

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК АЗЕРБАЙДЖАНА
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

На правах рукописи

СУХОСТАТ ЛЮДМИЛА ВАЛЕНТИНОВНА

**РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ДЛЯ СИНТЕЗА
СИСТЕМ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ
ЛИЧНОСТИ ПО ГОЛОСУ**

3338.01 – Системный анализ, управление и обработка информации

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
доктора философии по технике

БАКУ - 2015

Работа выполнена в Институте Информационных Технологий
Национальной Академии Наук Азербайджана

Научный руководитель:

доктор философии по технике

Я.Н.ИМАМВЕРДИЕВ

Официальные оппоненты:

доктор наук по технике, профессор

Ф.Г.МАМЕДОВ

доктор философии по технике, доцент

Н.Д.ДЖАФАРОВ

Ведущая организация: Институт Систем Управления Национальной Академии Наук Азербайджана

Защита состоится «25» июня 2015 года в 14⁰⁰ часов на заседании диссертационного совета FD.01.231 при Институте Информационных Технологий Национальной Академии Наук Азербайджана по адресу: AZ1141, г. Баку, ул. Б.Вахабзаде, 9.

С диссертацией можно ознакомиться в библиотеке Института Информационных Технологий НАН Азербайджана.

Автореферат разослан «12» мая 2015 г.

Ученый секретарь
диссертационного совета FD.01.231,
доктор философии по технике

Р.Г.ШЫХАЛИЕВ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Обеспечение информационной безопасности является наиболее сложной и актуальной задачей в современном информационном обществе, и эта задача является одним из приоритетных направлений государственной политики Республики Азербайджан. Эта политика отражена в концепциях, стратегиях, указах, государственных программах и других официальных документах, утвержденных Президентом Республики Азербайджан г-ном И.Алиевым. В этих документах, наряду с другими основополагающими задачами, ставится задача разработки и применения новых инновационных технологий в области информационной безопасности.

Биометрические технологии были определены в качестве одной из основных задач усовершенствования национальной системы безопасности страны согласно «Государственной программе по созданию системы биометрической идентификации в Азербайджанской Республике на 2007-2012 годы».

Отдельным пунктом была выделена задача создания системы идентификации и верификации по голосу в Азербайджанской Республике, выполнение которой было поручено Институту Информационных Технологий НАНА.

Системы идентификации и верификации востребованы во многих практических системах: системы аутентификации и авторизации электронных услуг э-правительства, в системах э-коммерции и э-банкинга, дистанционного обучения и т.д. Системы распознавания по голосу также используются правоохранительными органами при фоноскопической экспертизе.

Системы идентификации по голосу имеют ряд важных практических преимуществ по сравнению с другими широко распространенными биометрическими системами, такими как системы распознавания по отпечаткам пальцев и радужной оболочке глаз.

Национальный институт стандартов и технологий (National Institute of Standards and Technology, NIST) организывает конкурсы по оценке подходов для тексто-независимого распознавания диктора NIST SRE. Согласно его оценкам точность систем распознавания личности по голосу в значительной степени отстает от точности других систем распознавания.

Исследования существующих методов и алгоритмов показывают, что методы извлечения признаков речевых сигналов для распознавания личности по голосу в значительной степени чувствительны к изменениям состояния диктора, окружающей среде, условиям записи и идентификации. В реальных приложениях всегда существуют различия между этапами регистрации пользователя и условиями его идентификации. Например, регистрация может проводиться в офисных условиях, а идентификация - в условиях транспорта, по мобильным устройствам.

Исходя из вышеизложенного, настоящая диссертационная работа, посвященная разработке методов и алгоритмов извлечения признаков речевого сигнала и методов компенсации канальных искажений для синтеза систем идентификации личности по голосу, является весьма актуальной.

Можно отметить следующих ученых, которые проводят очень продуктивные исследования в области распознавания диктора: Douglas A. Reynolds, Jean-François Bonastre, Tomi Kinnunen, Luciana Ferrer, Nicolas Scheffer, Lukas Burget и др.

В русскоязычной научной литературе имеются работы в области распознавания диктора начиная с 70-х годов. Из авторов последних разработок можно выделить Симончика К.К., Первушина Е.А., Голубинского А.Н., Сорокина В.Н. и др.

Цель и задачи работы. Целью диссертационной работы является разработка методов и алгоритмов для синтеза систем биометрической идентификации личности по голосу с высокими показателями распознавания, устойчивых к изменениям канал/сессия и искажениям канала.

