Текстозависимая верификация диктора по голосу на основе коллектива решающих правил

Т.В. Левковская,
кандидат технических наук

Рассматриваются основные этапы и методы обработки речевых сигналов при решении задач автоматического распознавания личности по голосу. Описывается экспериментальная система текстозависимой верификации диктора на основе коллектива решающих правил, в которой принятие решения выполняется путём анализа и объединения оценок трёх классификаторов: на основе методов динамического программирования, векторного квантования и сравнения интегральных характеристик голоса. Приводятся результаты экспериментальных исследований надёжности верификации дикторов на речевом материале изолированно произнесённых названий цифр, показывающие эффективность применения принципов коллективного принятия решения в задачах автоматического распознавания диктора по голосу.

Abctract

The basic stages and methods of speech signals processing for automatic person recognition by his/her voice are considered. The experimental text-dependent speaker verification system based on multi-stream approach is described. The decision-making is carried out by the analysis and fusion of the scores of three classifiers based on dynamic time warping (DTW) methods, vector quantization (VQ) and comparison of integrated voice characteristics. Results of experimental researches of speaker verification reliability on the speech material of isolated digits are given. The obtained results show effectiveness of multi-stream approach in automatic speaker recognition tasks.

Введение

В настоящее время большое внимание уделяется развитию биометрических технологий, которые предназначены для получения и использования индивидуальных биологических данных человека, называемых биометриками, в целях его идентификации.

Наряду с такими биометриками, как отпечаток пальца, рисунок радужной оболочки глаза, структура ДНК и др., использование индивидуальных характеристик голоса предоставляет бесконтактный, этически корректный способ получения биометрической информации, позволяет осуществить скрытое наблюдение за человеком и его идентификацию, обеспечивает возможность удалённого доступа к конфиденциальной информации, в том числе по телефону.

Задача распознавания диктора по голосу может быть разделена на две подзадачи: идентификация и верификация. Идентификация диктора — это процесс определения говорящего из заданного набора дикторов. Распознаваемый голос сравнивается с эталонными голосами, и из набора выбирается тот диктор, голос которого в наибольшей степени соответствует данному. В случае верификации говорящий вначале предъявляет свой идентификатор (объявляет, кто он такой), а затем система определяет, принадлежит ли распознаваемый голос диктору с указанным идентификатором или нет. В задаче верификации при росте числа пользователей время принятия решения не увеличивается и является постоянным для различного числа пользователей. Это определяет возможность более широкого применения систем верификации, чем систем идентификации.

В последнее время разрабатывается множество экспериментальных и коммерческих приложений распознавания личности по голосу [1], которые применяются в банковских системах, системах безопасности и массового обслуживания в целях контроля доступа. Как правило, биометрические системы контроля доступа (пропускного контроля, доступа к информации, физической защиты закрытых объектов от несанкционированного проникновения посторонних лиц и т.п.) строятся на принципах автоматической верификации. Большинство верификаторов личности по голосу являются зависимыми от текста. Верификация диктора осуществляется по фиксированной парольной фразе, которая может быть изменена, или по конкретным словам, порядок произношения которых определяется самой системой случайным образом. В последнем случае снижается возможность фальсификации голосового пароля.

1. Принципы построения верификатора диктора по голосу

Основные компоненты и принципы функционирования систем автоматического распознавания личности по голосу детально описаны в ряде фундаментальных работ [2, 3]. Верификатор диктора по голосу функционирует в двух режимах: обучения и распознавания. Режим обучения предназначен для создания эталонных моделей голосов пользователей системы. Каждой модели ставится в соответствие идентификатор диктора, на основании речи которого была построена модель. Созданные эталоны сохраняются в базе данных с указанием персонального идентификатора (кода). В режиме распознавания пользователь вводит свой код и произносит пароль. Основные этапы обработки речевых сигналов при распознавании диктора по голосу таковы:

- предобработка и выделение информативных признаков, характеризующих индивидуальные особенности голоса человека;
- классификация, т.е. сравнение информативных признаков с эталонами и вычисление оценки соответствия;

- принятие решения об индивидуальности говорящего путём сравнения полученной оценки соответствия с заранее установленным пороговым значением.

