

УДК 628.440.22

М.В. Привалов, О.В. Кулибаба

Донецкий национальный технический университет, г. Донецк
кафедра автоматизированных систем управления
E-mail: olechka@mail.ua

ВЫБОР ПРИЗНАКОВ ДЛЯ АУТЕНТИФИКАЦИИ ПО ГОЛОСУ В КОМПЬЮТЕРИЗИРОВАННОЙ СИСТЕМЕ КОНТРОЛЯ ДОСТУПА

Аннотация

Привалов М.В., Кулибаба О.В. Выбор признаков для аутентификации по голосу в компьютеризированной системе контроля доступа. Рассматриваются преимущества системы аутентификации по голосу. Анализируется использование акустических признаков дикторов в различных системах. Предлагается выбор оптимального набора признаков.

Ключевые слова: аутентификация по голосу, акустические признаки, кластеризация, ошибки первого и второго рода.

Актуальность.

Информация в наше время – самый дорогой и востребованный товар. Она дает власть над людскими массами или позволяет обогнать конкурентов. На защиту информации мобилизованы самые высокие технологии. Первейшая их задача – не допустить к ней постороннего человека, то есть надежно распознать, кто «свой», а кто нет. Но так, если карту доступа элементарно можно украсть, а PIN-коды не особенно удобны, их приходится запоминать и каждый раз вводить. Очевидный выход – использовать биометрическую аутентификацию. Среди различных биометрических систем аутентификация по голосу демонстрирует очевидные преимущества:

- привычный для человека способ аутентификации;
- голос не отчуждает от человека;
- для аутентификации по голосу не требуются дорогостоящие считыватели биометрической информации.

Согласно международной конференции, посвященной «голосовой биометрии», которая ежегодно проводится в США, аутентификация по голосу стремительно развивается и с каждым годом пользуется все большим спросом [1]. Однако до сих пор нерешенным вопросом остается выбор оптимального набора признаков, которые бы минимизировали ошибки первого и второго рода.

Анализ последних исследований и публикации по рассматриваемому вопросу.

На сегодняшний день созданы десятки различных систем аутентификации по голосу, имеющих различные параметры и требования к процессу идентификации в зависимости от конкретных задач [2-4]. К сожалению, на сегодняшний день разработанные программы чаще всего применяются как дополнительные средства проверки подлинности там, где необходимо обеспечить высокую степень надежности систем аутентификации. Поэтому продолжаются работы по правильному выбору и обоснованию признаков, которые затем определяют принцип построения систем аутентификации.

В настоящий момент существует достаточно много различных методов расчета признаков, характеризующих личность диктора. Для вычисления признаков, связанных с физиологическими особенностями речевого тракта, наиболее часто используются методы спек-

трально-временного анализа. Такие методы анализа речевого сигнала адекватны природному механизму восприятия речи, что делает понятной тенденцию многих исследователей искать индивидуальные особенности в мгновенных спектральных распределениях отдельных фонем и в распределениях текущего спектра. В основе таких методов лежит классический Фурье-анализ [5] или параметрический авторегрессионный анализ (линейное предсказание как частный случай) [6]. Тесно связан со спектральным представлением речевого сигнала довольно часто применяемый в последнее время гомоморфный метод [7]. Что касается параметров сигнала возбуждения, то они могут быть рассчитаны одним из широко известных методов выделения частоты основного тона[8], но они обладают невысокой точностью и сильно зависят от формантной структуры самого сигнала. Существуют также методы определения собственно формы артикуляторного тракта[9], но они достаточно громоздки и обладают малой точностью

Постановка задач исследования.

Для построения компьютеризированной системы контроля доступа с использованием аутентификации по голосу необходимо решить следующие задачи:

- анализ уникальных индивидуальных признаков, характеризующих личность говорящего;
- обоснование и выбор этих признаков.

Решение задачи.

Аутентификация диктора - способ проверки подлинности, позволяющий достоверно убедиться в том, что субъект действительно является тем, за кого он себя выдает, на основании сравнения голоса с хранящимся в системе эталоном.

Под голосовой аутентификацией понимается следующая ситуация. Диктор произносит фразу, а компьютеризированная система распознавания характеристик голоса должна подтвердить или отвергнуть индивидуальность говорящего. В принципе произнести фразу может как истинный пользователь, так и злоумышленник. Задаваясь стоимостью возможных потерь в случае возможного несанкционированного доступа злоумышленника можно (для данной системы) рассчитать вероятность, с которой система не должна пропускать чужого.