Для достижения поставленной цели в работе решаются следующие задачи:

- анализ и исследование существующих методов и алгоритмов биометрической идентификации личности по голосу;
- разработка методов и алгоритмов извлечения признаков речевых сигналов для распознавания личности по голосу;
- разработка метода компенсации канальных искажений для распознавания личности по голосу;

- разработка методов распознавания личности по голосу путем объединения решений классификаторов признаков различных уровней;
- выбор речевого корпуса для экспериментальных исследований и оценка разработанных методов и алгоритмов для распознавания личности по голосу.

Методы исследования. При решении поставленных в работе задач использовались теория информационной безопасности, машинное обучение, теория распознавания образов, теория нечетких чисел, теория вероятностей.

Основные положения, выносимые на защиту:

- метод и алгоритмы извлечения спектральных признаков речевых сигналов на основе эмпирического вейвлет преобразования;
- метод извлечения периода основного тона, робастного к шумам;
- метод компенсации канальных искажений на основе модифицированного вероятностного линейного дискриминантного анализа;
- метод и алгоритм слияния решений классификаторов на основе нечетких интегралов.

Научная новизна работы состоит в разработке методов и алгоритмов распознавания личности по голосу, в том числе:

- разработан метод извлечения спектральных признаков речевых сигналов на основе эмпирического вейвлет преобразования и разработаны соответствующие алгоритмы;
- разработан метод извлечения периода основного тона;
- разработан метод компенсации канальных искажений с применением модифицированного вероятностного линейного дискриминантного анализа;
- разработаны метод нечеткого слияния классификаторов и алгоритм для распознавания личности по голосу.

Практическая ценность работы заключается в том, что разработанные методы и алгоритмы могут быть использованы в биометрических системах распознавания личности по голосу для повышения точности распознавания. Для синтеза системы

распознавания личности по голосу на основе предложенных методов и алгоритмов разработан прототип системы. Кроме того, для оценки систем распознавания личности по голосу была собрана речевая база для азербайджанского языка AZ-SRData (AZerbaijani language Speaker Recognition Data).

Полученные результаты диссертационной работы могут быть применены в области информационной безопасности к системам биометрической аутентификации.

Реализация и внедрение результатов работы. Разработанный прототип системы распознавания личности по голосу был внедрен в научную компьютерную сеть AzScienceNet и получен соответствующий акт.

Апробация работы. Основные научно-теоретические и практические результаты диссертации докладывались и обсуждались на: семинарах Института Информационных Технологий НАНА; III Международной конференции «Problems of Cybernetics and Informatics», Баку, 6-8 сентября, 2010 г.; V Международной конференции «Application of Information and Communication Technologies», Баку, 12-14 октября, 2011 г.; IV Международной конференции «Problems of Cybernetics and Informatics», Баку, 12-14 сентября, 2012 г.; научном семинаре, посвященном Международному дню защиты информации, Баку, 30 ноября, 2012 г.; Azərbaycan xalqının ümummilli lideri Heydər Əliyevin 90 illik yubileyinə həsr olunmuş İnformasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransı, Баку, 17-18 мая, 2013 г.; VII Международной конференции «Application of Information and Communication Technologies», Баку, 23-25 октября, 2013 г.

Соискатель принимал участие в конкурсах грантов № EIF-2011-1(3)-82/08/1, № EIF-RITN-MQM-2/ИКТ-2-2013-7(13)-29/18/1, проводимых Фондом Развития Науки при Президенте Азербайджанской Республики.

Публикации. По теме диссертации опубликовано 16 научных работ, из них 10 статей и 6 тезисов.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы из 204 наименований и трех приложений. Основное содержание работы изложено на 119 страницах, включая 32 рисунка и 8 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, сформулированы цель и задачи, определены научная новизна и практическая ценность полученных результатов.

В первой главе исследовано современное состояние проблем распознавания личности по голосу, также известное как верификация и идентификация диктора. Описаны основные компоненты системы биометрической идентификации личности по голосу, и проведен анализ методов извлечения признаков речевого сигнала, используемых для решения задачи распознавания личности по голосу.

Проведенный анализ научно-исследовательских работ показывает, что крупные исследовательские и коммерческие организации проявляют интерес к разработкам систем распознавания диктора, о чем свидетельствует большое число публикаций в научных журналах и докладов научных конференций посвященных этой теме, а также конкурсы по системам распознавания диктора NIST SRE.