Информативными признаками, наиболее часто используемыми в современных системах автоматического распознавания речи, являются кепстральные коэффициенты, рассчитанные по параметрам линейного предсказания речи, и мел-кепстральные коэффициенты, полученные с помощью дискретно-косинусного преобразования. Для сохранения информации о динамике речевых характеристик параметрическое описание обычно дополняется дельта-параметрами, которые представляют собой производные по времени от полученных признаков.

Последовательность векторов параметров является динамической моделью. Распространённым способом нелинейного во времени сравнения эталонной и анализируемой последовательностей векторов параметров является метод динамического программирования. Моделирование речи при помощи методов векторного квантования является более компактным описанием признакового пространства при формировании эталонных моделей голосов и эффективным способом распознавания диктора по голосу. Модель векторного квантования представляет собой конечное число типичных для конкретного диктора векторов параметров, совокупность которых образует кодовую книгу. Сравнение интегральных характеристик, вычисленных на продолжительных интервалах речи, таких как средний основной тон, средний спектр, кепстр сигнала и др., также достаточно эффективно используется для выявления индивидуальных особенностей голоса диктора.

В задачах распознавания диктора широко используются вероятностные модели, к которым относятся скрытые Марковские модели, модели Гауссовых смесей.

В последнее время при реализации систем автоматической верификации диктора применяются комбинации как разных наборов информативных признаков, так и разных способов классификации [4, 5]. Принятие решения осуществляется путём анализа и объединения полученных оценок используемых классификаторов.

2. Способы оценки эффективности систем верификации

Основным критерием качества биометрических систем являются вероятности ошибок первого и второго рода. Ошибка первого рода — это вероятность ложного отказа в доступе клиенту, имеющему право доступа (FRR — False Rejected Rate). Ошибка второго рода — вероятность ложного доступа, когда система ошибочно опознает чужого как своего (FAR — False Accepted Rate). Коэффициент равной вероятности ошибок (EER — Equal Error Rate) представляет точку совпадения вероятностей ошибок первого и второго рода. Изменение соотношения ошибок 1-го и 2-го рода достигается за счёт изменения порога принятия решения.

Система с двумя типами ошибок имеет много возможных уровней порога принятия решения. Для оценки качества таких систем традиционно используется кривая относительной характеристики функционирования (ROC — Receiver Operating Characteristics) [3]. В общем случае вероятность ложного срабатывания (FAR) откладывается по горизонтальной оси, а вероятность верного распознавания (1 — FRR) — по вертикальной оси.

Для задач верификации диктора также применяется оценка, отражающая компромисс ошибок детектирования (DET — Detection Error Tradeoff) [2]. В случае DET-графика значения

ошибки откладываются по обеим осям (FAR — по горизонтальной оси, FRR — по вертикальной оси), что позволяет чётко отличать эффективность распознавания одной системы от другой.

3. Структура верификатора диктора по голосу на основе коллектива решающих правил

Для повышения надёжности верификации предлагается использовать подход, основанный на принципах коллективного принятия решения. В разработанной экспериментальной системе автоматической верификации диктора по голосу (***рис. 1***) предлагается использовать три типа классификаторов: на основе методов динамического программирования (ДП), векторного квантования (ВК) и сравнения интегральных характеристик (интегральный классификатор — ИК).