Схема аутентификации представлена на рис. 1:

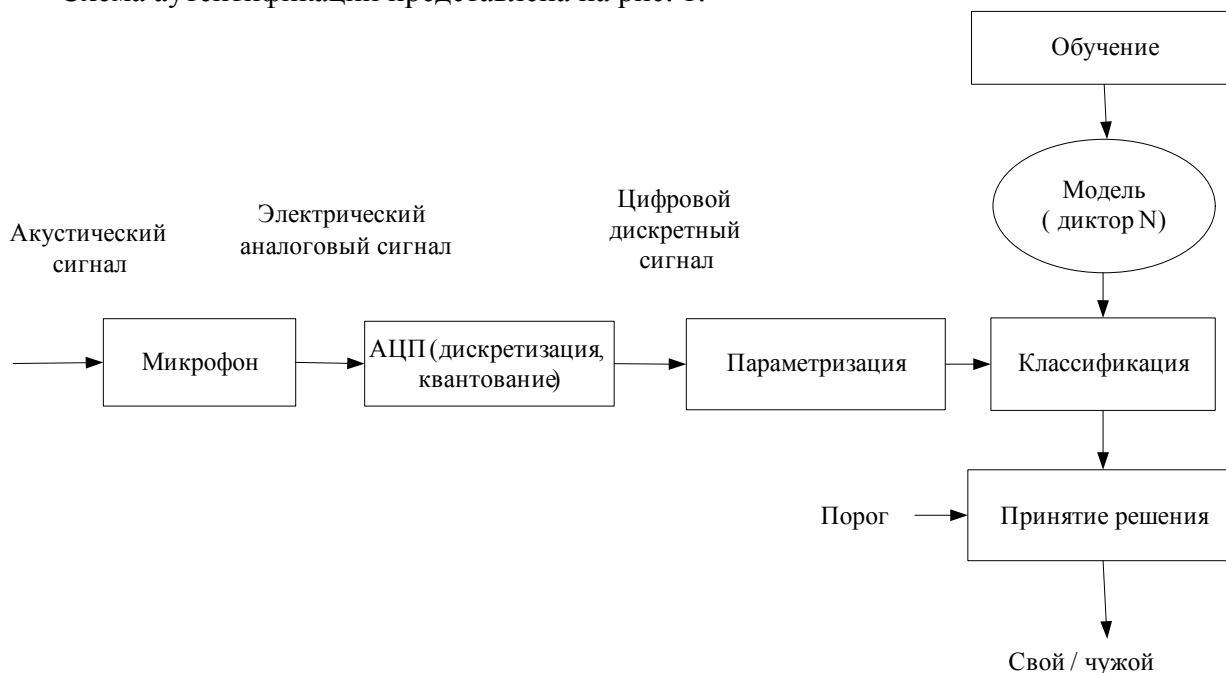


Рисунок 1 - Схема аутентификации по голосу

Задачей начального этапа аутентификации по голосу является преобразование в речевой сигнал генерируемых речевой системой человека звуков. Звук, как известно, представляет собой механические колебания, распространяющиеся в окружающей среде (средой распространения служит воздух). Давление звуковой волны воспринимается микрофоном и преобразуется им в электрический аналоговый сигнал.

Для дальнейшей обработки необходимо провести преобразование информационного образа речи из аналогового сигнала в дискретный. Эту задачу решает аналого-цифровой преобразователь (АЦП), который осуществляет дискретизацию и квантование речевого сигнала.

Важнейшим параметром систем аутентификации является коэффициент надежности - вероятность ошибок первого и второго рода:

- ошибка первого рода (FRR - False Rejection Rate) - "не узнать своего", т.е. принимается решение "чужой", хотя на самом деле субъект присутствует в списке зарегистрированных пользователей;
- ошибка второго рода (FAR - False Acceptance Rate) - "пропустить чужого", т.е. принимается решение "свой", хотя, на самом деле, субъект отсутствует в списке зарегистрированных пользователей.

Каждая данная система может перестраиваться таким образом, что ошибки одного рода могут быть уменьшены за счет увеличения ошибок другого рода (даже при сохранении всех других факторов, влияющих на вероятность ошибки: длительность и характер речевого сообщения, помехи и т.п.). Изменение соотношения ошибок первого и второго рода достигается за счет изменения порога принятия решения и выбора набора признаков.

Таким образом, важнейшим элементом успешного распознавания дикторов является выбор информативных признаков (речевых параметров), способных эффективно представлять информацию об особенностях речи конкретного диктора.