В главе было отмечено, что в настоящее время разработано большое количество методов распознавания диктора, которые показывают обнадеживающие результаты распознавания личности по голосу. Однако вопрос о создании систем идентификации диктора, обладающих высокой точностью, приемлемой для важных практических приложений, остается открытым.

Анализ существующих систем распознавания диктора показывает, что актуальными являются следующие проблемы: разработка методов и алгоритмов извлечения спектральных и просодических признаков речевых сигналов, учитывающих нестационарность сигналов; разработка методов компенсации канальных искажений и методов объединения решений классификаторов, повышающих точность распознавания.

Во второй главе предложены методы извлечения различных речевых признаков для распознавания личности по голосу. Для извлечения спектральных признаков речевых сигналов предложен метод эмпирического вейвлет преобразования (Empirical Wavelet Transform, EWT). В качестве признаков рассматриваются мгновенная частота и мгновенная амплитуда. Для их выделения применяется метод на основе оператора выделения энергии Тигера-Кайзера (Teager-Kaiser Energy Operator, ТКЕО). Данный подход адаптируется к реальным речевым сигналам, учитывая их нестационарность и нелинейность.

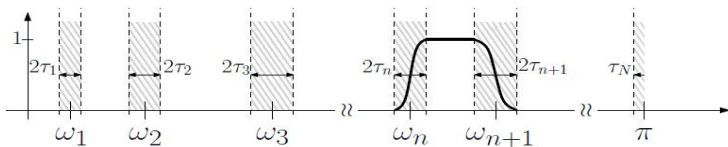


Рис. 1. Разбивка оси Фурье

Метод на основе адаптивных вейвлетов эквивалентен построению набора полосовых фильтров. EWT рассматривает спектр Фурье на отрезке $[0, \pi]$, который делится на N смежных сегментов (рис. 1). Границы между сегментами обозначаются через ω_n (где $\omega_0 = 0$ и $\omega_N = \pi$). Каждый сегмент обозначен через $\Lambda_n = [\omega_{n-1}, \omega_n]$. Вокруг ω_n определяется переходная фаза T_n шириной $2\tau_n$.

Внутренняя модовая функция (Intrinsic Mode Function, IMF) f_k определяется следующим образом:

$$f_0(t) = W_f^\varepsilon(0, t) * \phi_1(t), \quad (1)$$

$$f_k(t) = W_f^\varepsilon(k, t) * \psi_k(t), \quad (2)$$

где детализирующие коэффициенты представляются в виде скалярного произведения с эмпирическими вейвлетами:

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \psi_n(\tau - t) d\tau, \quad (3)$$

а аппроксимирующие коэффициенты в виде скалярного произведения с масштабирующей функцией:

$$W_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \phi_1(\tau - t) d\tau, \quad (4)$$

Алгоритм извлечения речевых признаков заключается в следующем:

- 1) **Выделение вокализованных/невокализованных участков.** Вектор признаков извлекается только из вокализованных участков.
- 2) **Получение IMF компонент на основе EWT:**
 - a. Применение быстрого преобразования Фурье
 - b. Вычисление локального максимума спектра сигнала на отрезке $[0, \pi]$ и нахождение множества $\{\omega_n\}$
 - c. Выбор параметра для определения границ между сегментами согласно $\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right)$

- d. Построение банка фильтров EWT
 - e. Фильтрация сигнала для получения каждого компонента IMF.
- 3) **Применение ТКЕО.** Нахождение мгновенной частоты и мгновенной амплитуды на основе оператора ТКЕО.
- 4) **Объединение векторов признаков.** Полученные вектора мгновенных частот и мгновенных амплитуд далее объединяются для получения нового вектора признаков.

Оценка периода основного тона является важным компонентом в различных приложениях обработки речи, таких как распознавание диктора, синтез речи и т.д. Несмотря на то, что существуют многочисленные методы оценки частоты основного тона речевых сигналов, они сильно подвержены влиянию шума.

Алгоритмы YIN, YAAPT, BaNa, PEFAC, AUTOC, AMDF, SWIPE', кепстральный метод и MBSC были оценены при различных условиях шума. BaNa оказался наиболее робастным к шуму в диапазоне от -5 до 15 дБ. PEFAC также показал высокую производительность.

В главе также предложен метод извлечения периода основного тона на основе EWT. Суть подхода: к сигналу применяется EWT преобразование, которое разлагает сегменты сигнала на серию IMF функций для дальнейшего извлечения мгновенной частоты. Из всего набора IMF выбирается та, чей период ближе к исходному сигналу. Была рассмотрена первая IMF, так как она содержит информацию о периоде основного тона и хорошо отображает форму сигнала.