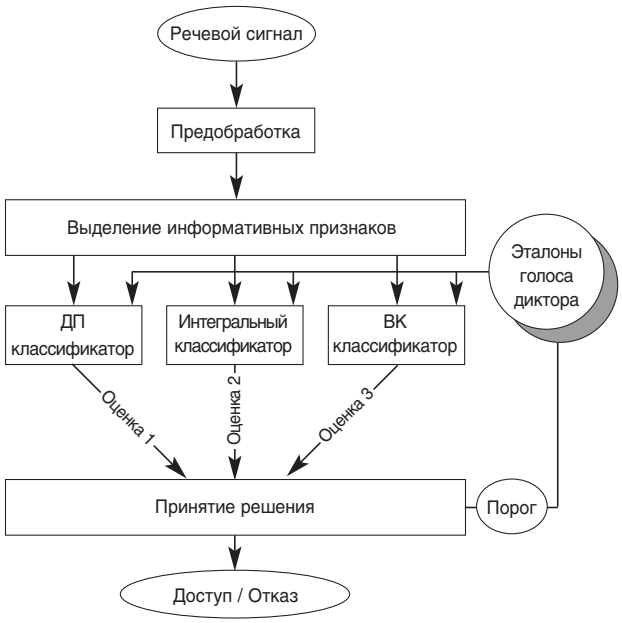

Рис.1. Общая схема верификатора на основе коллектива решающих правил

В качестве параметрического описания речевого сигнала использовались кепстральные коэффициенты, рассчитанные с помощью дискретно-косинусного преобразования, и их дельта-параметры. Нелинейное во времени сравнение параметрических описаний анализируемой речевой реализации и подготовленного в процессе обучения динамического эталона выполняется модифицированным ДП-методом [6, 7]. Главным достоинством этого метода является то, что он позволяет определить вероятность присутствия распознаваемых элементов речи в непрерывном речевом потоке и оценить их временное местоположение в условиях разного рода акустических помех.

На основе динамических параметров создаётся интегральная модель, представляющая собой средний кепстр по всей речевой реализации парольной фразы, а также формируется кодовая книга векторов параметров с помощью алгоритма векторного квантования. В режиме распознавания эталоны интегральной модели сравниваются с интегральными признаками входной речевой реализации, а кодовая книга сравнивается не с входными векторами, а с кодовой книгой входной реализации сигнала, подвергнутого такой же процедуре векторного квантования, как и при обучении.

Результатом сравнения каждого из классификаторов является расстояние между двумя моделями. Если расстояние меньше порогового значения, то диктор опознаётся как тот, за кого он себя выдаёт. Окончательное принятие решения может осуществляться как путём сравнения взвешенной суммы полученных оценок используемых классификаторов с порогом (рис. 1), так и путём объединения частных решений каждого классификатора в форме голосования.

4. Результаты экспериментальных исследований надёжности верификации дикторов

Исследования по надёжности верификации были проведены на одном и том же речевом материале для всех дикторов (названия цифр). Для тестирования была создана экспериментальная речевая БД, включающая образцы голосов 35 взрослых дикторов (13 женщин, 22 мужчин). Запись производилась с периодичностью 14 дней в офисных условиях. Каждый диктор произносил цифры от 0 до 9 по 2 раза. Речевые сигналы записывались в монорежиме с частотой дискретизации 11025 Гц, 16 бит на отсчёт. Всего было выполнено 5 серий записи.

БД условно была разделена на две части. Первая часть включала 15 дикторов (5 женщин и 10 мужчин). Эталоны каждого диктора создавались на речевом материале первых двух серий записи, после чего определялись пороговые значения для принятия идентификационного решения. Остальные серии записи использовались для тестирования. Во вторую часть входили оставшиеся дикторы, которые не участвовали в процессе обучения. Экспериментальная оценка эффективности системы проводилась на речевом материале обеих групп дикторов.

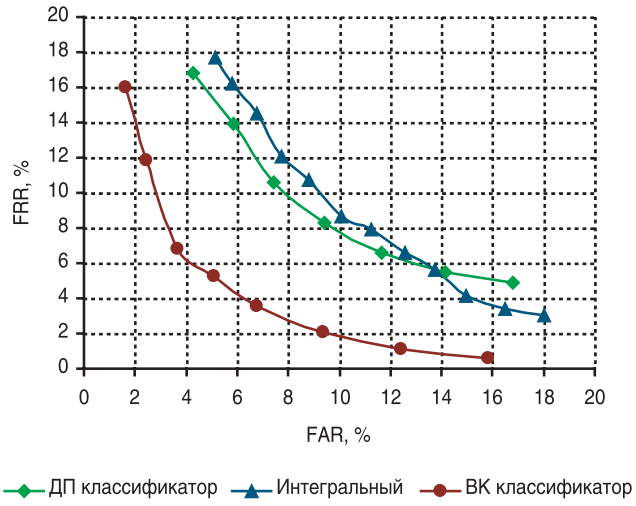

Рис. 2. DET зависимости ошибок верификации при использовании трёх типов классификаторов

На ***рис. 2*** показаны DET зависимости процента ошибок первого (FRR) и второго (FAR) рода при независимом использовании классификаторов ДП, ИК и ВК. Экспериментальные данные получены на речевом материале изолированно произнесённых названий цифр первой группы дикторов. Представленные результаты наглядно демонстрируют более низкий уровень ошибок верификации (примерно в 2 раза) при использовании ВК классификатора по сравнению с двумя остальными.