К ним предъявляются следующие требования:

- эффективность представления информации об особенностях речи конкретного диктора;
- простота измерения;
- стабильность во времени;
- частое и естественное появление в речи;
- невосприимчивость к имитации.

В качестве уникального вектора признаков можно использовать одномерный частотный вектор кепстральных коэффициентов, а также вектор, составленный из его производных. Кепстральные коэффициенты определяются в соответствии со схемой, представленной на рис. 2:

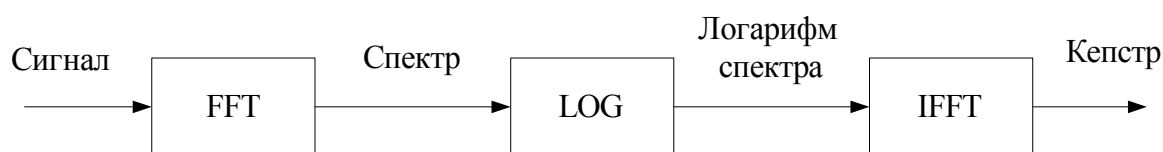


Рисунок 2 – Общая схема кепстрального анализа сигнала (FFT – блок быстрого преобразования Фурье сигнала, LOG – блок логарифмирования спектра, IFFT – блок обратного быстрого преобразования Фурье)

Линейное предсказание является одним из наиболее эффективных методов при оценке основных параметров речевого сигнала, таких как, например, период основного тона, функция площади речевого тракта и т.п. Важность метода обусловлена высокой точностью получаемых оценок и относительной простотой вычислений.

Основной принцип метода линейного предсказания состоит в том, что текущий отчет речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отчетов.

Коэффициенты предсказания при этом определяются однозначно минимизацией среднего квадрата разности между отчетами речевого сигнала и их предсказанными значениями (на конечном интервале).

Для нахождения коэффициентов предсказывающего фильтра используем автокорреляционный метод построения авторегрессионной модели [10]. Полученный фильтр может не предсказывать сигнал точно - даже в том случае, если он действительно является авторегрессионным процессом нужного порядка. Дело в том, что автокорреляционный метод неявно строит исходные данные, предполагая, что отсчеты сигнала за пределами входного вектора x равны нулю.

Нужно найти решение, дающее минимальную среднеквадратичную ошибку в системе уравнений:

$$x_a \approx b, \tag{1}$$

$$X = \begin{bmatrix} x(1) & 0 & \dots & 0 \\ x(2) & x(1) & \ddots & \vdots \\ \vdots & x(2) & \ddots & 0 \\ x(m) & \vdots & \ddots & x(1) \\ 0 & x(m) & \ddots & x(2) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & x(m) \end{bmatrix}, \quad a = \begin{bmatrix} 1 \\ a(2) \\ \vdots \\ a(p+1) \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{2}$$

где m - длина последовательности x .

Поиск решения, оптимального по критерию минимума среднеквадратичской ошибки, приводит к нормальным уравнениям:

$$x^H x_a = x_b \tag{3}$$

И системе линейных уравнений Юла-Уолкера:

$$\begin{bmatrix} r(1) & r(2)^* & \dots & r(p)^* \\ r(2) & r(1) & \dots & r(p-1)^* \\ \vdots & \ddots & \ddots & \vdots \\ r(p) & \dots & r(3) & r(1) \end{bmatrix} \begin{bmatrix} a(2) \\ a(3) \\ \vdots \\ a(p+1) \end{bmatrix} = \begin{bmatrix} -r(2) \\ -r(3) \\ \vdots \\ -r(p+1) \end{bmatrix}, \tag{4}$$

где $r = [r(1) \ r(2) \ \dots \ r(p+1)]$ - оценка автокорреляционной функции сигнала x . Уравнения Юла-Уолкера решаются за $O(p^2)$ операций с помощью алгоритма Левинсона-Дарбина [11].

Еще в качестве вектора признаков можно использовать коэффициенты отражения. Физический смысл коэффициентов отражения состоит в определении величины волны, отраженной на границе двух акустических труб.

Коэффициенты отражения рассчитываются путем преобразования вектора коэффициентов предсказывающего фильтра a в коэффициенты отражения соответствующей решетчатой структуры по следующему рекурсивному алгоритму:

$$k(n) = a_n(n) \tag{5}$$

$$a_{n-1}(m) = \frac{a_n(m) - k_n a_n(n-m)}{1 - k(n)^2}, \quad m=1,2,\dots,n-1 \tag{6}$$

Данные формулы основаны на рекурсивном алгоритме Левинсона. Для его реализации в цикле перебираются элементы вектора a , начиная с последнего и заканчивая вторым. Иногда используются также функции от коэффициентов отражения - логарифмические отношения площадей (Log-Area Ratio - LAP) [12]:

$$r_i = \log \frac{1 - k_i}{1 + k_i} \tag{7}$$

где k_i – коэффициенты отражения.