Проведено сравнение методов BaNa, PEFAC и предложенного подхода для оценки периода основного тона. Были получены результаты, представленные на графиках (Рис. 2) для баз данных CSTR и Keele.

Предложенный метод является более робастным по сравнению с другими методами и лучше работает при высоких соотношениях сигнал-шум (Signal-to-Noise Ratio, SNR).

Эффективность предложенного подхода может быть повышена при увеличении числа признаков. Однако это значительно повлияет на скорость работы алгоритма.

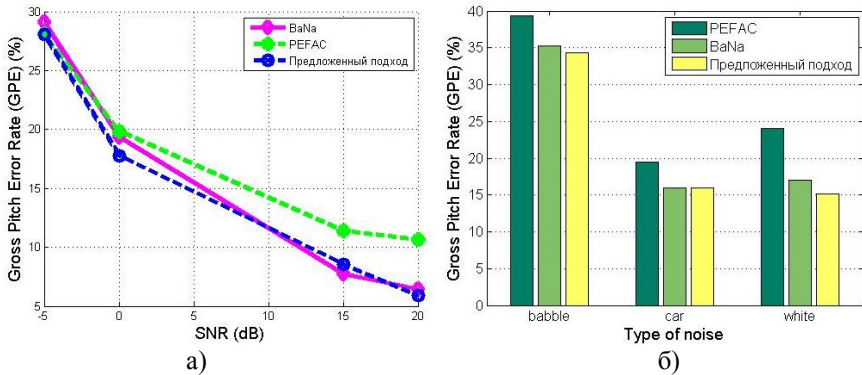


Рис. 2. Сравнение производительности методов оценки периода основного тона: (а) GPE гибридных методов (CSTR); (б) Сравнение гибридных методов при различных типах шума (Keele)

В третьей главе проанализированы различные методы компенсации и выравнивания канала, которые могут быть использованы в системах верификации и идентификации диктора, когда речь передается через каналы связи. Искажение канала является серьезной проблемой для этих типов систем, так как даже незначительное количество искажений в речи может вызвать изменение уникальных особенностей в голосе человека и уничтожение необходимой информации. Был предложен модифицированный вероятностный линейный дискриминантный анализ (Probabilistic Linear Discriminant Analysis, PLDA) для компенсации канальных искажений.

Суть предлагаемого подхода заключается в следующем: PLDA является производной моделью вектора x_{ij} размерности D_x , который может быть представлен в виде:

$$x_{ij} = \mu + Fh_i + Gw_j + \varepsilon_{ij}. \quad (5)$$

Эта модель состоит из двух частей: компонент $\mu + Fh_i$, который зависит только от диктора и компонент шума $Gw_j + \varepsilon_{ij}$. Остаточный член в (5) – гауссиан с нулевым математическим ожиданием и ковариационной матрицей Σ .

PLDA включает два этапа: обучение и вычисление правдоподобия.

Сложность на этапе обучения состоит в вычислении сложной обратной матрицы $(I + \tilde{A}^T \tilde{\Sigma}^{-1} \tilde{A})^{-1}$ на каждой итерации при нахождении моментов 1-го и 2-го порядка, т.к. она имеет большую размерность.

При вычислении правдоподобия согласно (6) сложность состоит в нахождении определителя и инверсии обратной матрицы $(\tilde{\Sigma} + \tilde{A}\tilde{A}^T)$ в последних двух членах уравнения.

$$\begin{aligned} \ln[\Pr(\tilde{x}_i)] = & -\frac{D_x J}{2} \ln[2\pi] - \frac{1}{2} \ln[\det(\tilde{\Sigma} + \tilde{A}\tilde{A}^T)] - \\ & -\frac{1}{2} \tilde{x}_i^T (\tilde{\Sigma} + \tilde{A}\tilde{A}^T)^{-1} \tilde{x}_i. \end{aligned} \quad (6)$$

Поэтому предлагается модифицированный подход на основе PLDA. Целью является достижение возможности обучения PLDA так, чтобы этот подход был масштабируем по отношению к количеству обучающих выборок. Для этого производится замена инверсии полной матрицы набором инверсий блочной матрицы, которые требуют меньших вычислений.