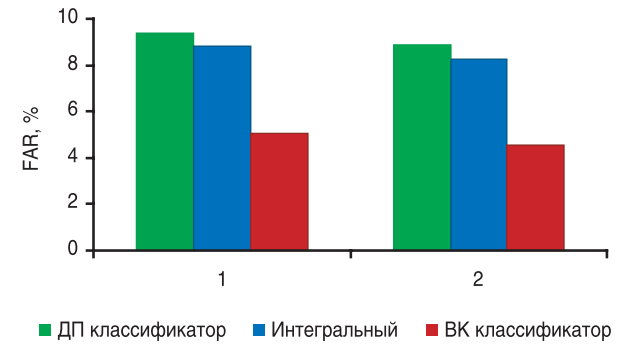

Рис. 3. Ошибки ложного доступа для своих (1) и чужих (2) дикторовпри использовании трёх типов классификаторов

Результаты ошибок ложного доступа (FAR) для двух групп дикторов представлены на ***рис. 3***. В первую группу входили свои дикторы, эталонные модели голосов которых были созданы в процессе обучения; во вторую — чужие дикторы, для которых эталоны отсутствовали. Из рис. 3 следует, что средние значения FAR практически не отличаются для своих и чужих дикторов при использовании каждого из классификаторов.

Во второй серии экспериментов были получены сравнительные результаты верификации на речевом материале разной длительности. Принятие решение выполнялось на основе анализа частных решений, полученных для каждой из произнесённых цифр, в форме голосования. Зависимость коэффициента равной вероятности ошибок (EER) от длительности анализируемой последовательности цифр, состоящей из одной, трёх, пяти, семи и десяти цифр соответственно, изображена на ***рис. 4***. Значение EER уменьшается примерно по экспоненте для каждого классификатора. Лучший результат получен для ВК классификатора.

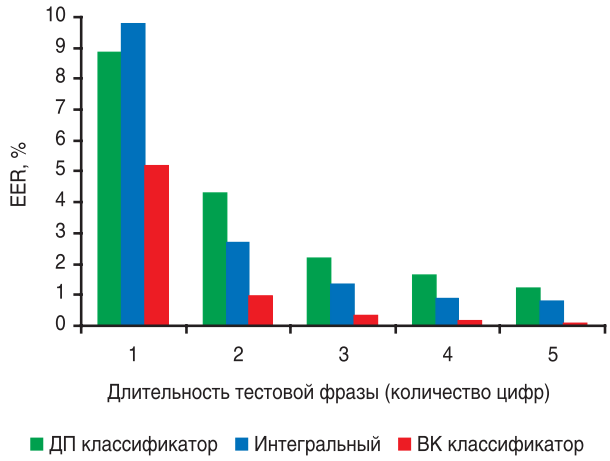

Рис. 4. Зависимость коэффициента равной вероятности ошибок (EER) от длительности анализируемой последовательности цифр (1, 3, 5, 7 и 10 цифр соответственно)

Последняя серия экспериментов связана с исследованием надёжности верификации при совместном использовании всех трёх классификаторов, а также их комбинаций. Средние значения ERR первой группы дикторов (***таб. 1***) и FAR двух групп дикторов (***таб. 2***) получены при анализе речевых сообщений разной длительности. Как и в предыдущей серии экспериментов, длительность тестовой фразы определялась количеством слов (1, 3, 5, 7 и 10 цифр).