Еще одним признаком являются площади поперечных сечений акустических труб. Голосовой тракт можно представить в виде последовательности p акустических труб одинаковой длины и различных диаметров, имеющие площади поперечных сечений A_i [5]. Представление голосового тракта в виде последовательности труб изображено на рис.3.

Площади поперечных сечений A_i акустических труб вычисляется через коэффициенты отражения:

$$A_{i+1} = \frac{1 - k_i}{1 + k_i} A_i, \quad A_1 = 1, \tag{8}$$

где p – порядок линейного предсказателя,

k_i – коэффициенты отражения.

(A_2, \dots, A_{p+1}) – вектор признаков, основанный на площадях акустической трубы.

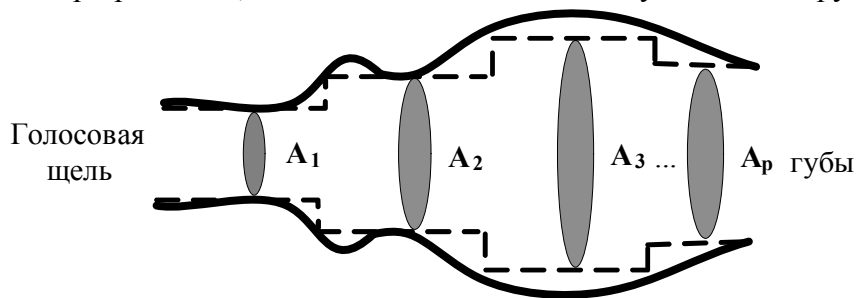


Рисунок 3– Представление голосового тракта в виде последовательности труб

Заметим, что коэффициенты отражения определяют соотношение площадей соседних секций. Таким образом, площади поперечного сечения не определяются абсолютно точно, но все-таки эти площади часто бывают сходными с конфигурацией голосового тракта, используемого человеком при речеобразовании.

Для выполнения этапа кластеризации выберем алгоритм Fuzzy c-means. Он хорошо работает с большим набором числовых данных и обладает нечеткостью при определении объектов, находящихся на границе, в кластеры [13]. Именно эти условия необходимы для решения нашей задачи.

Алгоритм Fuzzy c-means разделяет множество векторов $u_k = (k = 1, 2, \dots, K)$ на c нечетких кластеров, находит центры кластеров, минимизируя целевую функцию. При нечетком разбиении матрица принадлежности M принимает значения из интервала $[0, 1]$. Целевая функция вычисляется:

$$J(M, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{k=1}^K m_{ik}^q d_{ik}^2, \tag{9}$$

где m_{ik} - значение функции принадлежности, которое лежит в интервале от 0 до 1, c_i - центр i -го кластера, $d_{ik} = \|v_k - c_i\|$ - Евклидово расстояние между i -м центром кластера и k -ой точкой вектора и $q \in [1, \infty)$ - экспоненциальный вес.

Существует два необходимых условия для достижения минимума целевой функции:

$$c_i = \frac{\sum_{k=1}^K m_{ki}^q * u_k}{\sum_{k=1}^K m_{ki}^q}, \quad i = 1, c \tag{10}$$

$$m_{ki} = \frac{1}{\left(d_{jk}^2 * \sum_{j=1}^c \frac{1}{d_{jk}^2} \right)^{1/(q-1)}} \quad (11)$$

Экспериментальные исследования и анализ результатов.

Аутентификация по голосу с применением алгоритма кластеризации Fuzzy c-means проводилась следующим образом. Эксперимент проводился на голосовых сообщениях 18 дикторов. Исходная информация хранится в файлах .wav с форматом аудио PCM, 22050 кГц, 16 бит, моно. Для обучающей выборки были рассчитаны кепстральные коэффициенты, коэффициенты отражения, логарифмические отношения площадей (LAP) и площади поперечных сечений акустических труб по формулам (1) – (8). Показатель нечеткости в алгоритме Fuzzy c-means был выбран $q=2$, а начальные значения центров кластеров задавались как координаты центроидов.