Тогда оценка моментов 1-го порядка примет вид

$$E[h_i] = M[\sum_j F^T \tilde{\Sigma}^{-1} \tilde{x}_{ij} - \sum_j F^T \Lambda \tilde{\Sigma}^{-1} \tilde{x}_{ij}], \quad (7)$$

$$E[w_{ij}] = \Lambda(\tilde{x}_{ij} - FE[h_i]), \quad (8)$$

где $\tilde{x}_{ij} = x_{ij} - \mu$, $\Lambda = BG^T \tilde{\Sigma}^{-1}$, $M = (I + JF^T \tilde{\Sigma}^{-1} (F - G\Lambda F))^{-1}$ и $B = (I + G^T \tilde{\Sigma}^{-1} G)^{-1}$.

Оценка момента 2-го порядка при этом

$$E[y_{ij} y_{ij}^T] = \begin{bmatrix} M & -\Lambda F^T M \\ -\Lambda F M & B + \Lambda F^T M F \Lambda^{-1} \end{bmatrix} + E[y_{ij}] E[y_{ij}]^T. \quad (9)$$

В результате преобразований было получено следующее выражение для нахождения правдоподобия, что ускоряет вычислительный процесс:

$$\begin{aligned} \ln[\Pr(\tilde{x}_{ij})] = & -\frac{D_x J}{2} \ln[2\pi] - \frac{1}{2} \ln[\det(\tilde{\Sigma})^J \det(B^{-1})^J \det(M^{-1})] - \\ & -\frac{1}{2} \sum_{j=1}^J \tilde{x}_{ij}^T V \tilde{x}_{ij} + \frac{1}{2} \left(\sum_{j=1}^J \tilde{x}_{ij}^T V^T F \right) M \left(\sum_{j=1}^J F^T V \tilde{x}_{ij} \right). \end{aligned} \quad (10)$$

Эксперименты проводились на базе данных MIT Mobile Device Speaker Verification Database (MIT MDSVD).

На рис. 3 показаны параметрические кривые ошибок, полученные при трех различных условиях.

Равный уровень ошибок (Equal Error Rate, EER) составил 25,3%. На первом графике (рис. 3 (а)) показан результат применения модифицированного PLDA подхода. На рис. 3 (б) можно увидеть кривые, полученные в результате тестирования в условиях офиса, в коридоре и на перекрестке улиц на основе модифицированного PLDA.

В результате экспериментов EER был снижен, что показывает эффективность предложенного подхода.

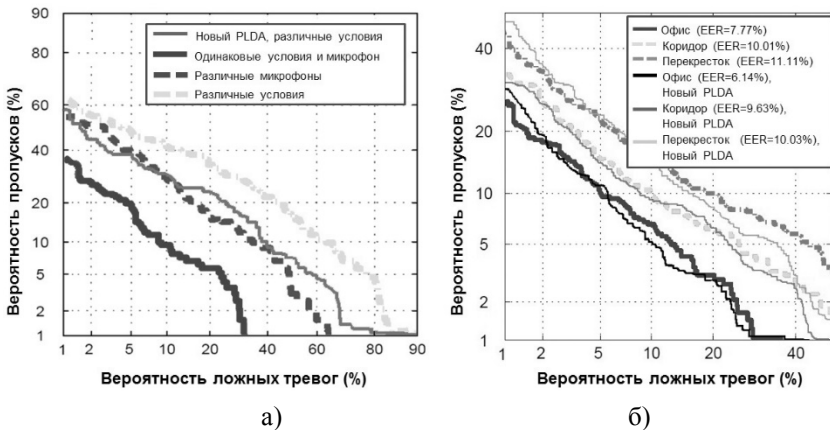


Рис. 3. Результаты проведенных экспериментов

Важным этапом систем распознавания диктора (говорящего) является классификация.

В главе была рассмотрена задача слияния решений классификаторов для повышения точности работы систем распознавания говорящего.

Был предложен подход на основе нечетких интегралов для объединения значений соответствия, полученных в результате классификации мгновенной частоты (y_1), мгновенной амплитуды (y_2) и частоты основного тона (y_3).

Подход для случая m речевых признаков включает следующие основные этапы: получение значений соответствия для каждого классификатора $y_i, i=1, m$. Далее определяются нечеткие меры для каждого признака речевого сигнала $g(A_i)$, где

$A_i = \{y_i, y_{i+1}, \dots, y_m\}$, а $g(A_i) = g_i + g(A_{i-1}) + \lambda g_i g(A_{i-1})$ при $1 \leq i \leq m$. Вычисляется постоянная λ согласно

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g_i). \quad (11)$$

Находятся нечеткие меры \overline{m} для всевозможных комбинаций речевых признаков $g(A_i, A_j)$, $i, j = \overline{1, m}$. Затем вычисляется нечеткий интеграл

$$\int_A f \circ g(\cdot) = \max_{i=1}^m [\min(f_i, g_i)], \quad (12)$$

где функция $f: Y \rightarrow [0, 1]$.

