. М.: СОЛОН-Пресс, 2004. 768 с.
10. Белоусова Е.Э., Пастушенко Н.С., Пастушенко О.Н. Анализ влияния частоты дискретизации на качество формирования квадратурной составляющей аналитического сигнала // Восточно-Европейский журнал передовых технологий. 2013. Т 1. № 9 (61). С. 8–13.
11. Белоусова Е.Э., Пастушенко О.Н. Анализ влияния частоты дискретизации на качество формирования квадратурной составляющей для некоторых сигналов // Радиотехника. 2013. № 172. С. 141–146.
12. Файзулаева О.Н. Автоматизация процедур принятия решения об исключении из обработки первой модовой функции при использовании преобразования Гильберта-Хуанга // Радиотехника. 2013. № 175. С. 147–153.

Файзулаева Ольга Николаевна	– аспирант, Харьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, Olga_Kharkov_2014@mail.ru
Невлюдов Игорь Шакирович	– доктор технических наук, профессор, заведующий кафедрой, Харьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, tapr@kture.kharkov.ua
Olha N. Faizulaieva	– postgraduate, Kharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, Olga_Kharkov_2014@mail.ru
Ihor Sh. Nevlyudov	– D.Sc., Professor, Department head, Kharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, tapr@kture.kharkov.ua

Принято к печати 18.06.14
Accepted 18.06.14