Таблица 1 Сравнение ошибок FRR и FAR для различных признаков

Название признака	FRR	FAR
Кепстральные коэффициенты	23%	37,5%
Коэффициенты отражения	20%	35%
LAP	18%	35%
Площади поперечных сечений акустических труб	33%	55%

Анализ результатов экспериментов показал, что наиболее эффективными признаками оказались коэффициенты отражения и функции от коэффициентов отражения - логарифмические отношения площадей (LAP). При этом неплохие результаты дали и коэффициенты отражения, но они, в отличие от двух предыдущих, слишком долго обрабатываются. Что касается площади поперечных сечений акустических труб, то этот признак неплохо работает только на маленьких выборках.

Выводы.

В данной работе были рассмотрены различные уникальные признаки, характеризующие личность дикторов. Был проведен эксперимент по кластеризации голосовых сообщений различных дикторов с применением алгоритма Fuzzy c-means. В результате исследований наилучшие результаты (наименьшие ошибки первого и второго рода) были получены при использовании признака LAP.

На основании полученных экспериментальных данных выяснено, что результат является удовлетворительным, как по времени выполнения, так и по значениям FRR и FAR, и его можно использовать для построения компьютеризированной системы контроля доступа с использованием аутентификации по голосу.

Литература

1. Г.Н. Зубов, М.В.Хитров, Состояние и перспективы голосовой биометрии, 2007. [Электронный ресурс]: Режим доступа: URL:
2. <http://www.speechpro.ru/files/publications/biometry.doc>
3. Ю.Н. Хитрова, Применение речевой биометрии в системах ограничения доступа. [Электронный ресурс]: Режим доступа: URL:
4. http://www.e-expo.ru/docs/sp/cat/data/media/18_ru.pdf

5. В.А.Свириденко, П.В.Мартынович, Системы верификации и идентификации диктора от SPIRIT Corp. [Электронный ресурс]: Режим доступа: URL: http://www.dancom.ru/rus/AIA/Archive/RUII_SPIRIT_DOKLAD_R.pdf
6. Официальный сайт американской компании Nuance Technology. [Электронный ресурс]: Режим доступа: URL: www.nuance-tech.com
7. Л.Р. Рабинер, Р.В. Шафер, Цифровая обработка речевых сигналов, М.: Радио и связь, 1981. - 495с
8. Марпл-мл. С. Л. Цифровой спектральный анализ и его приложение / Пер. с англ. М.: Мир, 1990.
9. Маркел Дж., Грей А. Х. Линейное предсказание речи / Пер. с англ. М.: Связь, 1980.
10. Арлазаров В.Л., Богданов Д.С., Розанов А.О., Финкельштейн Ю.Л. Методы выделения периодов основного тона в речевом сигнале. [Электронный ресурс]: Режим доступа: URL: <http://www.cognitive.ru/innovation/sbornic/rozanov.doc>.
11. J.Shroeter, M.M.Soudhi "Techniques for Estriruation Vocal-Tract Shape from Speech Signal" IEEE Trans.SAP-2, N1, pp.133-150(1994)
12. Консультационный центр MATLAB компании Softline.[Электронный ресурс]: Режим доступа: URL: <http://matlab.exponenta.ru/>
13. Ljung, L., System Identification: Theory for the User, Prentice-Hall, 1987, pp. 278-280.
14. А.С. Алексеев, Е.Е. Федоров, Количественный анализ систем признаков и методов идентификации, Штучний інтелект, Інститут проблем искусственного интеллекта,г.Донецк, № 3, 2005.
15. [Электронный ресурс]: Режим доступа: URL: http://www.iai.dn.ua/public/JournalAI_2005_3/Razdel7/02_Alekseev_Fedorov.pdf
16. Jan Jantzen «Neurofuzzy Modelling». [Электронный ресурс]: Режим доступа: URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.390&rep=rep1&type=pdf>

Abstract

Privalov M.V., Kulibaba O.V. Choice of features for authentication by voice in a computer access control system. Advantages of system using authentication by voice are considered. Use of acoustic features of speakers in different systems are analyzed. Choice of optimal set of features is proposed.

Keywords: authentication by voice, acoustic features, clusterization, FRR and FAR.

Анотація

Привалов М.В., Кулібаба О.В. Вибір ознак для аутентифікації по голосу в комп'ютеризованій системі контролю доступу. Розглядаються переваги системи аутентифікації по голосу. Аналізується використання акустичних ознак дикторів в різних системах. Пропонується вибір оптимально набору ознак.

Ключові слова: аутентифікація по голосу, акустичні ознаки, кластеризація, помилки першого і другого роду.

Здано в редакцію:
01.04.2010р.

Рекомендовано до друку:
д.т.н, проф. Зорі А.А.