Для проведения экспериментов были взяты речевые образцы из базы данных ТИМТ для 60 дикторов. В качестве классификатора был рассмотрен метод машин опорных векторов (Support Vector Machine, SVM) с радиальной базисной функцией ядра.

Таблица 1.
Результаты объединения значений классификаторов

Диктор	Правило максимума	Метод взвешенной суммы	Метод логарифмического объединения	Предложенный подход
falk0	0,5835	0,4673	0,4622	0,5815
fcmg0	0,5758	0,4274	0,4197	0,5699
flac0	0,6796	0,5905	0,5879	0,7256
makr0	0,5896	0,4464	0,4388	0,5908
mdhs0	0,5896	0,4377	0,4300	0,5908

Эксперименты проводились с применением четырех методов объединения: правила максимума, взвешенной суммы, логарифмического объединения и предлагаемого автором подхода на основе нечетких интегралов (Таблица 1). Результаты показали, что предложенный подход лучше по сравнению с другими методами.

В четвертой главе приводится описание прототипа системы распознавания личности по голосу, разработанного на основе предложенных методов и алгоритмов. Приведено также описание и сравнение основных речевых баз данных, которые используются для тестирования систем распознавания диктора. Описана речевая база данных азербайджанского языка для тестирования систем распознавания диктора, собранная автором.

Таблица 2.
Результаты экспериментов для различных баз данных

Речевая база данных	FRR	GAR	EER
TIMIT	6.56%	93.44%	3.97%
MIT MDSVD	7.27%	92.73%	4.15%
AZ-SRData	8.17%	91.83%	5.02%
NIST SRE (2010)	7.97%	92.03%	3.76%

Разработанная система была протестирована на речевых базах данных TIMIT, MIT MDSVD, AZ-SRData. В качестве эталонных значений были рассмотрены результаты конкурса NIST SRE (2010) (Таблица 2).

Таблица 3.
Результаты экспериментов для базы данных TIMIT

Метод	FRR	GAR	EER
GMM-SVM	7.87%	92.13%	6.31%
GMM-UBM	7.00%	93.00%	5.72%
VQ-UBM	9.10%	89.90%	10.17%
PLDA-FI	6.56%	93.44%	3.97%

Метод извлечения периода основного тона был протестирован в зашумленной среде (речевые корпусы Keele, CSTR).

Таблица 4.
Результаты экспериментов для базы данных NIST SRE (2010)

Метод	FRR	GAR	EER
GMM-SVM	10.83%	89.17%	6.84%
GMM-UBM	10.16%	89.84%	7.63%
VQ-UBM	11.37%	88.63%	7.56%
PLDA-FI	7.97%	92.03%	3.76%

В таблице 2 приводится сравнение полученных вероятности пропуска "своего" (Genuine Acceptance Rate, GAR), вероятности ложного отказа (False Reject Rate, FAR) и EER разработанной системы для предложенных подходов на различных базах данных в сравнении с NIST SRE 2010.

Результаты для собранной речевой базы данных AZ-SRData лишь незначительно уступают в точности распознавания речевым базам данных TIMIT и MIT MDSVD.

Было проведено сравнение различных методов (GMM-SVM, GMM-UBM, VQ-UBM) с предложенным подходом (PLDA-FI) на базах данных TIMIT (Таблица 3) и NIST SRE 2010 (Таблица 4).

Оценка разработанных методов и алгоритмов была реализована с использованием среды Visual Studio 2010 и языка программирования C++, а также Matlab 2011b.

В заключении отражены наиболее важные результаты диссертационной работы, а также сформулированы основные выводы, следующие из предложенных методов и алгоритмов и полученных результатов.

В приложениях приведены сравнение биометрических технологий, описание гауссовых смешанных моделей и программная реализация некоторых функций на Matlab 2011b.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Исследовано современное состояние проблем распознавания личности по голосу. Проведен анализ методов извлечения признаков речевого сигнала и исследованы методы оценки периода основного тона, используемые для решения задач распознавания диктора.
2. Разработан алгоритм извлечения спектральных признаков речевых сигналов на основе эмпирического вейвлет преобразования.
3. Для снижения помех в речевом сигнале разработан метод извлечения периода основного тона.
4. Разработан метод компенсации канальных искажений с применением модифицированного вероятностного линейного дискриминантного анализа.
5. Разработан метод нечеткого слияния классификаторов и соответствующий алгоритм для распознавания личности по голосу.
6. Проведены экспериментальные исследования разработанных методов и алгоритмов в зашумленной среде (Keele, CSTR), мобильной среде (MIT MDSVD) и на речевой базе данных для азербайджанского языка (AZ-SRData), собранной автором.