Таблица 1

Коэффициенты равной вероятности ошибок (%) текстозависимой верификации дикторов на основе коллектива решающих правил

Классификаторы	Длительность тестовой фразы (кол-во цифр)				
	1	3	5	7	10
ДП, ВК	5,249	2,271	1,686	0,271	0
ИК, ВК	6,88	1,849	0,832	0,33	0,067
ДП, ИК	6,597	1,822	0,599	0,253	0,202
ДП, ИК, ВК	5,377	1,313	0,387	0,119	0

Таблица 2

Ошибки ложного доступа (%) текстозависимой верификации своих (1) и чужих (2) дикторов на основе коллектива решающих правил

Классификаторы	Группа дикторов	Длительность тестовой фразы (кол-во цифр)				
		1	3	5	7	10
ДП, ВК	1	5,782	0,502	0,204	0,086	0
	2	5,221	0,803	0,55	0,501	0,404
ИК, ВК	1	6,779	0,963	0,474	0,283	0,067
	2	6,368	1,406	1,021	0,923	0,779
ДП, ИК	1	6,59	1,002	0,506	0,348	0,202
	2	6,274	1,522	1,138	1,031	0,966
ДП, ИК, ВК	1	5,849	0,615	0,265	0,16	0
	2	5,246	1,066	0,817	0,765	0,685

Сравнительный анализ полученных результатов показывает значительное снижение ошибок верификации при увеличении длительности анализируемой тестовой фразы. Результаты экспериментальных исследований позволяют сделать вывод, что использование нескольких классификаторов обеспечивает уменьшение коэффициента равной вероятности ошибок первого (ложный отказ в доступе) и второго (ложный доступ) рода и увеличивает надёжность верификации.

Заключение

Результаты экспериментальных исследований наглядно демонстрируют эффективность применения принципов коллективного принятия решения в задачах автоматического распознавания диктора по голосу. Коллектив решающих правил в разработанной системе может быть дополнен. Система открыта для использования дополнительных критериев принятия решения с целью дальнейшего повышения надёжности верификации и устойчивости системы в реальных условиях её функционирования.

Направления дальнейших исследований связаны с разработкой формантного анализатора речевых сигналов, анализом просодических характеристик речи, использованием статистических методов распознавания голосов дикторов, применением аппарата нечёткой кластеризации для решения задачи отбора наиболее информативных признаков и классификации в условиях присутствия шумов и помех, проведением экспериментальных исследований с использованием доступных речевых баз данных.

Основные результаты получены в ходе выполнения научно-исследовательской работы «Разработка экспериментальной бимодальной биометрической системы контроля доступа

на основе характеристик лица и голоса человека» государственной комплексной программы научных исследований «Научные основы информационных технологий и систем».

Литература

1. *Хитров М.В.* Речевые технологии на CEBIT 2008 // Речевые технологии. — С.-Петербург: Издательский дом «Народное образование», 2008, № 2. С. 79–80.

2. *Rosenberg A.E., Soong F.K.* Recent research in automatic speaker recognition, in Advances in Speech Signal Processing, Furui, S. and Sondhi, M.M., Eds.,Marcel Dekker, New York, 1991. — P.701–738.

3. *Furui S.* «An overview of speaker recognition technology», ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, 1994. — P.1–9.

4. *Rylov A.S., Chyzhdzenka V.A., Leukouskaya T.V.* The discriminant-stochastic approach of the speaker verification for entry control by the biometrical technologies // Proc. of the 9-th Intern. Conf. «Speech and Computer — SPECOM'2004, St.-Petersburg, 2004. — P. 377–381.

5. *Gupta H., Hautamaki, V., Kinnunen, T., Franti, P.* «Field Evaluation of Text-Dependent Speaker Recognition in an Access Control Application», Proc. of the 10th International Conference «Speech and Computer» — SPECOM'2005, Patras, Greece, 17–9 October, 2005. — P. 551–554.

6. *Lobanov B., Levkovskaia T.* Recognition of words and words-sequences in running speech // Proc Digital image processing. — Minsk: Institute Eng. Cybernetics, Academy of Science of Belarus, Minsk, 1997. — P. 154–161.

7. *Левковская Т.В.* Идентификация спектральных изображений речи // Анализ цифровых изображений. Мн.: ОИПИ НАН Беларуси, 2003. С. 186–193.

Левковская Т.В. —

кандидат технических наук, старший научный сотрудник Объединённого института проблем информатики НАН Беларуси. В 1997 г. защитила кандидатскую диссертацию «Исследование и разработка методов фонемного распознавания речи» (научный руководитель — доктор технических наук Б.М. Лобанов).
Область научных интересов — анализ сигналов, распознавание образов.