Результаты, полученные при экспериментальных исследованиях, показали преимущества разработанных методов и доказывают перспективы выбранных подходов.

Основные части содержания диссертационной работы изложены в следующих публикациях:

1. Imamverdiyev Y.N., Sukhostat L.V. Decision making in multibiometric systems based on fuzzy integrals / Proc. of Int. Conf. "Problems of Cybernetics and Informatics", 2010, p. 87-90.
2. Имамвердиев Я.Н., Сухостат Л.В. Речевые базы данных для систем распознавания диктора // Вопросы защиты информации, 2011, № 4, с. 27-32.
3. Imamverdiyev Y.N., Sukhostat L.V. SVM based recognition of Azerbaijani vowels / Proc. of Int. Conf. on Application of Information and Communication Technologies, 2011, Baku, p. 716-719.
4. Sukhostat L.V., Imamverdiyev Y.N. New feature vector extraction method for speaker recognition / Proc. of Int. Conf. "Problems of Cybernetics and Informatics", 2012, p. 99-102.
5. Имамвердиев Я.Н., Сухостат Л.В. О выборе признаков в задачах распознавания личности по голосу // Телекоммуникации, 2012, № 4, с. 28-32.
6. Имамвердиев Я.Н., Сухостат Л.В. Исследование методов распознавания личности по голосу // Телекоммуникации, 2012, № 10, с. 19-23.
7. Сухостат Л.В. Разработка методов и алгоритмов для синтеза систем биометрической идентификации личности по голосу / Материалы научного семинара. Баку: Институт информационных технологий НАНА, 2012, с. 29-30.
8. Сухостат Л.В. Разработка прототипа системы распознавания личности по голосу / Azərbaycan xalqının ümummilli lideri Heydər Əliyevin 90 illik yubileyinə həsr olunmuş informasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransın materialları. Bakı: AMEA İnformasiya Texnologiyaları İnstitutu, 2013, s. 151-154.
9. Имамвердиев Я.Н., Сухостат Л.В. Метод оптимизации показателя распознавания в мультибиометрических системах // Информационные технологии, 2013, № 1, с. 9-14.
10. Имамвердиев Я.Н., Сухостат Л.В. AZ-SRDATA – речевая база данных для азербайджанского языка // İnformasiya texnologiyaları problemləri, 2013, v. 7, № 1, с. 67-73.
11. Imamverdiyev Y.N., Sukhostat L.V. An advanced channel compensation method for speaker recognition / Proc. of Int. Conf. on Application of Information and Communication Technologies, 2013, p. 128-131.

12. Имамвердиев Я.Н., Сухостат Л.В. Разработка робастного метода извлечения речевых признаков на основе эмпирического вейвлет преобразования // Информационные технологии, 2015, № 1, с. 19-23.
13. Имамвердиев Я.Н., Сухостат Л.В., Подход для слияния решений классификаторов для задачи распознавания диктора // Информатика (ОИПИ НАН Беларуси 50 лет), 2015, № 45, с. 17-25.
14. Sukhostat L.V., Imamverdiyev Y.N. A comparative analysis of pitch detection methods under the influence of different noise conditions // Journal of Voice, 2015, (ISI Impact Factor: 0,944), <http://www.sciencedirect.com/science/article/pii/S0892199714002021>.
15. Сухостат Л.В. Об одном подходе для нахождения периода основного тона речевых сигналов с применением адаптивных вейвлетов // Informasiya texnologiyaları problemləri, 2015, № 1, с. 33-41.
16. Имамвердиев Я.Н., Сухостат Л.В. Компенсация канальных искажений на основе вероятностного линейного дискриминантного анализа // Телекоммуникации, 2015 (в печати).

Роль соискателя в трудах, опубликованных в соавторстве:

- [1] – предложен метод нечеткого слияния классификаторов;
- [2] – проведен анализ основных речевых баз данных;
- [3] – предложен метод распознавания гласных азербайджанского языка;
- [4] – предложен новый метод извлечения признаков для задачи распознавания личности по голосу;
- [5] – проведен анализ методов и алгоритмов извлечения признаков речевых сигналов;
- [6] – исследованы методы распознавания личности по голосу;
- [9] – предложен метод оптимизации показателей распознавания;
- [10] – собрана речевая база данных для азербайджанского языка;
- [11] – предложен метод компенсации канальных искажений;
- [12] – предложен метод извлечения речевых признаков на основе эмпирического вейвлет преобразования;
- [13] – предложен метод и алгоритм слияния решений классификаторов для задачи распознавания говорящего;
- [14] – проведен сравнительный анализ методов извлечения периода основного тона речевого сигнала;
- [16] – предложен метод компенсации канальных искажений с применением модифицированного вероятностного линейного дискриминантного анализа.

Suxostat Lyudmila Valentinovna

Səsə görə şəxsin biometrik identifikasiyası sistemlərinin sintezi üçün metod və alqoritmlərin işlənməsi

Xülasə

Dissertasiya işi səsə görə şəxsin biometrik identifikasiyası sistemlərinin sintezi üçün metod və alqoritmlərin işlənməsinə həsr edilmişdir. Səsə görə şəxsin tanınması üçün mövcud yanaşmalar tədqiq edilmiş və perspektiv tədqiqat problemləri müəyyən edilmişdir.

Dissertasiya çərçivəsində nitq signalı əlamətlərinin çıxarılması metodları öyrənilmiş və diktorun tanınması məsələlərinin həlli üçün istifadə edilən əsas ton periodunun qiymətləndirilməsi metodları tədqiq edilmişdir. Tədqiqatın nəticələrinə əsasən nitq signallarından əlamətlərin çıxarılması üçün empirik veyvlet çevirməsi əsasında metod təklif edilmiş və onun reallaşdırılması üçün alqoritmlər işlənmişdir. Səsin əsas tonunun periodunun hesablanması metodu təklif edilmişdir.

Kanal təhriflərinin kompensasiyası üçün ehtimallı xətti diskriminant analizi metodunun modifikasiya olunmuş variantı işlənmişdir. Səsə görə şəxsi tanıma klassifikatorlarının qeyri-səlis aqreqasiyası metodu və müvafiq alqoritm işlənmişdir.

Səsə görə tanıma üçün nitq verilənləri bazaları təhlil edilmişdir. Azərbaycan dili üçün nitq bazası üçün məlumatlar toplanmışdır. Təklif edilmiş metod və alqoritmlərin eksperimental tədqiqatları aparılmışdır. Metod və alqoritmlər Visual Studio 2010 və Matlab 2011b istifadə edilməklə C++ proqramlaşdırma dilində reallaşdırılmışdır.

Sukhostat Lyudmila Valentinovna

Development of methods and algorithms for synthesis of speaker identification systems

Summary

The thesis is dedicated to the development of methods and algorithms for synthesis of speaker identification systems. Existing approaches for speaker recognition has been investigated and perspective research problems have been identified.

During the research the methods of speech feature extraction and investigated methods of pitch period estimation for speaker recognition have been developed. Based on the results of the study, an algorithm for spectral feature extraction based on the empirical wavelet transformation has been proposed. New method for pitch period extraction has been developed.

A modified method for channel distortion compensation that can be used in systems of speaker identification and verification based on probabilistic linear discriminant analysis has been developed. A method for fuzzy classifiers fusion and the corresponding algorithm for speaker recognition have been developed.

Popular speech databases have been analysed. Speech database for Azerbaijani language has been collected and, based on the developed methods and algorithms, experiments have been performed. The evaluation has been implemented using the C++ programming language with Visual Studio 2010 and Matlab 2011b.

Çapa imzalanıb: 08.05.2015. Tirağı 100 nüsxə
AMEA İnformasiya Texnologiyaları İnstitutunun
“İnformasiya Texnologiyaları” nəşriyyatı

**AZƏRBAYCAN MİLLİ ELMLƏR AKADEMİYASI
İNFORMASIYA TEXNOLOGİYALARI İNSTİTUTU**

Əlyazması hüququnda

SUXOSTAT LYUDMİLA VALENTİNOVNA

**SƏSƏ GÖRƏ ŞƏXSİN BIOMETRİK İDENTİFİKASIYASI
SİSTEMLƏRİNİN SİNTEZİ ÜÇÜN METOD VƏ
ALQORİTMLƏRİN İŞLƏNMƏSİ**

3338.01 – Sistemli analiz, idarəetmə və informasiyanın işlənməsi

Texnika üzrə fəlsəfə doktoru elmi dərəcəsi
almaq üçün təqdim edilmiş dissertasiyanın

A V T O R E F E R A T I

BAKI – 2015